le cnam

Paul Lazarsfeld and latent classes: some history

Gilbert Saporta

gilbert.saporta@cnam.fr
http://cedric.cnam.fr/~saporta

This talk is about
# Old Data, Old Paradigms

1.Introduction to latent variable models

- Latent class analysis (LCA) may be considered as a clustering model similar in its spirit to factor analysis when compared to PCA.

- LCA is a particular mixture model

- We will also focus on the personality of its inventor Paul Lazarsfeld (1901-1976) considered by many as the father of mathematical thinking in sociology

- A common idea in social sciences is that many phenomenons or concepts are not directly observable like intelligence, conservatism, business confidence, quality of life etc.

- Distinction between observed or **manifest** variables (eg survey data) and unobservable or **latent** variables

- Latent variables models suppose that manifest variables may be explained by a **few** latent ones: hence a goal of reducing dimensionality.

- LCA provides a **single** categorical latent variable, like in non hierarchical cluster analysis

Factor analysis

- Charles Spearman 1904
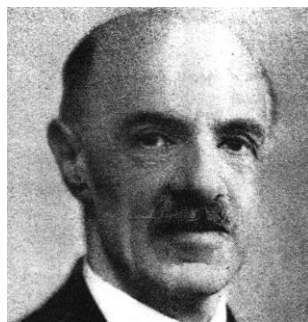
- Louis Leon Thurstone 1931

$$x_i = \lambda_i f + u_i \quad i = 1, \ldots p$$

$$x_i = \sum_{j=1}^{k} \lambda_{ij} f_j + u_i \quad i = 1, \ldots p$$

- If each **manifest variable** is caused by latent variables, covariations between manifest variables are explained by the relationship of each one with the **latent variables**

- **Fundamental hypothesis** : manifest variables are independent conditionally to latent variables

The four models

Manifest variables	Latent variables	
	categorical	numerical
categorical	Latent classes	Latent traits
numerical	Latent profiles	Factor analysis

Latent traits : Item Response Theory
particular case: Rasch models

Latent profiles : specific mixture models

- Latent Class Analysis (LCA) is a model for finding groups of cases (latent classes) from multivariate categorical data

- The latent variable is categorical and unique

  – 2 latent variables with p and q categories are equivalent to a unique one with pq categories

- LCA is a clustering technique such that within each latent class, each variable is statistically independent of every other variable.

- A particular mixture model

Comparison with exploratory techniques

	Latent variables	
Manifest variables	**categorical**	**numerical**
categorical	**Latent classes**	Latent traits
numerical	Latent profiles	Factor analysis

models

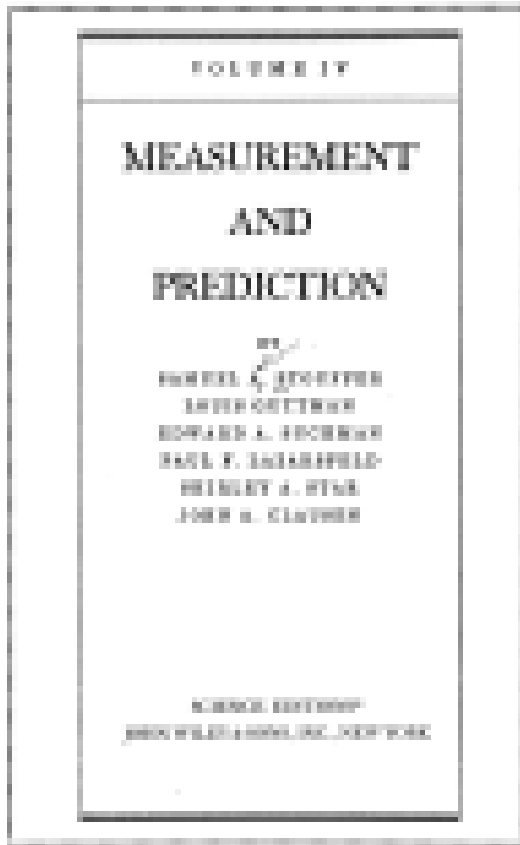|	Components or summary variables	
Observed Variables	**categorical**	**numerical**
categorical	**clustering**	MCA
numerical		PCA

« Data Analysis »

PCA versus Factor Analysis

- exploration
- unique solution
- explains variance
- units oriented

- modelisation
- indetermination
- explains correlations
- variables oriented

However frequently close results if specific variances are near equal , or if the p-k last eigenvalues are close to 0

2. Latent class analysis

- **Latent classes and latent traits** :
  - "Latent class analysis was developed mainly within the social and political sciences, whereas latent trait models have a clear psychometrical background." (Heinen, 1996, ix)
  - Both method developped by P.Lazarsfeld from the 40's till the 60's under the common name of latent structures.
  - Context: psycho-social studies on soldiers during 2nd world war funded by the american war department .

VOLUME IV

MEASUREMENT AND PREDICTION

BY

SAMUEL A. STOUFFER
LOUIS GUTTMAN
EDWARD A. SUCHMAN
PAUL F. LAZARSFELD
SHIRLEY A. STAR
JOHN A. CLAUSEN

Stouffer, Samuel A.; Louis Guttman; Edward A. Suchman; Paul F. Lazarsfeld; Shirley A. Star; John A. Clausen
Published by Princeton University Press (1950)

From « trace lines » to latent classes

First example

Let us look at specific test items. Suppose in connection with studying the effect of an indoctrination film we want to test the "ethnocentrism" of American soldiers. By this we mean vaguely the extent to which they think that only Americans are of value and that the actions and ideals of the war allies are of little value or importance for the war effort. Suppose the following three items have been proposed for inclusion in the test:

1. I believe that our European allies are much superior to us in strategy and fighting morale. Yes _____ No _____

2. The majority of all equipment used by all the allies comes from American lend lease shipment. True _____ False _____

3. Neither we nor our allies could win the war if we didn't have each other's help. Agree _____ Disagree _____

All three items were included in the test because they are related and partially express ethnocentrism the way we have conceived it to begin with. This fact can be formulated more precisely in the following way: We assume that the probability y, that respondents check the first alternative in the ith question, is a function $f_i(x)$ of their position on the continuum x, their degree of ethnocentrism. This formulation can be graphically represented in Figure 1. The graphical picture of our functions $f_i(x)$ we shall call the *trace line* of item i.
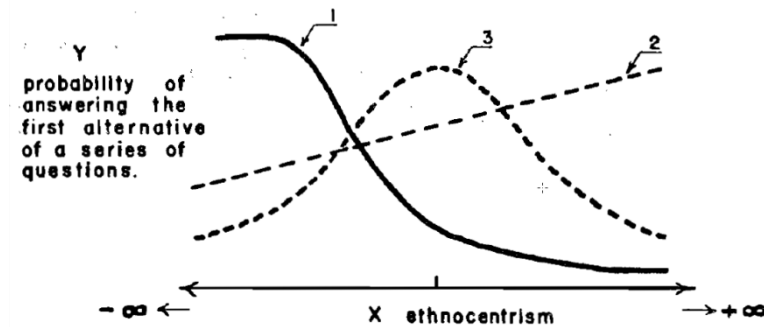

Figure 1. Trace lines of 3 items in a test of "ethnocentrism."

Lazarsfeld, Paul F. [1950a], The Logical and Mathematical Foundations of Latent Structure Analysis, dans Stouffer, S.A., et al.
chapter 10, 362-412
Measurement and Prediction, Volume IV of The American Soldier : Studies in Social Psychology in World War II.
Princeton University Press.

From « trace lines » to latent classes

Second example

A 2 classes model

Computed Latent Structure for Attitude toward the Army

Latent Class Frequencies	Item 1 Probability +	Item 2 Probability +	Item 3 Probability +	Item 4 Probability +
424.3	.9240	.6276	.5704	.5125
575.7	.4324	.1871	.1008	.0635

Manifest Data of Four Items on Attitude toward the Army

In general how do you feel the Army is run?	Do you think when you are discharged you will [have] a favorable attitude toward the Army?	In general do you feel you yourself have gotten a square deal from the Army?	Do you feel that the Army is trying its best to look out for the welfare of enlisted men?	Count
+	+	+	+	75
+	+	+	-	69
+	+	-	+	55
+	-	+	+	42
-	+	+	+	3
+	+	-	-	96
+	-	+	-	60
+	-	-	+	45
-	+	+	-	16
-	+	-	+	8
-	-	+	+	10
+	-	-	-	199
-	+	-	-	52
-	-	+	-	25
-	-	-	+	16
-	-	-	-	229

- **1950:** no goodnes of fit test, no convenient estimation technique

- **1968**: Identifiability, moments estimation

- **1974:** Leo Goodman identifies the latent class model as a log-linear model. Treating latent classes as missing data lead to new interpretations and EM estimation procedures

The binary case (may be generalized to polytomous manifest variables)

- p binary variables X_1, X_2,..,X_p with 0 1 values

- Y latent variable with k classes,

- p_{ij}=P(X_i=1/ i belongs to class j).

- π_j prior probability of belonging to latent class j

- **The conditional (local) independence hypothesis**

$$f(\mathbf{x}) = \sum_{j=1}^{k} \pi_j \prod_{i=1}^{p} p_{ij}^{x_i} (1 - p_{ij})^{1-x_i}$$

A mixture model

- Posterior probability that x belong to latent class j

$$h(j / \mathbf{x}) = \pi_j \prod_{i=1}^{p} p_{ij}^{x_i} (1 - p_{ij})^{1-x_i} / f(\mathbf{x})$$

Parameter estimation

Maximum likelihood with EM:

$$l = \sum_{h=1}^{n} \ln \left(\sum_{j=1}^{k} \pi_j \prod_{i=1}^{p} p_{ij}^{x_{ih}} (1 - p_{ij})^{1 - x_{ih}} \right)$$

Goodness of fit and model choice

- G^2 statistic

$$G^2 = 2\sum_x n(\mathrm{x})\ln(\frac{n(\mathrm{x})}{n\hat{f}(\mathrm{x})})$$

- number of free parameters :

   k – 1 class probabilities π_j and kp conditional probabilities p_{ij}

   = k(p+1) – 1 parameters.

- G^2 compared to a chi-square with

  $\nu = 2^p - k(p+1) + 1$ degrees of freedom

- Nested hypotheses with k

- However , for large p, observed frequencies of the 2^p response patterns are too small or even null.

  – Chi-square distribution not applicable.

  – Simulations, or bootstrap

- Choice of the number of classes: minimizing Akaiké AIC, or Schwartz's BIC:

$$AIC = -2\ln(L) + 2(k(p+1) - 1)$$

$$BIC = -2\ln(L) + \ln(n).(k(p+1) - 1)$$

Example (Bartholomew & Knott)

Survey about British Social Attitudes in1990, 1077 respondents, 10 binary questions on sexual attitudes

1. Should divorce be easier here?

2. Do you support the law against sexual discrimination?

3. View on pre-marital sex: not at all wrong, ..., always wrong.

4. View on extra-marital sex: not at all wrong, ..., always wrong.

5. View on sexual relationship between individuals of the same sex: not at all wrong, ..., always wrong.

6. Should gays teach in schools?

7. Should gays teach in higher education?

8. Should gays hold public positions?

9. Should a female homosexual couple be allowed to adopt children?

10. Should a male homosexual couple be allowed to adopt children?

Among $2^{10}=1024$ response patterns, only 147 were observed:

```
 1      90      0110011100
 2      11      0110011000
 3       9      0110111000
 4     117      0110000000
 5      18      0100000100
 6      93      0100000000
 7      19      0111111100
 8      35      0010000000
 9      21      0110001100
10       6      0111111110
```

Nb. of classes	AIC	BIC
2	9328	9432
3	8946	9105
4	**8850**	**9064**
5	8852	9121

$$\hat{\pi}_1 = 0.4611 \quad \hat{\pi}_2 = 0.0139 \quad \hat{\pi}_3 = 0.4169 \quad \hat{\pi}_4 = 0.1081$$

Conditional probabilities of yes response to each question

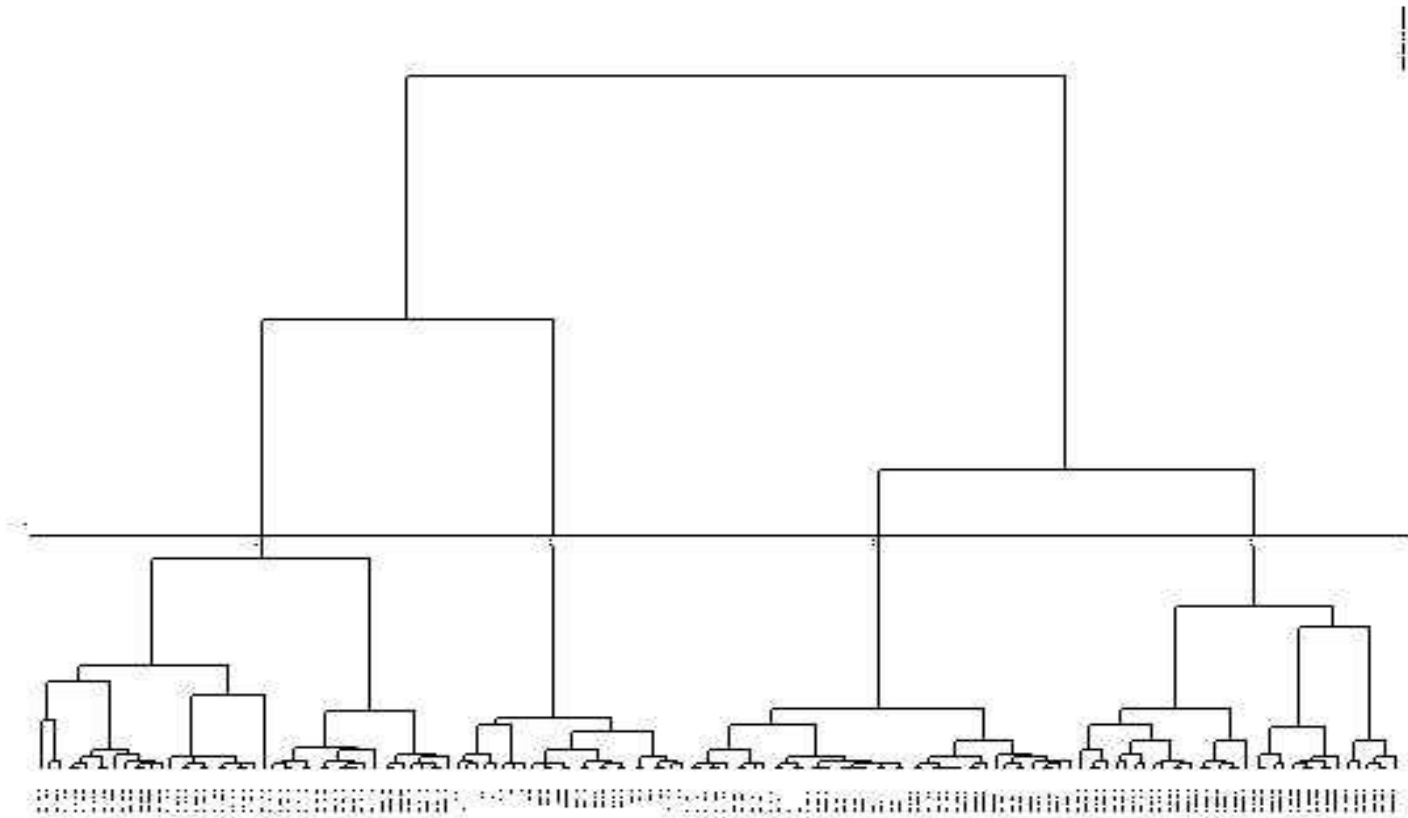|	classe 1	classe 2	classe 3	classe 4
X1	0.1360	0.0667	0.0947	0.2144
X2	0.7656	0.6000	0.8712	0.9246
X3	0.6319	0.8667	0.8620	0.9635
X4	0.0822	0.2667	0.1319	0.3089
X5	0.0681	0.6000	0.3822	0.8299
X6	0.0081	0.0000	0.8721	1.0000
X7	0.0589	0.2000	0.9829	1.0000
X8	0.2077	0.2667	0.9141	1.0000
X9	0.0463	1.0000	0.1071	0.9790
X10	0.0000	1.0000	0.0000	0.8505

Class 1 non-permissive
Class 2 whose distinguishing feature is very positive support for homosexual adoption, combined with a strong negative attitude to items 6, 7 and 8, concerning teaching and holding public positions.
Class 3 permissive on everything except adoption
Class 4 permissive on all issues

OBS.FREQ.	E(FREQ)	LAT. CLASS	RESPONSE VECTOR
90	114.608	3	0110011100
11	10.825	3	0110011000
9	6.661	3	0110111000
117	125.255	1	0110000000
18	19.173	1	0100000100
93	72.966	1	0100000000
19	10.831	3	0111111100
35	38.337	1	0010000000
21	18.862	3	0110001100
6	4.348	4	0111111110
14	16.949	3	0010011100
1	2.737	3	0111001100
2	0.315	3	0111001110
15	11.223	1	0111000000
11	9.161	1	0110100000
1	1.583	3	0010101100
3	1.549	3	0110101000
32	18.344	3	0100011100
1	0.142	1	1011000100
27	33.119	1	0110000100
8	7.973	4	0110011111
95	71.011	3	0110111100
7	3.887	3	0100001100
40	38.911	4	0110111111
2	2.252	3	0100011110
13	15.139	3	0110011110

Comparison with hierarchical clustering

Split into 4 classes followed by a k-means partitioning

	CLASSE HIER. 1	CLASSE HIER. 2	CLASSE HIER. 3	CLASSE HIER. 4	ENSEMBLE
CLASSE LAT. 1	3	0	67	421	491
CLASSE LAT. 2	0	14	0	1	15
CLASSE LAT. 3	419	0	42	0	461
CLASSE LAT. 4	6	99	5	0	110
ENSEMBLE	428	113	114	422	1077

3. Paul Felix Lazarsfeld 1901-1976

- «One of the four or five sociologists of his generation who will have most strongly marked the history of their discipline» (Encyclopedia Universalis, R.Boudon)

- Born in Vienna 1901. Studies in law, economics and mathematics at University of Vienna.

- In 1922-23, Lazarsfeld spent a year studying at the Sorbonne, during which he was enrolled in the SFIO section of the 5th arrondissement of Paris, and from where he returned francophile enthusiastic, although very insightful to France and the French, for the rest of his life. *
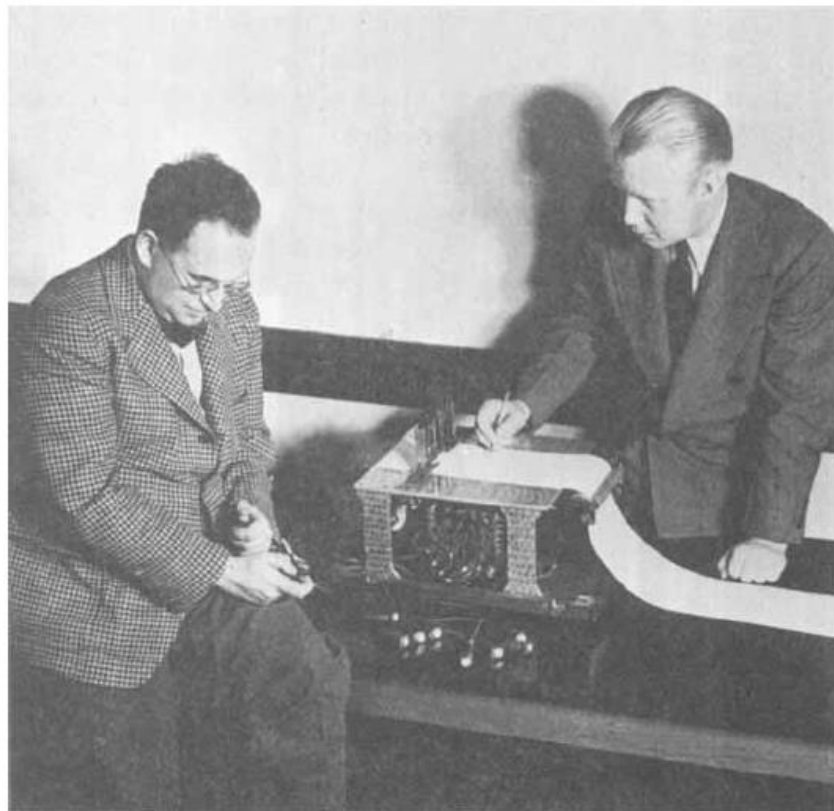
* Translated from P.L. Lécuyer, Mathématiques et sciences humaines, 157, 2002, p. 49-104

- Ph.D. in applied mathematics 1924 : «Über die Berechnung der Perihelbewegung des Merkur aus der Einsteinischen Gravitationstheorie» *ie* an application of Einstein's theory of gravitation to the movement of the planet Mercury.

- Goes to psycho-social surveys. 1932: «Marienthal: The Sociography of an Unemployed Community » (french edition 1981, preface written by Pierre Bourdieu).

- 1933 Invited in the USA with a Rockefeller scholarship. When his scholarship expires, he returns to Europe

- In 1934, the Socialist Party of Austria was declared illegal. Most of his family members are arrested. He loses his job. Decides to use the remaining dollars on the purchase of a third-class ticket for the United States. He settled and became «**a marxist on leave**»

- Established in Newark a contract-based institute: consumption studies for the dairy industry, Cosmopolitan, etc. To make believe that his institute had a larger staff, Lazarsfeld also published under a pseudonym

- Professor of sociology for nearly 30 years at Columbia University (1940-1969), He was appointed in 1962 as the first « *Quetelet Professor of Social Science* » a chair that has been created for him. He considered Adolphe Quetelet as the real founder of sociology, rather than Auguste Comte.

- After his retirement from Columbia he continued to teach at Pittsburgh University until his death.

Opinion and mass media

- Thanks to a new funding of the Rockefeller Foundation he created in 1937 at Princeton University the *Office of Radio Research* in order to look into the effects of mass media on society

- With Frank Stanton, director of research at CBS (then CEO 1946-1971), he designed in 1938 the first program analyzer measuring satisfaction with push buttons the «**Lazarsfeld-Stanton Program Analyzer**» also known as *Little Annie*

- Transferred to the University of Columbia with Lazarsfeld, the ORR became in 1944 the *Bureau of Applied Social Research,* further remaned *Center for the Social Sciences,* then *Paul F. Lazarsfeld Center for the Social Sciences* after his death in 1976

Paul Lazarsfeld and Frank Stanton with an early model of their program analyzer, circa 1938

Chart from an analysis of program analyzer problems, CBS, 1942
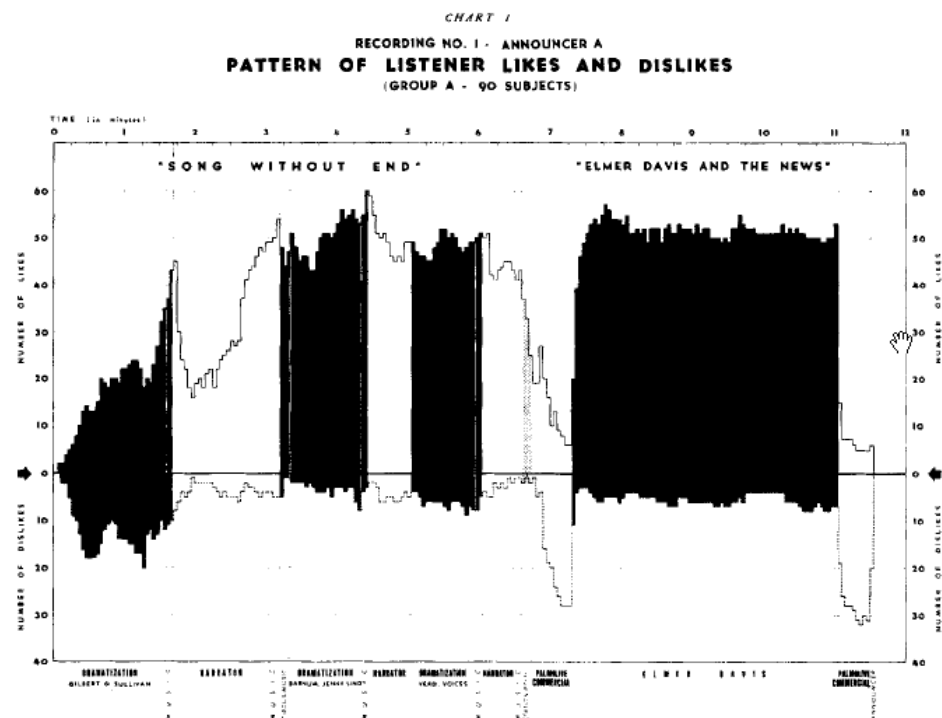
Contributions to opinion surveys

– Paul F. Lazarsfeld studied the determinants of voting and developed for his first major election study, during the 1940 presidential elections (F.D.Roosevelt vs Willkie), an method (called «**panel**») conducting repeated interviews with a representative sample of voters throughout the election campaign*The People's Choice* [1944]

– On the occasion of the 1948 elections (Truman vs Dewey), Paul F. Lazarsfeld proposed using Markov chains to model changes of opinion in a panel. He entrusted the resolution to T.W.Anderson

- Founder of 4 research institutes, true scientific entrepreneur and propagandist of the empirical approach in sociology, P.F. Lazarsfeld considered that he had only 4 original ideas in his life including **panel analysis** and **latent structures**

Latent classes and social classes

American sociologists have been fascinated from the very beginning with the idea of social class. Whether social status is enabled by membership in a class of persons or should be conceptualized as a continuous variable has been hotly, warmly and coolly debated throughout this century. Lazarsfeld's latent class analysis was taken up by the class proponents as a method that would allow the scientific assessment of a person's class membership. (Paul Henry, Latent Structure Analysis at Fifty, 1999)

P.F. Lazarsfeld and french sociologists

- Long collaboration with Raymond Boudon

- Honorary doctorate or « honorary » or «emeritus» professor * (a title then unregulated) from Université René Descartes in 1972 on a proposal from Jean Stoetzel (professor of sociology and founder of IFOP)

- Complex relationships with Pierre Bourdieu…

* Contradictory sources

PAUL LAZARSFELD
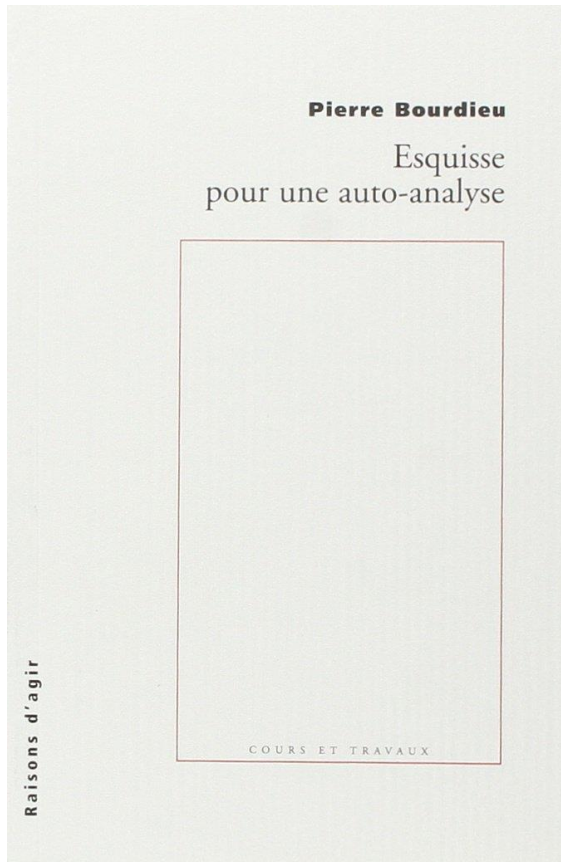MARIE JAHODA - HANS ZEISEL

LES CHÔMEURS DE MARIENTHAL

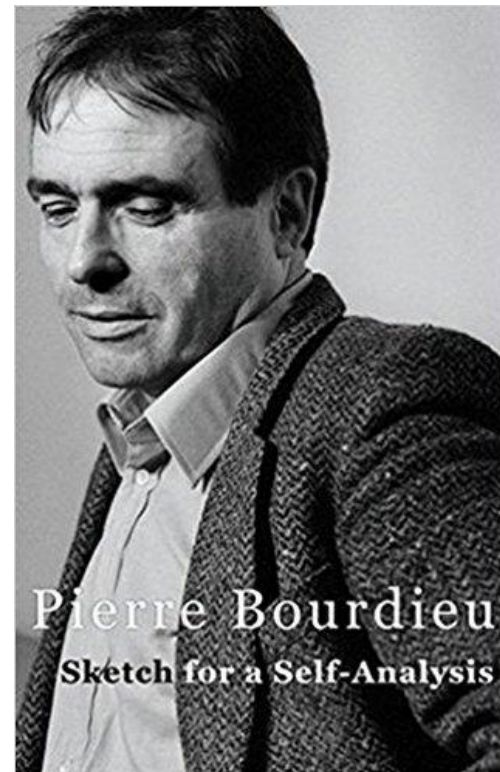Traduit de l'allemand par Françoise Laroche

PRÉFACE DE PIERRE BOURDIEU

1981

☆m

LES ÉDITIONS DE MINUIT

Pierre Bourdieu

Esquisse
pour une auto-analyse

Raisons d'agir

COURS ET TRAVAUX

Pierre Bourdieu
Sketch for a Self-Analysis

2004   Polity Press 2008

If, in the early 1960s, despite weekly reminders from the authorized representatives of the master in a missionary land, I had stubbornly refused to attend the lectures that Paul Lazarsfeld gave at the Sorbonne, before the assembled world of French sociology, it was because it had all seemed to me more like a collective ceremony of submission than a simple technical enterprise of scientific training and updating.

Page 73

The history of my (at first sight hopeless) confrontation with Paul Lazarsfeld, who at that time exercised a social and scientific hegemony over world sociology that is difficult to imagine, found, for me, something like a satisfying conclusion, one day in the late 1960s, when, being then at the height of his renown, he had literally 'summoned' Alain Darbel and me to the Hotel des Ambassadeurs, where he habitually stayed when visiting on behalf of the Ford Foundation, to tell us his criticisms of the mathematical model of gallery visiting that we had published in *L'Amour de lart.* He arrived with a copy of the book on which he had scrawled in blue ink and, with a big cigar in his mouth, he pointed out, with some brutality, what he regarded as unforgivable errors. …

When these corrections were granted, Lazarsfeld declared with some solemnity that 'nothing so good had ever been done in the United States'. But he took care never to put it in writing, and continued to give his spiritual investiture to Raymond Boudon, the *accredited agent* *of his scientific multinational.

Pages 74-75

* « *chef de comptoir* » in french edition

michael pollak

PAUL F. LAZARSFELD FONDATEUR D'UNE MULTINATIONALE SCIENTIFIQUE*

Actes de la recherche en sciences sociales. 25, 45-59, 1979

Il n'est pas facile, à première vue, de comprendre que Paul F. Lazarsfeld, symbole d'un empirisme a-politique, pour ne pas dire «antipolitique», ait pu avoir, au début de sa carrière, à la fin des années 20, le projet de construire une psychologie sociale compatible avec le marxisme, et qu'il ait été aussi un militant très actif dans les organisations de jeunesse socialistes en Autriche. La reconstruction de sa biographie intellectuelle et politique, qui est présentée ici, vise à rendre compte de la continuité qui se maintient à travers les ruptures associées à l'immigration, en même temps que des discontinuités évidentes de son projet intellectuel.

L'engagement politique

Lazarsfeld n'a jamais nié l'influence que son activité militante au sein du parti social-démocrate a eue sur sa décision de s'engager dans la recherche sociale, mais rares sont les témoignages d'après-guerre qui révèlent ce passé de militant et d'actif responsable politique (1). Le père de Paul F. Lazarsfeld, un avocat qui militait au parti social-démocrate, donnait des consultations juridiques gratuites aux militants inculpés de délits politiques. Il était en personnel de presque tous les théoriciens de l'austromarxisme du début du siècle. C'est ainsi qu'en 1916 les enfants Lazarsfeld passèrent leurs vacances dans un hôpital dirigé par Rudolf Hilferding (2). La mère de Lazarsfeld, Sophie, réussit à faire de son foyer un centre de rencontre d'intellectuels proches du parti, qu'ils fussent ou non militants. A côté de la revue *Der Kampf* qui regroupait les plus importants théoriciens «austro-marxistes» (Otto Bauer, Karl Renner, Rudolf Hilferding, Max Adler), ce salon était un des plus importants endroits de rencontre d'intellectuels

socialistes. L'engagement socialiste n'était pas la seule raison de cohésion de ce groupe d'intellectuels. Il faut mentionner l'origine juive de la plupart d'entre eux. L'association entre les intellectuels juifs et la social-démocratie en Autriche résulte pour une forte part de l'évolution spécifique du mouvement libéral autrichien au 19ème siècle (3).

Quand, en 1867, le parti libéral accéda pour la première fois au pouvoir, les espoirs réformistes les plus divers se concentraient sur son action en faveur d'un régime constitutionnel dans la monarchie austro-hongroise : limiter le pouvoir impérial, réduire l'influence de l'Église catholique, établir

*Ce travail n'aurait pas pu être mené sans les entretiens que des amis et collègues de Paul Lazarsfeld m'ont accordés. Je tiens à remercier tout particulièrement Freda Meissner Blau et Paul Blau qui m'ont aidé à connaître le climat intellectuel autrichien et le milieu dans lequel Lazarsfeld passa sa jeunesse, et Rose K. Goldsen, collaboratrice de Lazarsfeld pendant de longues années à l'université de Newark et au Bureau of Applied Social Research, actuellement professeur de sociologie à l'université de Cornell. Je tiens à remercier V. Karady et D. Merllié de leurs critiques et de la relecture minutieuse qu'ils ont faite de ce texte. G. Stourzh a eu la gentillesse de me communiquer la liste non publiée des intellectuels autrichiens émigrés aux États-Unis.

1.—A l'exception de la préface au livre d'Yvon Bourdet (Y. Bourdet éd., *Otto Bauer et la Révolution*, Paris, EDI, 1968, pp. 7-8) où Lazarsfeld rapporte les liens officieux qu'il entretenait avec la 5ème section de la SFIO et avec Léo Lagrange lors de son séjour d'études à Paris en 1922-1923.

2.—Il s'agit du médecin et économiste, auteur du livre *Finanzkapital*, théoricien austromarxiste qui avait souvent publié dans *Der Kampf* avant de s'installer en Allemagne où il était un des ministres des finances de la République de Weimar. Pour cette période de la vie de Lazarsfeld, voir J. Maitron, G. Haupt, éd., *Dictionnaire biographique du mouvement ouvrier, I, Autriche*, Paris, Les éd. ouvrières, 1971, pp. 182-184. (Ces informations furent complétées par des entretiens avec des amis autrichiens de Lazarsfeld).

3.—Pour l'histoire des juifs d'Autriche et de la relation entre juifs et autrichiens, voir : Institut für Judaistik der Universität Wien éd., *Das österreichische Judentum, Voraussetzungen und Geschichte*, Wien-München, Jugend und Volk, 1974 ; H. Gold, *Geschichte der Juden in Wien*, Tel Aviv, 1966.

CONCLUSION

- LCA deserves to be better known from statisticians

- Provides an interesting model for cluster analysis and gives a solution to the choice of the number of classes

- LCA does not have the main drawbacks of Factor Analysis (factor indeterminacy)

- Do latent variables exist or not is an other issue…

?

A few software and websites

- XLSTAT-LatentClass https://www.xlstat.com/fr/solutions/lg
- SAS Proc LCA : http://methodology.psu.edu/downloads/proclcalta
- R package poLCA

- "Latent Structure Analysis at Fifty" by Neil W. Henry (1999) http://www.people.vcu.edu/~nhenry/LSA50.htm
- " Latent Structure Analysis Website ", by John Uebersax : http://www.john-uebersax.com/stat/index.htm

References

- Aitkin, M.; Francis, B.; Raynal, N.: Une étude comparative d'analyses des correspondances ou de classifications et des modèles de variables latentes ou de classes latentes , *Revue de Statistique Appliquée*, 35, 2, p. 53-81, 1987

- Bartholomew,D.J., Knott, M., Moustaki,I.: *Latent Variable Models and Factor Analysis* , 3rd Edition, J.Wiley, 2011.

- Celeux, G.: Classification et modèles. *Revue de Statistique Appliquée*, 36, 4, p. 43-57, 1988

- Droesbeke,J.J., Saporta,G., Thomas-Agnan, C. (eds.) *Modèles à variables latentes et modèles de mélange*, éditions Technip, 2013

- Everitt, B.S.: *An Introduction to Latent Variable Models* , Chapman & Hall, London, 1984.

- Hagenaars J.A., McCutcheon A.L. (eds.): *Applied latent class analysis,* Cambridge University Press, 2002.

- Collins L., Lanza S.: *Latent class and latent transition analysis*, Wiley, 2010

- Lautman, J. et Lécuyer, B.P. (éditeurs): *Paul Lazarsfeld (1901-1976), la sociologie de Vienne à New-York,* Paris, L'Harmattan,1998.

- Lazarsfeld, P.F., Henri, N.W.: *Latent Structure Analysis* , Houghton Mifflin, Boston, 1968.

- McCutcheon, A.L,: *Latent Class Analysis* , (Sage University Paper Series on Quantitative Applications in the Social Sciences n°64) , SAGE publications, 1987.