Eventually, three GDA paradigms associated with three types of data tables emerged:

— CA (with χ^2–metric): *two–way frequency, contingency tables*;

— PCA (various metrics): *Individuals×Numerical Variables* tables;

— MCA: *Individuals×Categorized Variables* tables.

Moving on from CA on frequency tables to MCA on Individuals×Variables tables means a shift in methodological emphasis. In MCA as well as in PCA, the individuals carry all the information, and it is on the cloud of individuals that interpretation and exploration deserve to be concentrated.

1.4 Historical Sketch

Predecessors and contemporaries. At this point, a historical sketch of GDA is in order. In a well thought–out text, entitled "Histoire et Préhistoire de l'*Analyse des Données*", Benzécri (1982a) recalls the great historical figures of statistics, especially K. Pearson (1901) — "Should we need an Anglo–Saxon patronage for *Analyse des Données*, we would be pleased to turn to the great Karl Pearson." — and Fisher; he situates his approach in the *psychometric tradition*, with *Factor Analysis* from Spearman to Thurstone and Burt (1950), and *scaling methods* with Hirschfeld (1935), Eckart & Young (1936) and especially Guttman (1941). He also makes due reference to related contemporary works, such as the *quantification method* developed around Hayashi (1952) in Japan, and proximity analysis — also known as *MultiDimensional Scaling* (MDS) — developed by Torgerson (1958), Shepard (1962), and others[17]. Admittedly, the list of references could be enlarged to Guttman (1959) for a synthesis of early literature, to Dempster (1969) for the formal key idea, or to Tukey (1960) for the Data Analysis Philosophy. But the conclusion that clearly emerges from this review is that, beyond its similarities with several anterior or contemporary undertakings, the geometric construction around CA was most original and brought an in–depth renewal of multivariate statistics.

As far as the *history* (strictly speaking) of GDA is concerned, it can be divided into three periods (landmark years are indicative).

First period: Emergence (1963–1973). The hard core of CA was achieved in the mid–sixties, with the six lectures given by Benzécri at the

[17] MDS (see Shepard, 1966, 1980) — with its variants like Small Space Analysis (SSA) — is a case in point; it unquestionably belongs to GDA, without being an outgrowth of CA.

Collège de France, B. Cordier–Escofier's dissertation (1964, 1969), and a host of mimeographed reports that were widely circulating. A brief account in English of the first developments can be found in Benzécri (1969). Toward 1973, the emergence of "Analyse des Données" around CA, combined with classification methods, was completed with the publication of the monumental treatise by Benzécri & Coll. (1973) in two volumes (Taxinomy and Correspondence Analysis). Meanwhile, the first statistical textbook incorporating CA had appeared: Lebart & Fénelon (1971), followed by many others: Berthier & Bouroche (1975), Cailliez & Pagès (1976), Lebart, Morineau & Tabard (1977), etc. (all of them in French).

Second period: Splendid isolation (1973–1980). The movement of "Analyse des Données" enjoyed a golden age in France. Benzécri's laboratory at the "Institut de Statistique de l'Université de Paris" (ISUP) was for many years a place of creative dialogue between statisticians and researchers. An innovative statistical tradition developed, around a body of expert knowledge: contributions, supplementary elements, Guttman effect, disjunctive coding, Burt table, etc. Procedures were implemented in software (whose sources were free). Statistical work was published mostly in *Cahiers d'Analyse des Données* and *Revue de Statistique Appliquée*. "Analyse des Données" began to be taught in graduate statistics curricula. In applications, CA (especially MCA) became a major tool for analyzing multivariate data such as questionnaires (cf. §1.5).

Throughout the first two periods, in the relevant literature in English, there were very few published reactions to the work done in France, even though some "joint display spirit" came to be floating in the air, as reflected in Gower (1966), Good (1969), Gabriel (1971). In the seventies, CA is ignored in MDS publications, such as Shepard, Romney, Nerlove (1972), Kruskal & Wish (1978), Shepard (1980); also ignored in the encyclopedic treatise by Kendall & Stuart (1976). The silence about CA was conspicuously broken by Hill (1974), who — encouraged by Gower? — launched the English phrase "Correspondence Analysis" (perhaps its first appearance in print) and emphatically announced that CA was a "neglected method".

Third period: Bounded international recognition (since 1981). International recognition eventually came in the eighties. Books in English were published that directly stemmed from the work done in France: Greenacre (1984) was specifically devoted to CA, whereas Lebart & al (1984) dealt more generally with "multivariate descriptive analysis"; then came Jambu (1991); then Benzécri (1992): the translation of the introductory book of 1984. In the meantime, Malinvaud (1980) and Deville & Malin-

vaud (1983) had discussed CA in official statistics and referred to "Data Analysis" as "Econometrics without stochastic models". In psychometry, there was the valiant paper by Tenenhaus & Young (1985), etc.

Recognition also came from other lines of work that incorporated CA into their own systems, especially *Dual Scaling*, developed by Nishisato (1980), who compiled a giant bibliography on "quantification of categorical data" (Nishisato, 1986), and *Homogeneity Analysis*, developed by the Leyden group led by de Leeuw and reflected in Gifi (1981/1990).

Most important, CA penetrated such areas as marketing research (Hoffman & Franke, 1986), where *MultiDimensional Scaling* (MDS) techniques had long been dominating. In the late eighties, the MDS group came to adopt CA as an authentic (even though "less conventional") MDS method, as reflected in Carroll & Green (1988)[18], Weller & Romney (1990), etc. In the 1984 edition of Kendall & Stuart, a casual reference to Lebart & Fénelon (1971) can be spotted in Volume 3 (p.418).

This sort of recognition continued in the nineties, with e.g. Gower & Hand (1996) presenting CA and MCA as "biplot" methods among others.

Where do we stand now? In 2003, the situation calls for a mixed assessment.

On the positive side, the phrases "Correspondence Analysis" and even "Multiple Correspondence analysis" are well rooted in English. The basic procedure of CA can be found in international statistical software. CA is definitely renowned for the visual exploration of data. It is now commonplace to discuss topics explicitly related to CA, such as stability, choice of metrics (Rao, 1995), canonical analysis (Goodman, 1991), etc. International conferences specifically oriented to "CA and related methods" are organized outside France; see e.g. Blasius and Greenacre (1998), and the recent conference organized by Greenacre (2003).

On the other side, CA still remains isolated in the field of Multivariate Statistics. In spite of increasing demand from users, popular books and international software all too often offer imperfect versions of the method. Frankly speaking, for MCA the situation is really defective. This method, which is so powerful for analyzing large–scale questionnaires, is still hardly ever discussed and therefore remains underutilized, as does most GDA expert knowledge.

[18]The concluding sentence of this paper is that INDSCAL (an MDS variant) should be used "if only as an adjunct to the more conventional MCA"; a sentence that speaks volumes for the increasing popularity of CA and MCA in marketing research.

To sum up, in the international scientific community, CA is now recognized and used, but GDA largely remains to be discovered.

1.5 Methodological Strong Points

We will now consider GDA from the *user's viewpoint*, and discuss GDA as a geometric frame model (including Euclidean classification).

In social sciences, GDA has generated a statistical practice in sharp contrast with the conventional one. In the latter, *numerical indicators* (regression coefficients, etc.) occupy center–stage, together with significance levels (the *star system*: * significant at .05 level, ** significant at .01, etc.). In GDA, *clouds of points* are central. In social sciences, this contrast reflects two distinct conceptions of the role of statistics, namely sustaining a "sociology of variables" versus *constructing a social space*.

1.5.1 GDA as a Frame Model

Any empirical study of a research field involves some theoretical framework, which, by summarizing relevant knowledge, guides the collection of data and the interpretation of results. When this framework is formalized in mathematical terms, we call it a *mathematical frame model*. For example, regression or ANOVA models are frame models that summarize the relevant knowledge in terms of variables; whereas GDA models summarize this knowledge in geometric terms. The frame model concept serves as a reminder that GDA methods, like all statistical ones, can only be fruitful if they deal with *relevant data*; performing a geometric analysis does not mean gathering disparate data and see 'what comes out' from the computer.

Homogeneity and Exhaustiveness. According to Benzécri (1973, p.21), two principles should be fulfilled. i) *Homogeneity*: All variables in the table should be of the same nature. ii) *Exhaustiveness*: The data should constitute an exhaustive or at least a representative inventory of a real research field. The exhaustiveness principle is seen to be quite at odds with the *parsimony* principle often advanced in the conventional statistical methodology, especially in connection with regression procedures.

Large–size tables. GDA is eminently apt at revealing the structural complexities of large–size tables[19]; all authors stress this fact[20]. Small tables

[19] "Large–size tables" should of course not be confused with "large frequency tables"; a 5×4 contingency table involving millions of individuals is still a small–size table!

[20] See Benzécri (passim). Lebart & al. (1984): "Large data sets often contain so many