This chapter appeared in *Vizualisation of Categorical Data*,
Blasius & greenacre (Ed), Academic Press, 1998.

Note of March 2019.

The *data* of "*The French Worker Survey*" can be found on my web site under the folder "Data - R scripts - Texts".

`http://helios.mi.parisdescartes.fr/~lerb/Logiciels/software.html`

The strategy of the analysis presented in this chapter can easily be performed using Coheris-SPAD software. A SPAD project is available on my web site.

An *Appendix* that explains the computation of the contribution of a ternary comparison has been added after the *References* (p. 223).

Examples of application of the method to several data sets can be found in Bernard et al. (1989); Bonnet et al. (1996); Hjellbrekke et al. (2007); Gounelle and Le Roux (2007); Le Roux et al. (2008) Le Roux and Rouanet (2004, 2010, 2013); Le Roux (2014); Lebaron and Le Roux (2015); Borjesson et al. (2016).

Brigitte le Roux

Chapter 16

Interpreting Axes in Multiple Correspondence Analysis: Method of the Contributions of Points and Deviations

Brigitte Le Roux and Henry Rouanet

1 Introduction

In geometric data analysis, once a "cloud" of points has been constructed, as the outcome of correspondence analysis (CA), for example, or principal component analysis, the phase of interpretation follows. This phase is always a delicate one; at this point, the need to fill the gap between theory and practice appears essential—a need well reflected in the book edited by Greenacre and Blasius (1994a). In the French tradition of data analysis, *aids to interpretation* have been devised, such as the familiar table of contributions and supplementary elements. The method we will present in this chapter, namely the method of the *contributions of points and deviations*, directly extends the existing aids to interpretation. It stems from the following remark: in analysis of variance (ANOVA) terms, contributions of points to an axis are simply *parts of variance* accounted for by points. This leads to considering other parts of variance that are also used in ANOVA; for example, those that express *contrasts* among groups of observations. That is, it leads us to study the contributions of *deviations* between points. Indeed, all those who practice geometric data analysis are accustomed to think intuitively in such terms ("axis 1 opposes rich vs. poor, axis 2 old vs. young, etc."). From a theoretical viewpoint, the statistical interpretation of CA in ANOVA terms is well known; see Fisher (1940) and Tenenhaus and Young (1985). But we feel that the idea deserves to be fully elaborated.

This chapter will be mainly devoted to the first and basic phase of interpretation, namely that of the principal axes, in the case of multiple correspondence analysis (MCA). Henceforth we assume the data structure of a *questionnaire* in standard form; that is, there is a set of *questions*, together with, for each question, a set of *response modalities* (also called response categories)—including nonresponse whenever relevant—and each individual chooses one (and only one) modality of each question. Then consider the following two ideas taken from the nested designs in ANOVA:

1. With each modality is associated one and only one question; in ANOVA terms, this means that the set of all modalities is *nested* in the set of questions. This prompts us to investigate—in addition to contributions of modalities—the *contributions of questions to axes*, and also the *contributions of modalities to questions*.
2. For each question, each individual chooses one and only one modality, which means that for each question the set of individuals is *nested* in the set of the observed modalities of the question. In other words, each question generates a partition of individuals indexed by the modalities of the question. This suggests that we investigate the *cloud of individuals* and its *subclouds* associated with modalities of interest.

The method of the contributions of points and deviations will be illustrated with data taken from the The French Worker Survey.

2 The French Worker Survey

2.1 The Survey

The *French Worker Survey* (Adam *et al.*, 1970) was conducted in July 1969, on a representative sample of French workers—unskilled, specialized and technicians—using a thorough battery of 70 questions, with the overall objective of "analyzing the political and social behavior of the working class".

At the time of the survey, presidential elections had just taken place, opposing the candidates of the four main political families, along the traditional range from left to right: Communist (Duclos), Socialist (Defferre), Center (Poher), Gaullist (Pompidou, who won the election). One objective of the survey was to inquire about this traditional dimension in the specific population of workers; other objectives were to identify and interpret other important dimensions, possibly specific to this population. For instance, the communist dominance among workers was beyond doubt (although already on the decline), but the influences and roles of the noncommunist left and of center were not so well delineated. Also, what needed to be clarified were the relations and interplay between political attitudes and attitudes toward trade unions. The leading trade unions were the CGT (with notorious links with Communist party), then—far behind—CFDT, FO (both loosely linked with the noncommunist left), and "autonomous" (inclined toward right wing).

Table 1: The four basic questions and their relative frequencies

Professional elections (q1). In professional elections in your firm, would you rather vote for a list supported by:		**Union affiliation (q2).** At the present time, are you affiliated to a Union, and in the affirmative, which one:	
1. CGT	.3298	1. CGT	.2107
2. CFDT	.0877	2. CFDT	.0524
3. FO	.0782	3. FO	.0229
4. CFTC	.0248	4. CFTC	.0048
5. Autonomous	.1077	5. Autonomous	.0210
6. Abstention	.1525	6. CGC	.0114
7. NonAffiliated list	.1049	7. NotAffiliated	.6663
8. NR	.1144	8. NR	.0105
Presidential election (q3). On the last presidential election [1969], can you tell me the candidate for whom you have voted?		**Political sympathy (q4).** Which political party do you feel closest to, as a rule?	
1. Jacques Duclos (Comm.)	.2221	1. Communist [PCF]	.1935
2. Gaston Defferre (Soc.)	.0467	2. Socialist [SFIO+PSU+FGDS]	.1697
3. Alain Krivine	.0095	3. "Left" ("Party of workers",...)	.0429
4. Michel Rocard	.0286	4. Center [+MRP+RAD.]	.1192
5. Alain Poher (Center)	.1420	5. RI	.0086
6. Louis Ducatel	.0067	6. Right [+INDEP.+CNI]	.0381
7. Georges Pompidou (Gaullist)	.2336	7. Gaullist [UNR]	.1335
8. NRAbst	.3108	8. NR	.2946

Note. Within union questions $q1$ and $q2$, there are correspondences between modalities (except 6), reflected by label numberings. Similarly for modalities 1,2, 7 and 8 of political questions $q3$ and $q4$, the other label numbers being arbitrary. There are no such correspondences between union and political parties, except for the well-known affinities between CGT and Communist (modalities 1).

The analysis to be presented in this chapter is based on 1049 respondents and concentrates mainly on two questions about trade unions and two questions about political preferences. The four basic questions, each with eight modalities of response are presented in Table 1, together with the associated relative frequencies. In Table 2, the 319 observed *response patterns* are given with their frequency counts.

(Let us briefly comment on the one–way tables.) From the union questions ($q1$ and $q2$), we see that 63% of workers vote for a list sponsored by some Union ($q1$, modalities 1–5 and 7), more than half of them for CGT; 67% of workers, however, are not affiliated with any union ($q2$, modality 7). From the two political questions ($q3$ and $q4$), we see the high percentages of nonresponses, NR (31% and 29%). Among expressed sympathies, the communist party indeed comes first (19%), but is exceeded by noncommunist left-wing sympathies pooled together (21%, $q4$, modalities 2 and 3); also, the Gaullist pooled with other right-wing parties ($q4$, modalities 5 and 6) come up to 18%. Duclos' score (22%) is exceeded by Pompidou's (23%), and so on.

Table 2: 319 response patterns with frequency counts

1111	81	1712	4	2234	1	3356	1	4722	1	5751	1	6776	4	7787	4
1112	9	1717	1	2242	2	3357	1	4732	1	5752	3	6777	19	7788	16
1113	7	1718	7	2251	1	3358	1	4753	1	5754	10	6778	5	8111	1
1114	2	1721	1	2252	6	3374	1	4756	1	5756	3	6781	1	8113	1
1118	7	1722	5	2254	8	3377	2	4766	1	5757	3	6782	5	8152	1
1122	5	1728	1	2258	2	3378	1	4773	1	5758	7	6783	4	8154	1
1126	1	1738	1	2261	1	3384	1	4774	2	5772	1	6784	8	8181	2
1128	2	1742	2	2274	3	3388	2	4777	7	5774	5	6786	4	8182	1
1132	1	1748	1	2276	2	3554	1	4778	3	5775	1	6787	4	8188	2
1142	4	1751	3	2282	3	3614	1	4782	1	5776	2	6788	50	8288	1
1146	1	1752	5	2284	1	3662	1	5113	1	5777	14	7111	2	8322	1
1148	2	1754	3	2285	1	3711	2	5132	1	5778	4	7112	1	8588	1
1151	3	1757	1	2286	1	3712	1	5142	1	5781	2	7154	1	8677	2
1152	3	1758	4	2287	1	3713	1	5161	1	5782	2	7177	1	8678	1
1153	2	1771	1	2288	1	3714	2	5174	1	5784	1	7181	1	8711	3
1154	2	1772	3	2711	3	3722	3	5184	2	5787	3	7522	1	8712	4
1158	3	1774	3	2728	3	3724	1	5187	1	5788	9	7582	1	8713	1
1161	1	1775	1	2737	1	3732	1	5354	1	5876	1	7588	1	8718	4
1162	1	1776	2	2738	1	3751	1	5382	1	6116	1	7711	9	8741	1
1171	1	1777	7	2742	3	3752	2	5512	1	6172	1	7712	2	8742	1
1172	3	1778	5	2744	1	3754	4	5513	1	6178	1	7713	1	8751	1
1177	5	1781	8	2752	1	3755	1	5518	1	6181	1	7716	1	8752	1
1178	3	1782	9	2754	3	3756	1	5522	2	6182	1	7718	1	8753	1
1181	10	1783	3	2756	2	3758	5	5548	1	6188	2	7722	2	8754	1
1182	7	1784	4	2772	1	3774	4	5574	2	6528	1	7742	2	8757	1
1183	5	1786	1	2774	3	3775	2	5575	1	6676	1	7752	2	8758	4
1184	1	1787	2	2777	7	3776	1	5577	4	6711	8	7754	6	8765	1
1188	13	1788	26	2778	5	3777	7	5584	1	6712	1	7756	1	8774	2
1218	1	1858	1	2782	1	3778	4	5588	1	6714	1	7758	5	8776	2
1272	1	1881	2	2784	1	3782	3	5672	1	6718	5	7772	1	8777	12
1288	1	2111	1	2787	2	3783	1	5674	1	6722	3	7774	1	8778	9
1311	1	2132	1	2788	3	3784	3	5677	1	6742	2	7775	1	8781	2
1381	1	2154	1	3122	1	3787	1	5711	1	6752	3	7776	2	8782	2
1418	1	2178	1	3182	1	3788	6	5712	5	6753	3	7777	22	8783	5
1481	1	2211	2	3277	1	4241	1	5713	2	6754	6	7778	11	8784	3
1552	1	2214	1	3311	2	4254	1	5722	1	6756	2	7781	2	8788	37
1611	1	2218	1	3312	1	4274	1	5728	1	6758	4	7782	3	8822	1
1673	1	2222	7	3322	3	4441	1	5732	1	6771	1	7783	2	8878	1
1677	1	2223	1	3342	2	4477	2	5742	1	6772	3	7784	3	8888	5
1711	33	2224	1	3354	1	4712	1	5744	1	6774	5	7786	1		

We might continue by commenting on two-way and higher way tables. Looking at the four-way table amounts to considering response patterns (Table 2). The most frequent pattern (81 individuals) is 1111, describing the CGT-Communist “hard core”: CGT vote and affiliation, Duclos vote and communist sympathy. Next comes the pattern 6788 (50 individuals), that is, abstention and nonaffiliation for union questions, and nonresponse for the political ones. The 14 most frequent patterns together represent about one third of the total number of respondents.

In the book by Adam *et al.* (1970), the reader will find one-way and two-way tables for the most important questions, with extensive sociological comments, organized by topics—for example, attitudes toward unions, electoral behavior—based on careful examination of tables. The alternative approach that we will follow in this chapter, along the line of geometric data analysis, is to construct a relevant "social space" (as Bourdieu would call it), a "union–political space" for the French workers in 1969, applying MCA to the responses to the four questions. The study of maps yielded by MCA amounts to a synthesis of analyses of the conventional kind.

2.2 Multiple Correspondence Analysis (MCA)

From the responses of the individuals, we construct the *disjunctive table* (Benzécri, 1992, p. 392; Lebart *et al.*, 1995, p. 108), also called an indicator matrix, crossing the 1049 individuals and the $8 \times 4 = 32$ modalities. The principle of construction of this table is recalled by Table 3.

Correspondence analysis of the disjunctive table, that is, multiple correspondence analysis, yields two clouds of points, namely the cloud of 32 modalities, and the cloud of 1049 individuals—or equivalently of 319 weighted response patterns. In numerical terms, each cloud is defined by a table of principal coordinates, where for each axis the weighted average of the squares of principal coordinates is equal to the eigenvalue associated with the axis.

Here, we will interpret the first four axes; the corresponding eigenvalues are given in the first row of Table 4. For the cloud of modalities in the plane 1-2, see Figure 1.

- On the left, a compact group of four modalities emerges: vote and affiliation CGT, Duclos, Communist.
- On the lower right, there are the various NR and abstention modalities, together with the two nonaffiliated modalities, Pompidou and Gaullist. Moving up, we find Center and Poher, Socialist and Defferre, then CFDT vote and affiliation.

Table 3: Disjunctive table

Patterns			$q1$	$q2$	$q3$	$q4$
1 1 1 1	81		10000000	10000000	10000000	10000000
...			...	...	...	...
1 1 1 1			10000000	10000000	10000000	10000000
		Disjunctive encoding $\Longrightarrow$				
8 8 8 8	5		00000001	00000001	00000001	00000001
...			...	...	...	...
8 8 8 8			00000001	00000001	00000001	00000001

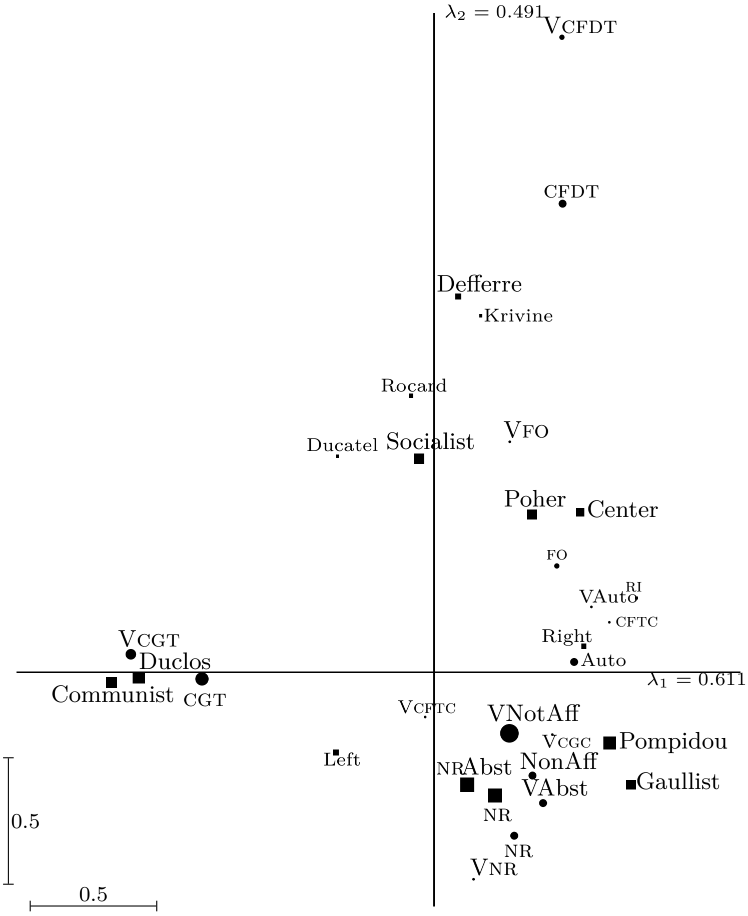

Figure 1: Cloud of 32 modalities in plane 1-2. Modalities that contribute most to axes 1 and 2 are in large characters; modalities of the two union questions are represented by circles and those of the two political questions by squares, whose areas are proportional to frequencies. CGT voting is denoted VCGT, as distinct from CGT affiliation, denoted CGT, and so on.

3 Contributions of Points and Deviations

3.1 Basic Formulas

A cloud of weighted points being given, the *variance* (also called inertia) of the cloud is the weighted mean of the squares of the distances between the points and the mean point of the cloud (Benzécri, 1992, p. 36). The *absolute contribution* of a point to the cloud is defined as the product of the weight of the point by the square of its distance from the mean point (Benzécri, 1973a, p.38; 1992, p. 61). In this chapter, we will be mainly interested in contributions to an axis; accordingly, distances will be measured along the axis under consideration.

1. *Contribution of a point* (Cta). Let us consider a point of weight, or mass, p and coordinate y along the axis. The *absolute contribution* of the point to the axis will be denoted Cta; it is given by the formula (Benzécri, 1992, p. 340; Greenacre, 1984, p. 67):
$$\mathrm{Cta} = p\,y^2 \qquad \text{(point)}$$
2. *Contribution of the deviation between two points* (Cti). Let us now consider two points. Let p and p' denote the weights of the points and y and y' their coordinates along the axis. The absolute contribution of the deviation, also called the *intra (within) contribution*, will be denoted Cti and is given by the following formula (Rouanet, Le Roux, 1993, p. 268):
$$\mathrm{Cti} = \frac{pp'}{p+p'}\,(y-y')^2 \qquad \text{(deviation)}$$

These notions of contribution readily extend to a subset of points, or subcloud. With a subcloud are associated its *weight* (sum of the weights of its points), its *weighted mean point* (barycenter), and its *variance*, and the following three types of contribution:

- Its (global) contribution (Cta), which is the sum of the contributions of its points
- The absolute contribution (Cta) of its mean point, which is the product of its weight by the square of the principal coordinate of its mean point
- Its intra-contribution (Cti), which is the weighted sum of squared distances from the point to their mean point

By the classical Huyghens property, the Cta of a subcloud is the sum of its Cti and the Cta of its mean point, which shows that Cta and Cti are equal if and only if the mean point of the subcloud coincides with the mean point of the cloud (Rouanet & Le Roux, 1993, p. 118).

3.2 Application to the Cloud of Modalities

In MCA, the weight of a modality is the relative frequency of this modality divided by the number of questions. Hereafter we illustrate the calculations for axis 1.

- *Contribution of modality* (Cta). Taking CGT vote (denoted VCGT) as an example: the relative frequency is 0.3298 (Table 1), hence the weight $p = 0.3298/4 = 0.0825$. The coordinate along axis 1 is $y = -1.090$ (see Table 5). Hence the absolute contribution of VCGT: $p\,y^2 = 0.0825 \times (1.090)^2 = 0.0980$.
- *Contribution of deviation between modalities* (Cti). Take VCGT (coordinate $y = -1.090$, weight $p = 0.0825$) on the one hand and VAuto and VAbst on the other hand; the barycenter of VAuto and VAbst has a weight equal to $p' = 0.0269 + 0.0381 = 0.0650$ (weights add up), and its coordinate is $y' = (0.0269 \times 0.659 + 0.0381{\times}0.513)/0.0650 = 0.573$ (coordinates average up). One has $pp'/(p+p') = (0.0825{\times}0.0650)(0.0825+0.0650) = 0.0364$. Hence the absolute contribution of the deviation $(-1.090 - 0.573)^2{\times}0.0364 = 0.1014$.
- *Contribution of modality to axis* (Ctr). Let us divide the contribution of VCGT, namely 0.0980, by the *sum of the contributions of all modalities*, that is, $\lambda_1 = 0.6113$; we get $0.0980/0.6113 = 0.160$, which means that VCGT contributes to 16% of Axis 1. This ratio is often denoted by Ctr.

We further define two other ratios that will be directly useful in the interpretation process; for clarity, we will always express them as percentages.

- *Contribution of question to axis.* By definition, the Cta of a question is the sum of the Ctas of its modalities. For example, the Cta of $q1$ (Professional Elections) for axis 1 is the sum of the eight Ctas: $0.0980 + \ldots + 0.0041 = 0.1482$ (see Table 5). If we now divide the contribution of $q1$ by the sum of the contributions of questions, that is, $\lambda_1 = 0.6113$, we get $0.1482/0.6113 = 0.24$; accordingly, we state that question 1 *accounts for* 24% of axis 1.
- *Contribution of modality (and of deviation) to question.* If we divide the contribution of VCGT by the *contribution of the question* it belongs to, namely Professional Elections ($q1$), we get $0.0980/0.1482 = 0.66$; therefore we state that VCGT contributes to 66% of question $q1$ (for axis 1). Similarly for the contributions of deviations. The deviation VCGT versus VAuto and VAbst contributes to $0.1014/0.1482 = 68\%$ of the question $q1$ (for axis 1).

3.3 Cloud of Individuals and Cloud of Modality Mean Points

In the cloud of individuals, with each observed modality is associated the subcloud of the individuals who have chosen that modality. The mean point of this subcloud will be called *modality mean point*. For each axis, the coordinate of the modality mean point is the mean of the principal coordinates of the individuals who have chosen this modality, and this can be shown to be equal to $\sqrt{\lambda}\,y$, where y is the principal coordinate of the corresponding modality (Benzécri, 1992, p. 410).

The cloud of all modality mean points can be obtained from the CA of the *Burt table*, which has as eigenvalues the squares λ^2. As a consequence, if one divides the contribution of a modality mean point—or of a deviation between modality mean points—by λ^2, one again finds the relative contribution (Ctr) of modality, or of deviation, and consequently, the relative contribution (Ctr) of a question to an axis.

Each question q induces a partition of individuals into as many subclouds as there are observed modalities for that question. Consider the derived cloud of the modality mean points for question q. For each axis, the variance of this cloud, or *interclass (between-class) variance*, is equal to λ times the absolute contribution (Cta) of question q in the cloud of modalities. As a consequence, if one divides the contribution of a modality mean point—or of a deviation between modality mean points—by the interclass variance, one again finds the relative contribution of modality—or of deviation—to the question.

As a conclusion, it will be equivalent to interpret axes in the cloud of modalities or in the cloud of modality mean points.

4 Interpreting Axes

Benzécri (1992, p. 405) gives the following guideline: "Interpreting an axis amounts to finding out what is similar, on the one hand, between all the elements figuring on the right of the origin and, on the other hand between all that is written on the left; and expressing with conciseness and precision, the contrast (or opposition) between the two extremes." The method of contributions of points and deviations has been devised as a guide along this line.

4.1 The Method of Contributions of Points and Deviations

As far as MCA is concerned, the method consists in the following four steps.

Step 1. *Important questions.* In the cloud of modalities, look for the questions whose contributions to the axis are important. This leads to a first overall interpretation of the axis.

Step 2. *Important modalities.* Select modalities—or groups of modalities of the same question that are close on the axis—whose contributions to the axis exceed some threshold (average contribution is a rule of thumb, but when the cumulated amount is not sufficient, a less severe threshold may be in order).

Step 3. *Contributions of modalities to questions.* For each question retained at step 1, calculate the relative contribution to the question (on the axis) accounted for by the modalities retained in step 2. When for the question under study, those modalities separate into several groups—often, for the first axes, into two groups on the two sides of the origin—determine the barycenters of groups, then their intra-contribution, and express this contribution as a percentage of the contribution of the question. For each question, the content of groups is a concise summary of the interpretation of the axis, whereas the relative intra-contribution to the question is a quantitative appraisal of the precision of that summary.

Table 4: Contributions (Cta) of the four questions

	Axis 1	Axis 2	Axis 3	Axis 4
Eigenvalue λ	**0.611**	**0.491**	**0.416**	**0.373**
$q1$ Professional elections	0.148	0.149	0.078	0.162
$q2$ Union affiliation	0.137	0.141	0.049	0.162
$q3$ Presidential election	0.157	0.105	0.148	0.024
$q4$ Political sympathy	0.169	0.096	0.141	0.026

Step 4. *Composite modalities or patterns.* The interpretation will be usefully complemented by the examination, in the cloud of individuals, of the composite modalities or patterns brought out at step 3. When interpreting a specific response pattern, be aware that its frequency count can be quite low.

4.2 First Overview

In the cloud of modalities, the contributions of the four questions to the first four axes are given in Table 4. The relative contributions of $q1$ through $q4$ to axis 1 lie between 22% and 28%; for axis 2, they lie between 19% and 31%. Therefore the interpretation of axes 1 and 2 will be based on the four questions. For axis 3, questions $q3$ and $q4$ contribute to 70% of the axis; therefore the interpretation of axis 3 will be based predominantly on the two political questions. For axis 4, $q1$ and $q2$ contribute to 87% of the axis; therefore the interpretation will be essentially based on the two trade union questions.

4.3 Interpretation of Axis 1

The interpretation of axis 1, in the cloud of modalities, is based on the results shown in Table 5, which may be used for checking the numerical values, with Figure 1 serving as an intuitive guide.

Step 1. *Important questions.* All four questions are important for axis 1; axis 1 is a general axis, that is, its interpretation involves all four questions.

Step 2. *Important modalities.* There are four very important modalities, namely Communist (Cta = 0.1111, i.e., 18% of axis), CGT (17%), Duclos (17%), VCGT (16%). Those four modalities together account for 69% of axis 1. They are all on the left side of axis 1. Three other modalities have contributions exceeding average (0.6113/32 = 0.0191), namely Pompidou (7%), Gaullist (5%) and NotAff (3%), all three on the right side of the axis. The previous seven modalities together contribute to 84% of axis 1. Let us add to them the two modalities VAuto and VAbst, which are close to each other on the axis and together contribute 3%, with the nine modalities we come up to 87%.

Table 5: Axis 1: weights, coordinates and absolute contributions (Cta) of modalities($\lambda_1 = 0.61132$)

Professional elections				Union affiliation			
q1	**Weight**	**Coord.**	**Cta**	***q2***	**Weight**	**Coord.**	**Cta**
1. VCGT	.0825	−1.090	.0980⋆	1. CGT	.0527	−1.425	.1069⋆
2. VFDT	.0219	0.605	.0080	2. CFDT	.0131	0.602	.0047
3. VFO	.0195	0.578	.0065	3. FO	.0057	0.356	.0007
4. VCFTC	.0062	0.824	.0042	4. CFTC	.0012	−0.040	.0000
5. VAuto	.0269	0.659	.0117+	5. Auto	.0053	0.741	.0029
6. VAbst	.0381	0.513	.0100+	6. CGC	.0029	0.557	.0008
7. VNonAff	.0262	0.463	.0056	7. NotAff	.1666	0.355	.0210⋆
8. VNR	.0286	0.377	.0041	8. NR	.0026	0.186	.0001
	.2500		.1482		.2500		.1373
Presidential election				**Political sympathy**			
q3	**Weight**	**Coord.**	**Cta**	***q4***	**Weight**	**Coord.**	**Cta**
1. Duclos	.0555	−1.387	.1069⋆	1. Comm.	.0484	−1.516	.1111⋆
2. Defferre	.0117	0.114	.0001	2. Soc.	.0424	−0.069	.0002
3. Krivine	.0024	0.221	.0001	3.“Left”	.0107	−0.460	.0027
4. Rocard	.0072	−0.108	.0001	4. Center	.0298	0.687	.0140
5. Poher	.0355	0.461	.0075	5. RI	.0022	0.950	.0019
6. Ducatel	.0017	−0.452	.0003	6. Right	.0095	0.705	.0047
7. Pompidou	.0584	0.826	.0398⋆	7. Gaull.	.0334	0.926	.0286⋆
8. nrAbst	.0777	0.156	.0019	8. NR	.0737	0.286	.0060
	.2500		.1568		.2500		.1690

Stars (⋆) refer to modalities whose contributions exceed the average of the axis (.61132/32 = .0191.) Plus (+) either refer to modalities close (on the axis) to a starred modality, or to clustered modalities whose grouped contribution exceeds average.

Step 3. *Contributions of modalities to questions.*

Professional elections ($q1$). The sum of the Cta of VCGT (on the left side), VAuto, and VAbst (on the right side) is $0.0980 + 0.0117 + 0.0100 = 0.1197$, that is, those three modalities contribute to $0.1197/0.1482 = 81\%$ of the question on the axis. The intra-contribution (Cti) of the deviation VCGT vs. VAuto with VAbst is found to be 0.1006; that is, it accounts for 68% of the question on the axis.

Union affiliation ($q2$). CGT (left) and NotAffiliated (right) together contribute 93% to the question on axis 1. The opposition between these two modalities accounts for 92% of the question.

Presidential election ($q3$). Duclos (left) and Pompidou (right) contribute to 94% of the question. The opposition Duclos vs. Pompidou accounts for 89%.

Political sympathy ($q4$). Communist (left) and Gaullist (right) contribute to 83% of the question. The opposition Communist vs. Gaullist accounts for 70%.

Step 4. *Relevant patterns.* The foregoing results suggest considering the composite modalities that emerge for axis 1. Since all (four) questions are involved in the interpretation of the axis, the relevant composite modalities are patterns, obtained by combining the cells of Table 6.

Hence the three relevant patterns (with frequency counts, out of a total of 1049): 1111 (81); 5777(14); 6777 (19). Figure 2 gives the simultaneous representation of relevant modalities and patterns for axis 1. It provides a graphical summary of the interpretation of axis 1, and the summary in words may read as follows. Axis 1 opposes the left profile VCGT-CGT-Duclos-Communist (1111) vs. the right profile [VAuto or VAbst]-nonaffiliated-Pompidou-Gaullist (5777, 6777).

4.4 Interpretation of Axis 2

Applying our four-step interpretation to axis 2 leads to the following results.

Step 1. Axis 2 is also a general axis (involving all four questions).

Step 2. Important modalities are CFDT, VCFDT, Socialist and Defferre (upper side of axis), then NR to $q4$, NRAbst, Poher, VNR (i.e. NR to $q1$), Center. Adding VAbst and NotAff (which are nearly average) and Gaullist (near NR to $q4$ on axis), one arrives at 90% of Axis 2.

Step 3. For $q1$, $q2$ and $q4$, there are well–marked oppositions. VCFDT (upper side) vs. VNR and VAbst (lower side) (92% of $q1$); CFDT (upper) vs. NotAff (lower) (92% of $q2$); Socialist and Center (upper) vs. NR and Gaullist (lower) (97% of $q4$).

Question $q3$ (presidential election) calls for a more detailed interpretation. Defferre, Poher and NRAbst together contribute 74% of $q3$. However, Poher lies halfway between Defferre and NRAbst, which means that those three modalities do not lend themselves easily to a grouping into two opposed

Table 6: Relevant modalities for axis 1

	Prof. Vote	**Union Aff.**	**Pres. Vote**	**Polit. Symp.**
Left	1. VCGT	1.CGT	1.Duclos	1.Comm.
Right	5.VAuto 6.VAbst	7.NotAff	7.Pompidou	7.Gaull.

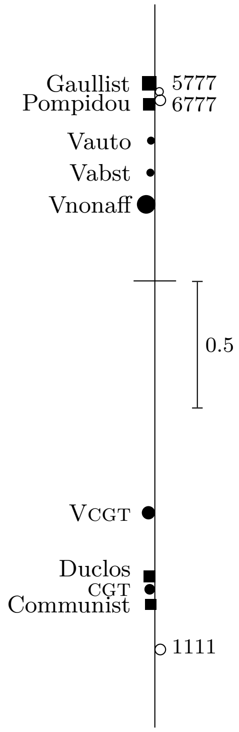

Figure 2: Axis 1: simultaneous representation of relevant modalities and patterns.

classes. This difficulty is confirmed by the weakness of the contribution of the opposition Defferre and Poher vs. NRAbst (only 65%). To get a more substantial contribution to the question, one must resort to the "ternary" comparison between Defferre vs. Poher vs. NRAbst, which accounts for 74% of question $q3$.

Step 4. The composite modalities that emerge from the analysis of axis 2 are obtained by combining the cells of Table 7.

Hence there are eight patterns (with frequency counts): 2222 (7); 2224 (1); 2252 (6); 2254 (8); 6787 (4); 6788 (50); 8787 (0) (a nonobserved pattern!); 8788 (37). Figure 3 gives the simultaneous representation of relevant modalities and patterns for axis 2. On the whole, axis 2 reflects the opposition between noncommunist left workers, with CFDT vote and affiliation, and nonrespondent nonaffiliated workers.

Table 7: Relevant modalities for axis 2

	Prof. Vote	**Union Aff.**	**Pres. Vote**	**Polit. Symp.**
Above	2.VCFDT	2. CFDT	2. Defferre 5. Poher	2. Socialist 4. Center
Below	6. Vabst 8. VNR	7. NotAff	8. NRAbst	7. Gaull. 8. NR

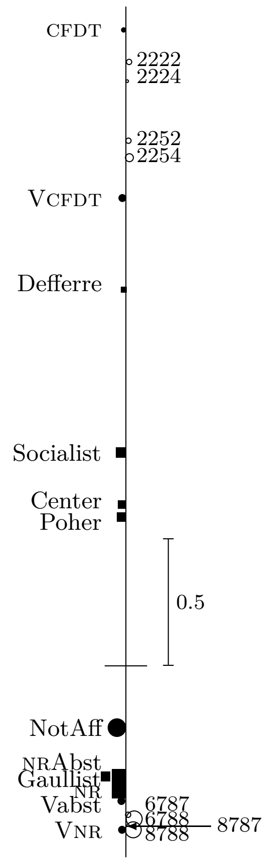

Figure 3: Axis 2: simultaneous representation of relevant modalities and patterns.

4.5 Interpretation of Axis 3

We summarize the results.

Step 1. Axis 3 is predominantly a political axis.

Step 2. The important modalities are Gaullist, Pompidou (on one side of the axis), NRAbst, and NR (on the other side), all four belonging to $q3$ and $q4$; then come three modalities of $q1$: VCFTC and VAuto (on the Gaullist side) and VNR (on the NR side). Those seven modalities together account for 76% of axis 3.

Step 3. The opposition Pompidou vs. NRAbst contributes to 90% of $q3$; the opposition Gaullist vs. NR to 89% of $q4$.

Step 4. In the cloud of individuals, the important modalities of questions $q3$ and $q4$ induce a subcloud of 114 Pompidou-Gaullists (patterns xx77), and a subcloud of 177 "political nonrespondents" (patterns xx88). Figure 4 shows the simultaneous representation of important modalities and of those two subclouds with their mean points. As may be seen, the separation between the two subclouds is perfect. Notice the "union–committed" patterns 4x77 and 5x77 (among Pompidou-Gaullist), and the noncommitted patterns8x88 (among political nonrespondents).

Axis 3 is predominantly political and opposes politically committed Pompidou-Gaullist workers to political nonrespondents.

4.6 Interpretation of Axis 4

Step 1. Axis 4 is predominantly a union axis.

Step 2. The important modalities are FO, VFO (on one side of the axis), VCFDT and CFDT (opposite side), Socialist, and Defferre: together 87% of the axis.

Step 3. The opposition VFO vs. VCFDT contributes to 86% of $q1$; FO vs. CFDT to 91% of $q2$.

Step 4. The important modalities of $q1$ and $q2$ induce a subcloud of 19 FO-affiliated voters (patterns 33xx) and a subcloud of 47 CFDT affiliated voters (22xx). Figure 5 shows the simultaneous representation. Again, the separation between the two subclouds is perfect.

Axis 4 is union dominated and opposes CFDT-affiliated voters to FO ones.

4.7 Synopsis

The synopsis is shown in Table 8.

4.8 Plane 1-2

The interpretation of axes 1 and 2 leads to allocating the relevant modalities for those axes to three classes corresponding to three polar areas: A (communist left CGT

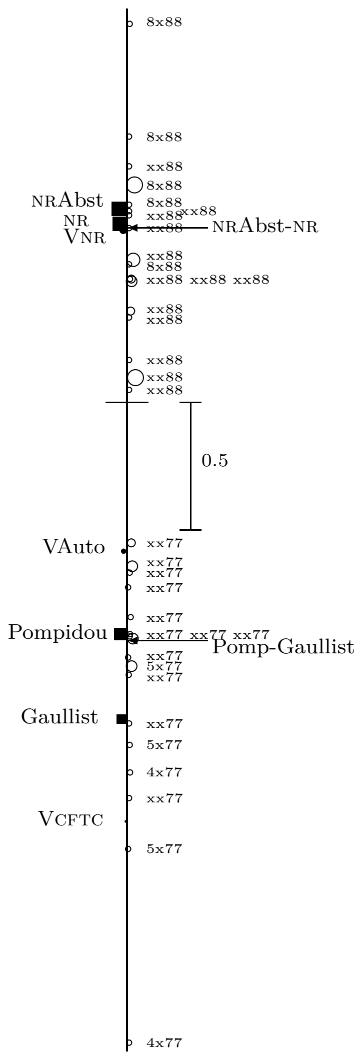

Figure 4: Simultaneous representation on axis 3 with patterns Pompidou-Gaullist (xx77) and NRAbst-NR (xx88).

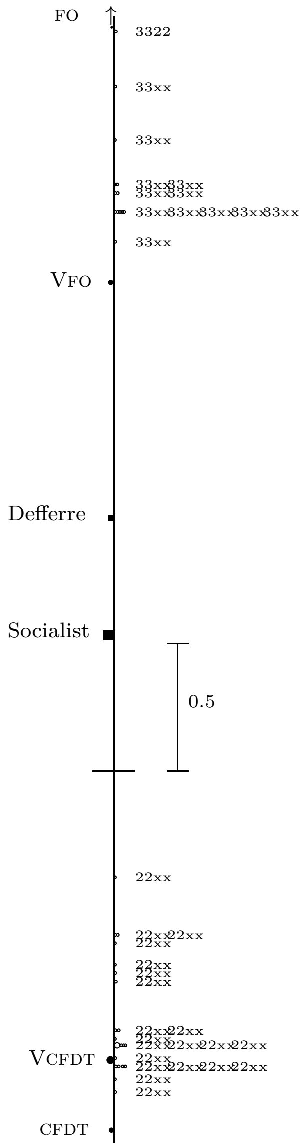

Figure 5: Simultaneous representation on axis 4 with patterns FO-VFO (33xx) and CFDT-VCFDT (22xx).

Table 16.1: Synopsis

	Axis 1: $\lambda_1 = 0.611$	**Axis 2:** $\lambda_2 = 0.491$	**Axis 3:** $\lambda_3 = 0.416$	**Axis 4:** $\lambda_4 = 0.373$
$q1$	24% of axis VCGT vs. VAuto-VAbst: 68% of question	30% of axis VCFDT vs. VAbst-VNR: 92% of question	[19% of axis]	43% of axis VCFDT vs. VFO: 86% of question
$q2$	22% of axis CGT vs. 7 NotAff: 92% of question	29% of axis CFDT *vs* NotAff: 92% of question	[12% of axis]	43% of axis CFDT vs. FO: 91% of question
$q3$	26% of axis Duclos vs. Pompidou: 89% of question	21% of axis Defferre vs. Poher *vs* NRAbst 74% of question	36% of axis Pompidou vs. NRAbst: 90% of question	[6% of axis]
$q4$	28% of axis Commun. vs. Gaullist 70% of question	20% of axis Socialist-Center vs. Gaullist-NR 97% of question	34% of axis Gaullist vs. NR: 89% of question	[7% of axis]

All comparisons are oppositions (1 d.l.) except the ternary comparison (2 d.l.) for question $q3$ and axis 2.

affiliated), B (Gaullist together with nonaffiliated and NR), and C (noncommunist left CFDT affiliated).

The modalities in Table 9 lead to defining the $1 + 12 + 4 = 17$ following response patterns (with their frequency counts, total 269): 1111 (81); 5777 (14); 5778 (4); 5787 (3); 5788 (9); 6777 (19); 6778 (5); 6787 (4); 6788 (50); 8777 (12); 8778 (9); 8787 (0); 8788 (37); 2222 (7); 2224 (1); 2252 (6); 2254 (8).

Figure 6 shows the simultaneous representation on plane 1-2, with the relevant modalities and patterns used as landmarks. This figure shows all 1049 individuals in their 319 unique positions.

Table 9: Relevant modalities for plane 1-2

	Prof. Vote	**Union Aff.**	**Pres. Vote**	**Polit. Symp.**
A	VCGT	CGT	Duclos	Communist
B	Auto Abst NR	NotAff	Pompidou NRAbst	Gaullist NR
C	VCFDT	CFDT	Defferre Poher	Socialist Center

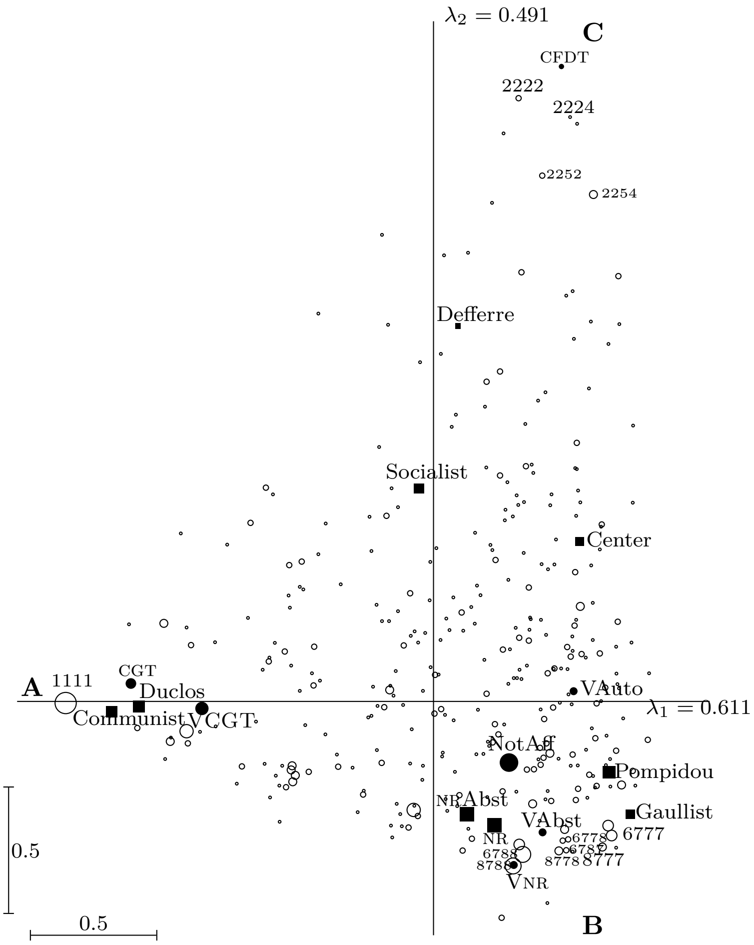

Figure 6: Simultaneous representation in plane 1-2: cloud of 319 weighted patterns (1049 individuals) and 16 relevant modalities.

5 From Interpretation to Exploration

Henceforth we place ourselves in the cloud of individuals. Considering this cloud opens new opportunities for interpretation—to begin with, the possibility of representing any patterns of interest, for instance, those that contribute most to an axis or typical patterns chosen by the specialist as landmarks to enhance interpretation. Further, the interpretation of axes may be prolonged by the exploration of the cloud and enlarged to planes or higher order spaces, always making use of the structures of the questionnaire. In this section, we will suggest—without trying to be systematic—some lines for cloud exploration. Exploration will often be motivated by specific interrogations (i.e., pertaining to parts or to groupings of data), which may be raised either before gathering data or when examining results.

5.1 Composite Modalities

The cloud of individuals enables one to go farther than the cloud of modalities, because individuals carry all the information of the data (see Chapters 15 and 20). In particular, the concept of the subcloud associated with a modality also applies to a *composite modality* (also called an "interactively coded modality"). That is, with each observed pair of modalities (k, k') (with k belonging to question q and k' to question q') is associated the subcloud of the individuals who have chosen both k and k'. The derived cloud of mean points now corresponds to the composite modalities of questions q and q'.

For example, the two political questions $q3$ et $q4$ induce 51 subclouds (among $8^2 = 64$ possible subclouds). The derived cloud of 51 mean points contributes to 81% of axis 3. Now consider the deviation between the mean points of the two composite modalities: Pompidou-Gaullist vs. political nonrespondents (see Figure 4). It is found that this deviation contributes 73% to the variance of this cloud (Rouanet and Le Roux, 1993, p. 295). This result reinforces and refines the interpretation of axis 3.

5.2 Correlation Ratios and Supplementary Questions

Every question of a questionnaire generates a partition of individuals, with a cloud of modality mean points, whose variance defines the interclass (between class) variance of the question. For each axis, dividing the interclass variance by the total variance yields a ratio denoted by η^2, which expresses the correlation between the question and the numerical variable of principal coordinates of individuals on the axis. The η^2 ratios can be calculated for active questions, as well as for *supplementary questions*.

For example, for the supplementary question "personal political situation" with five modalities after recoding—left communist (175), left noncommunist (237), center (254), right (154) and NR (229)—the graphs of Figure 7 show, in plane 1-2, the derived cloud of the five mean points (Figure 7a) and the five subclouds (Figures 7b through 7f).

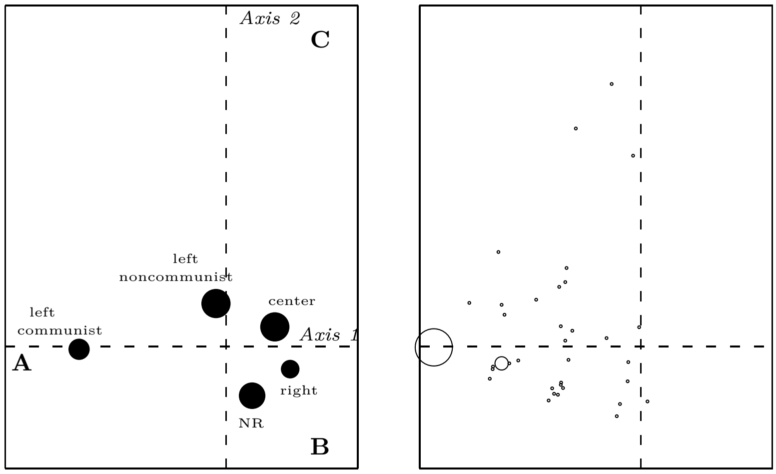

Figure 7: Personal political situation (plane 1-2). (a) five mean points; (b) left communist (175).

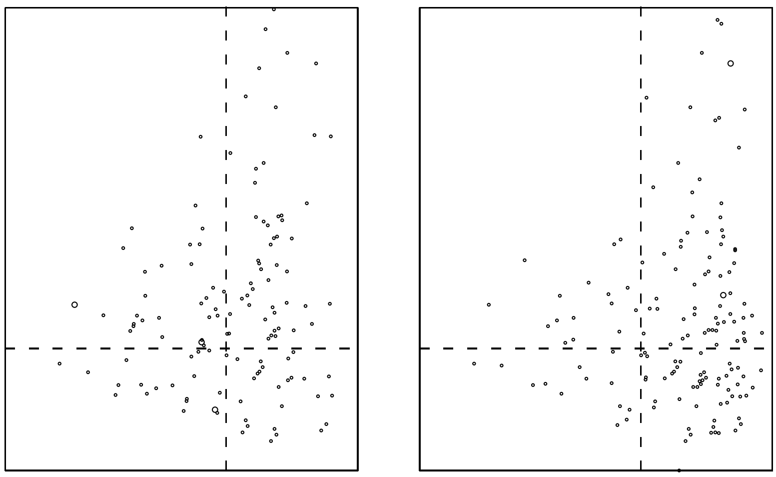

Figure 7: Personal political situation (plane 1-2). (c) left noncommunist (237); (d) center (254).

Along axis 1, the interclass variance of this supplementary question is found to be 0.347. Dividing by $\lambda_1 = 0.6113$ yields $\eta^2 = 0.57$. Then calculating the contribution of the deviation between the mean points left communist vs. right and center yields the value 0.332; that is, this opposition accounts for $0.332/0.347 = 96\%$ of the correlation ratio η^2 between axis 1 and this question.

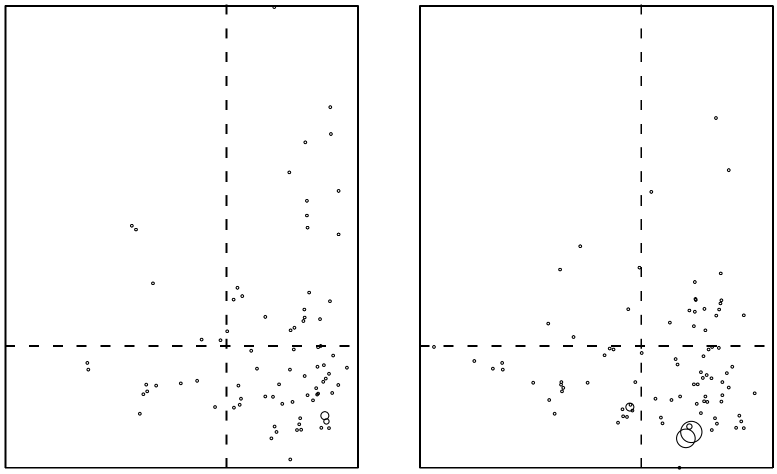

Figure 7: Personal political situation (plane 1-2). (e) right (154); (f) NR (229).

The exploratory process may extend beyond mean points. For instance, a look at the five subclouds reveals striking disparities among dispersions in plane 1-2. The most concentrated subcloud is left communist, whose variance (in plane 1-2) is equal to 0.395; the most scattered subcloud is left noncommunist, whose variance is equal to 1.018.

5.3 Crossing Relationship and Interaction

When all pairs of modalities of two questions (whether active or supplementary) are observed, it may be said (adopting ANOVA language) that there is a "crossing relationship" between the questions. Then the concept of *interaction between questions* may be formally defined as in ANOVA with unbalanced designs (Bernard *et al.*, 1989; Le Roux, 1991; Le Roux & Rouanet, 1984).

For example, let F denote the question "personal political situation" and L denote the question "trust toward unions", with three modalities high, moderate, and low or none or NR. For any axis, a diagram akin to the interaction diagrams familiar in experimental data analysis can be constructed. Figure 8 shows the interaction diagram for axis 1. Abscissas correspond to the five modalities of question F. Ordinates are the coordinates along axis 1 of the $3 \times 5 = 15$ mean points corresponding to the crossing of questions F and L. For each modality of L, the points of the five modalities of F have been joined. The three lines appear to be nearly parallel, which means that there is virtually *no interaction* between the two questions F and L with respect to axis 1. The η^2 ratio associated with the crossing $F \times L$ for axis 1 is equal to 0.64; calculation shows that the interaction accounts for only 1% of η^2.

In plane 1-2, the visualization of interaction—or the weakness of interaction, for that matter—can be performed similarly by constructing the modality mean points

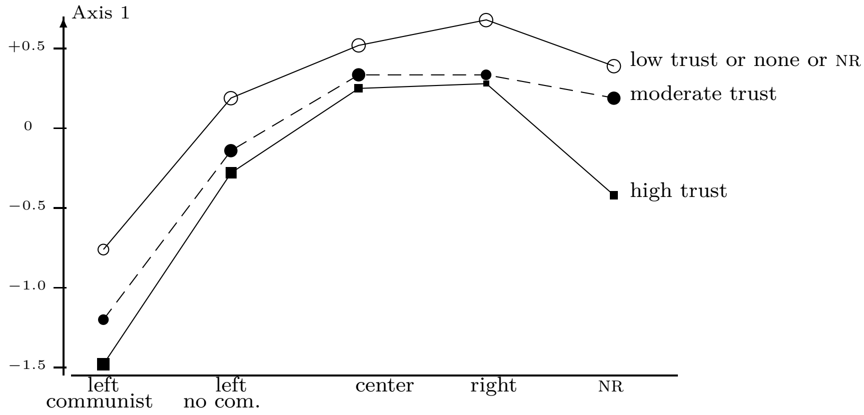

Figure 8: Interaction for axis 1.

corresponding to the crossing and joining the points corresponding to one of the questions. In Figure 9, the points of the five modalities of question F have again been joined. The quasi-parallelism of the three lines now means that there is virtually no interaction between the two questions with respect to the plane. The η^2 ratio associated with the crossing for plane 1-2 is equal to 0.45, and calculation shows that the interaction accounts for only 1% of this η^2.

6 Concluding Comments

After presenting this guide for interpretation of axes in MCA, several points are worth stressing, all directly bearing on the topic of the visualization of data.

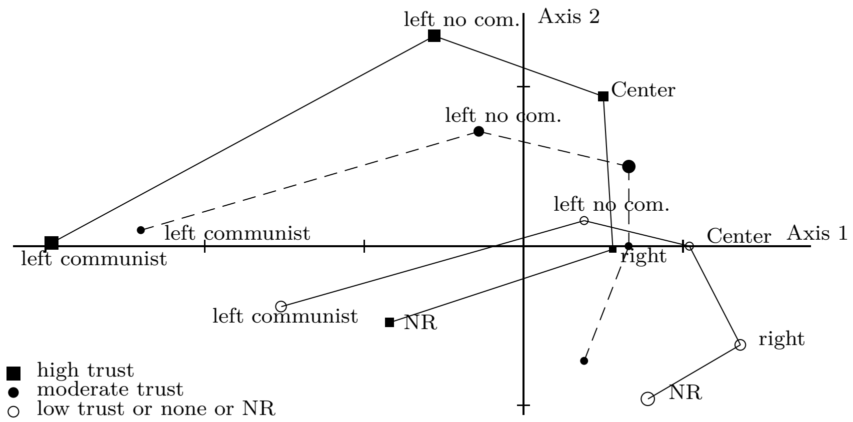

Figure 9: Interaction in plane 1-2.

1. The method of the contributions of points and deviations, developed in this chapter for MCA, readily applies, with appropriate modifications, to the interpretation of principal axes of all kinds of structured multidimensional data (Le Roux & Rouanet, 1984).
2. The interpretation of axes of higher order may reveal important findings.
3. Simultaneous representation in CA has been recognized as a most powerful visualization tool to sustain interpretations; see Benzécri (1969, 1973a, especially pp. 330–331 and pp. 468–469). This is all the more important in the case of MCA, where simultaneous representation brings together two radically different entities, namely individuals and modalities—or in other terms, objects and descriptors of objects.
4. In MCA, investigating the cloud of individuals, together with its subclouds and derived clouds (modality mean points), leads to detailed interpretations, in the first place by the examination of composite modalities.
5. A general claim underlying this chapter is that the use of specific comparisons, a tool borrowed from ANOVA, should considerably enrich the usual aids to interpretation in geometric data analysis. The method of the contributions of points and deviations provides a first step in this direction. Another step would be the investigation of the interactions between questions, a topic we have just touched upon in this chapter.

References

Adam, G., Bon, F., Capdevielle, J., and Mouriaux, R. (1970). *L'ouvrier français en 1970.* Paris: FNSP.

Benzécri, J.-P. (1969). Statistical analysis as a tool to make patterns emerge from data. In Watanabe, S., editor, *Methodologies of pattern recognition*, pages 35–74. New York: Academic Press.

Benzécri, J.-P. (1992). *Correspondence Analysis Handbook.* Dekker: New York. (adapted from J.-P. & F. Benzécri, Paris: Dunod, 1984).

Benzécri, J.-P. & *coll.* (1973). *L'Analyse des Données. 1 Taxinomie, 2 L'analyse des correspondances.* Paris: Dunod.

Bernard, J.-M., Le Roux, B., Rouanet, H., and Schiltz, M.-A. (1989). L'analyse des données multidimensionnelles par le langage d'interrogation des données (lid) au delà de l'analyse des correspondances. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 23(1):3–46.

Bonnet, P., Le Roux, B., and Lemaine, G. (1996). Analyse géométrique des données: une enquête sur le racisme. *Mathématiques et sciences Humaines*, 136:5–24.

Borjesson, M., Broady, D., Le Roux, B., Lidegran, I., and Palme, M. (2016). Cultural capital in the elite subfield of swedish higher education. *Poetics*, 56:15–34.

Fisher, R., A. (1940). The precision of discriminant functions. *Ann. Eugen.*, 10:422–429.

Gounelle, C. and Le Roux, B. (2007). Etude de la structure d'un test de jugement pratique par l'analyse géométrique des donnés. *Revue européenne de psychologie appliquée*, 57:107–1117.

Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis.* London: Academic Press.

Greenacre, M. and Blasius, J., editors (1994). *Correspondence Analysis in the Social Sciences.* London: Academic Press.

Hjellbrekke, J., Le Roux, B., Korsnes, O., Lebaron, F., Rosenlund, L., and Rouanet, H. (2007). The Norwegian Field of Power Anno 2000. *European Societies*, 9(2):245–273.

Le Roux, B. (1991). Sur la construction d'un protocole additif de référence. *Mathématiques et sciences humaines*, 114:57–62.

Le Roux, B. (2014). *Analyse géométrique des données multidimensionnelles.* Paris: Dunod.

Le Roux, B. and Rouanet, H. (1984). L'analyse multidimensionnelle des données structurées. *Mathématiques et sciences humaines*, 85:5–18.

Le Roux, B. and Rouanet, H. (2004). *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis (Foreword by P. Suppes).* Dordrecht: Kluwer.

Le Roux, B. and Rouanet, H. (2010). *Multiple Correspondence Analysis, 163.* QASS. Thousand Oaks (CA): Sage Publications.

Le Roux, B. and Rouanet, H. (2013). Geometric data analysis of gifted students' individual differences. In Suppes, P., editor, *Differences in Online Computer-based Learning: Gifted and Other Diverse Populations*, chapter 3. Stanford, CA: CSLI Publications.

Le Roux, B., Rouanet, H., Savage, M., and Warde, A. (2008). Class and cultural division in the UK. *Sociology*, 42:1049–1071.

Lebaron, F. and Le Roux, B., editors (2015). *La méthodologie de Pierre Bourdieu en action: espace culturel, espace social et analyse des données.* Paris: Dunod.

Lebart, L., Morineau, A., and Piron, M. (1995/2006). *Statistique exploratoire multidimensionnelle.* Paris: Dunod.

Rouanet, H. and Le Roux, B. (1993). *Analyse des données multidimensionnelles.* Paris: Dunod.

Tenenhaus, M. and Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, 50(1):91–119.

Appendix

Computation of the Contribution of a "Ternary" Comparison

In this appendix, we explain how to compute the absolute contribution of the "ternary" comparison between the three modalities of question $q3$ that contribute the most to axis 2, namely Defferre, Poher and NRAbst (see p. 209).

The weights, coordinates and absolute contributions (Cta) of modalities of the question *Presidential election* (***q3***) on axis 2 are given in the Table 10.

Table 10: Axis 2; weights, coordinates and absolute contributions (Cta) of modalities of question $q3$.

q3	**Weight**	**Coord.**	**Cta**
1. Duclos	.0555	−0.024	.0000
2. Defferre	.0117	1.767	.0365⋆
3. Krivine	.0024	1.677	.0067
4. Rocard	.0072	1.301	.0121
5. Poher	.0355	0.742	.0195⋆
6. Ducatel	.0017	1.016	.0017
7. Pompidou	.0584	−0.333	.0065
8. nrAbst	.0777	−0.530	.0219⋆
	.2500		.1049

The absolute contribution of question ***q3***—sum of the absolute contributions of its modalities—is equal to .1049. The sum of the contributions of the three modalities Defferre, Poher and NRAbst is equal to $.0365 + .0195 + .0219 = 0.0779$, hence together they contribute $0.0779/0.1049 = 74\%$ to question ***q3***.

The weight of the mean point of the three modalities is equal to $.0117 + 0.0355 + .0777 = 0.1249$ and the coordinate on axis 2 is equal to $(.0117 \times 1.767 + 0.0355 \times 0.742 - 0.0777 \times 0.530)/0.1249 = 0.0464$.

The contribution of the "ternary" comparison (or intra-contibution) between the three points is equal to the sum of the contributions of the three points minus the contribution of their mean point, that is, $0.0779 - 0.1249 \times 0.0464^2 = 0.0776$. Hence, it accounts for $0.0776/.1049 = 74\%$.

Remark: by definition (see p. 203), this intra-contribution (Cti) is equal to the weighted sum of the squared distances from points to their mean point.