Esa Reunanen & Risto Suikkanen, Tampere

Latent Class Analysis: Wandering in Latent Space

As it is often mentioned, Bernard Berelson's *Content Analysis in the Communication Research* became the fundament of defining the self-understanding of the research technique called *content analysis*. Berelson (1952,18) defined the technique as an instrument for objective, quantitative and systematic description of the manifest content of communication. Since that the meaning of the term *content analysis* has referred to the research tradition based on the definition given by Berelson.

Berelson was, of course, well aware of the so called *qualitative content analysis* (there is a chapter in his book reflecting qualitative content analysis). When he used the term the former part of the expression was always embraced by quotation marks: he spoke always (in his book) about ”qualitative” content analysis. In spite of that he wasn't very hostile to this kind of research. It was useful - as Pietilä (1997,152) states - at least as a basis for formulation of hypothesis. Obviously Berelson felt some kind of uneasiness because of the low level of formalization and quantification which led to not very self evident modes of argumentation. But after all he was ready to express respect to this old art of interpretation drawing on glorious and authoritative sources like philosophy, rhetorics or literary criticism.

The basic topics in the dispute concerning the quantitative content analysis were of course 'quantitaveness' and 'manifest content'. While presenting his definition Berelson at the same time made clear that there were limits in the application of the method (1952, 13 -20). It could be used only if the ”units of the content” were as significant. One couldn't do quantitative content analysis if one word or one statement was as important as all the rest of the material. As such it wasn't necessary to use numbers, it was enough to use expressions which had some kind of connection to quantitative relations (as the words *more* or *less* have.).[1]

Another limitation was the axiom that only those texts which were based on common universe of discourse between the sender and the audience could be taken as an object of content analysis. This would be a quarantee that the meanings encoded and decoded were common enough. As an example Berelson mentioned news story about railway accidents. As an opposite example he mentioned 'modern' poetry which couldn't be taken as an object of content analysis at least as far as the meanings were concerned[2]. Berelson hardly denied the possibility that a news story about an accident in railroads could generate dozens of other meanings (”the railroads are not well kept”, ”We can loose our life in every second”). He probably wanted only to emphasize what is common to everyone: a message informing of an accident in the railroads. The aim was to exclude pragmatics which was beyond the scope of content analysis.

A well known reaction to Berelson's book was Sigfried Kracauer's article *The challenge of qualitative Content Analysis* (Public Opinion Quarterly 52/53 nr 4). According to Kracauer the typical form of the quantitative analysis had been to cut continuous dimensions into pieces by fullpacked classification scales in order to reveal if the content of communication was ”for” or ”against” this or that. This wasn't, however, by any means a quarantee of exact analysis because by doing so...

> „the atomistic character of the resulting data precludes a relevant examination of the relations within the text. It is worth of noticing that often exactly these relations play decisive role in determining the orientation of texts. Everyone who is used to interpretation of texts knows - all the same if he supports gestalt psychology or not -

> that due to their structure texts have often ”orientation” which can’t be concluded merely by listing their elements“. (1990, 340.)

In the end of his article (1990, 350) Kracauer still repeated that texts are meaningful wholes and not mere agglomerations of facts. Content of their elements...

> „is no longer their content if it is detached from the texture of intimations and implications to which it belongs and taken literally; it exists only with and within this texture - a still fragmentary manifestation of life, which depends upon response to evolve its properties. Most communications are not so much fixed entities as ambivalent challenges. They challenge the reader or the analyst to absorb them and react to them. Only in approaching these wholes with his whole being will the analyst be able both to discover and determine their meaning - or one of their meanings - and thus help them to fulfil themselves“.

Kracauer’s idea is no doubt already too familiar to the readers of this presentation: observing and listing the elements of a text doesn’t necessarily reveal the totality build upon these elements.

The thoughts of Paul F. Lazarsfeld may have taken completely different course as Kracauer’s in other respects except in one. Using Hempel’s expression Lazarsfeld - who often referred to the texts of the Vienna Circle - stated that the social sciences were in the ”pre-theoretical stage of research” Lazarsfeld 1959, 485). According to Lazarsfeld the task at that time was to produce empirical results and knowledge as well as to have methodological discussion in order to be clear how the methods have been constructing theoretical concepts used and studied. Thus one would build a storehouse of empirical results and their production processes which both could have their ”systematic use in a ”theory” which we hope will one day develop” (1959, 485).

The main obstacle in producing knowledge for logistic purposes was the underdevelopment of methodological capacities and abilities. According to Lazarsfeld researchers were helpless in front of empirical indicators referring to different directions. They ought to be seen as meaningful wholes ”but it is surprising how grievously the issue can be missed by scholars when they leave the field of their special research experience” (1959,482).

Lazarsfeld himself was developing a multivariate method similar to factor analysis but based on probabilities and thus directly applicable to data produced by using nominal scales (Lazarsfeld 1950, 469). The name he used was Latent Structure Analysis (LSA). In outlining LSA Lazarsfeld aimed at producing a method which would be very widely applicable in the social sciences. The basic idea was that survey results or other observations form the manifest object. This object is a product of some latent aspects or dimensions (attitudes, mental states or characteristics) which can’t be directly observed but which can be mediated to the field of sense experience by using tests, surveys or taking every day events as a point of departure for producing data. This data based on observations ought to be reduced to the latent structure which is a description of how persons or other objects under study are situated in the latent dimensions. If there would be several dimensions somehow connected to each other one could speak about a ‘latent space’ in which the units of observation were situated in a certain way. (Lazarsfeld 1950, 1959.)

There are several latent structure methods that try to find out latent variables that structure the data and ‘cause’ the dependencies between manifest variables. The methods differ from each other in what kind of variables the manifest and latent variables are thought to be. For example *factor analysis* assumes both manifest and latent variables to be continuous. *Latent class analysis*, then, assumes both latent and manifest variables to be discrete (ordinal or nominal). Under the title latent structure methods one has still to add two more techniques: *latent trait analysis* enables the characterization of continuous latent variables from discrete observed variables and *latent profile analysis* enables the characterization of discrete latent variables from continuous observed variables. (McCutcheon 1987, 7.)

The problem in 1950's was that algorithms needed were not developed. Mathematical problems were solved in 1970's. Computer technique on its part has developed rapidly enough in 1980s. As far as we understand since that the infrastructure for LCA-studies has been ready. Until today LCA hasn't, however, become such a widely used and widely known method as factor analysis. We don't know, of course, why. One reason may be the slow and 'step by step' progress of mathematical and technical arsenal needed. When this process was going on factor analysis already occupied the field and some kind of 'critical mass' developed on it. Also the rise of qualitative methods since 1970's and 1980's hasn't been likely to direct attention to new forms of quantitative methods among students.

In 1950's LCA (or rather LSA at that time) obviously has been used in attitude and opinion polls and studies on how people read newspapers or listen to the radio. There are also traces of using LCA in what could be called 'critical social research' in 1980's (see Aitkin & Anderson & Hinde 1981). In 1990s the method has been used also in content analysis. Most of this kind of work has been done in the Peace Research Group in the University of Konstanz under the management of Wilhelm Kempf who has referred to LCA (Kempf 1994,6) as a possibility of doing statistical analysis overcoming some aspects of Kracauer's critic of quantitative content analysis. In Finland Ari Heinonen (1997) has used LCA in analysing survey data.

In this paper we try to demonstrate the main principles and procedures of LCA. First we try to find out what is meant with the 'principle of local independence'. After that we shall demonstrate the procedures of LCA by using a data produced in a query of what Finnish editors think about Internet (Heinonen 1997). In the third part of our presentation we try to bring out the characteristics of LCA by comparing it to some other methods of multivariate analysis. In the end we shall shortly reflect LCA as a method of content analysis.

LCA and the principle of 'local independence'

LCA has emerged from the 'empirical social research' proposed by Paul F. Lazarsfeld, who formed his basic views (as a psychologist) in Vienna of the 1920's and 1930's. Lazarsfeld adopted a standpoint, according to which the aim of social research - in any field of research - was to study and understand human action. The theoretical concepts needed (attitude, status, motivation...) were typically not in the reach of direct observation, and Lazarsfeld's passion was - in the spirit of the general science - theorising that aimed at controlled ways of connecting empirical data to what it should reveal. This involves methods of collecting and interpreting observed data. The meaning of the word 'latent' in the 'Latent Structure Analysis' seems twofold. First, it indicates the 'latent space' described by the theoretical concepts. Second, it points to mathematical relations implicit in empirical data. In the end, then, the word 'latent' tries to catch especially the relation of these two (the latent space and the mathematical relations).

The principle of local independence is very central from the point of view of the connection of latent space and the empirical material describing it. The principle is kind of ‘if-then’ -axiom. According to it, the coding units are alike in regard to this or that latent property (latent continuum), if they produce a statistically unrelated distribution in tests measuring this continuum. This can be demonstrated by observing answers of two different groups to two questions.

TABLE 1. *Statistically unrelated and related distributions.*

Statistically unrelated distribution

		Question	*2*	
		+	-	
Quest.	+	75	15	90
1	-	15	3	18
		90	18	108

Statistically related distribution

		Question	*2*	
		+	-	
Quest.	+	35	19	54
1	-	19	35	54
		54	54	108

In the first group 90 respondents of 108 have answered ‘yes’ to the first question. 75 of these 90 respondents have answered ‘yes’ also to the second question, while 15 of them have answered ‘no’. The ratio is, then, 75:15. Within those who have answered ‘no’ to the first question, the ratio concerning the other question is exactly the same, 15:3. This kind of distribution, in which the ratio stays same when adding further questions, is statistically unrelated. According to the principle of local independence, the group of coding units that produce this kind of unrelated distribution is homogeneous in regard to the latent property measured by the variables.

In the latter group, the distribution of the answers given is not statistically unrelated. Here are 54 positive and 54 negative answers to the first question. Within the group that has answered ‘yes’ to the first question the ratio concerning the other question is 35:19, while corresponding ratio within the group that answered ‘no’ to first question is 19:35. In this ‘statistically related’ group the answer to the first question clearly ‘affects’ what the answer to the second question will be, and this is why this group is heterogeneous.

The statistical relatedness implicates that the questions (variables) have one or more ‘common factors’, that distinguish respondents (coding units) from each other. Relatedness is, firstly, an indication that the questions have something ‘in common’, and secondly, that the respondents are different in regard to that common factor. In the first group of table 1 the questions may well have something in common because so many respondents have answered positively to both of them. This common factor doesn’t anyhow distinguish respondents from each other because answers to the first question seem not ‘affect’ to the answers to the second question.

In the statistically unrelated distribution it is also true, that the mathematical probability of the joint occurrence of certain answers is same as the their real share in the data. For example, within the first group in table 1 the mathematical probability for a positive answer to both of the questions is 90/108*90/108 = 0,69, which is exactly the real share of the joint occurrences of positive answers (75/108 = 0,69). In the latter group the real share of ‘yes-yes’ -answers (35/108 = 0,32) is considerably greater than the mathematical probability (54/108*54/108 = 0,25), and this is a proof of statistical dependence.

The aim of LCA is to divide heterogeneous groups to homogeneous subgroups. To demonstrate this, we follow loosely Lazarsfeld’s example (1959, 496-500), and assume, that the heterogeneous group presented in table 1 had answered to the next two questions:

Question 1: Do the big oil companies control too much of the oil business?
Question 2: Is the oil industry wasteful of our natural resources?

In this example the respondents are coding units, the questions are variables and the answers given are categories of the variables. The positive answers can be thought to implicate negative attitude to oil industry and the negative answers accordingly positive attitude. Although the questions strictly reading are about different things, they both can be seen testing also more general attitude to oil industry. Exactly this ‘general attitude’ would be the latent factor that could explain observed statistical dependence, and in regard to which the respondents would be different.

General attitude to oil industry can be understood as a continuum on which each respondent is at certain point. Those situating near the negative end of the continuum very probably answer positively to the questions, while those situating near the positive end of the continuum are more likely to answer negatively. Some respondents have hardly any attitude to oil industry, and their answers are as probably negative as positive. On the basis of the principle of local independence the respondents can be divided to three homogeneous subgroups as is presented in table 2:

TABLE 2. *Heterogeneous data divided in homogeneous subgroups.*

Do the big oil companies control too much of the oil business?

Total data

		+	-	
Is the oil industry wasteful of our natural resources?	+	35	19	54
	-	19	35	54
		54	54	108

Do the big oil companies control too much of the oil business?

		Subgroup 1				*Subgroup 2*				*Subgroup 3*		
		+	-			+	-			+	-	
Is the oil industry wasteful of our natural resources?	+	1	5	6	+	9	9	18	+	25	5	30
	-	5	25	30	-	9	9	18	-	5	1	6
		6	30	36		18	18	36		30	6	36

According to the logic of LCA the heterogeneity of total data is caused by the fact that it is a mixture of two or more inherently homogeneous subgroups. This mixture can be unmixed by dividing the data to statistically unrelated subgroups. On the basis of the principle of local independence, these subgroups should be homogeneous in regard to those latent factors that caused the dependency in the total data.

Each of the three subgroups in table 2 should consist of respondents who have similar attitude to oil industry. The attitudes within subgroup 1 are most positive which can be clearly seen because the answers to the questions are most probably negative. Subgroup 2, then, consists of respondents, who have no clear attitude to oil industry: the answers are as probably positive or negative. In the third group, the answers are very probably positive, which implicates negative attitude.

When talking about homogeneity of subgroups it is important to remember that the question is about homogeneity in regard to the latent factor (common to the variables), not about similarity of manifest answers. For example in the subgroup 2 all possible answer-combinations are equally represented. Anyhow, this is a very indication of the subgroup's homogeneity: answers of uncertain or indifferent respondents even should distribute equally to all alternatives.

Although Lazarsfeld stresses strongly this kind of interpretation, where latent variable is unidimensional (positive or negative attitude or something in between), LCA's latent variable can be interpreted to be also multidimensional. In extreme case the latent variable is nominal in the sense that the classes represent separate types that are homogeneous, but do not situate themselves on any common latent dimension or scale. According to McCutcheon (1987, 8), the multidimensional interpretation is more general, and it can be used either as a method for empirically characterizing a set of latent types within a set of observed indicators, or as a method for testing whether a theoretically posited typology adequately represents the data.

Of the subgroups one should, finally, note that on the basis of groupings by LCA it is not possible to say by sure which single coding units belong to which subgroups. For example, a person who has answered positively to both of the questions presented above can belong to any of the three subgroups. However, it is still possible to say, that she belongs most probably to subgroup 3, second probably to subgroup 2, and least probably to subgroup 1.

The construction of LCA-classes in practice: an example analysis

Let us next examine the procedures and mathematical principles of LCA with a concrete example data. We have chosen as an example a survey of Finnish chief editors about their attitudes to Internet (Heinonen 1997). A survey is as good an example as is a content analysis because LCA works similarly with both kinds of data. In our example the chief editors were presented statements, and they had to say if they agreed or disagreed with them. We have chosen three such statements, that make up the variables of our example analysis. The categories of variables are 0 (agree), 1 (can't say), or 2 (disagree):

> **Statement 1:** Technology is proceeding so fast, that the newspapers must already now prepare themselves for net-publishing.
>
> **Statement 2:** Technicality and expensiveness of Internet-use will still a long time restrict its popularity among the public.

> **Statement 3:** Net-publications that are read on the computer-screen will not interest general public.

The Lacord-program (developed by Jürgen Rost) computes first the distributions of variables in total data. In LCA these distributions are called ‘1-class solution’ (table 3). The only difference to normal percentage tables is that Lacord gives the shares as probabilities, not as percentages. Probability of 1.000 equals with the share of 100 %, probability of 0.500 equals with the share of 50 % and so on.

TABLE 3. *Opinions on the development of Internet and its significance to the newspapers.*

```
                                        Newspapers    Technicality  Net-publications
                                        must prepare  will restrict don’t interest
                                                      already now   still long
           general public

1.CLASS  1.000  *0* agree               0.861         0.750         0.556
                *1* can’t say           0.056         0.056         0.056
                *2* disagree            0.083         0.194         0.389
```

It can be seen in the table that the size of the class 1 is 1.000 which means, that 100 percent of the answers belong to this class. Numbers 0 - 2 denote categories of the variables. It can be seen in the table, for example, that 75,0 percent of the respondents have agreed with the statement that ”technicality and expensiveness of Internet-use will still a long time restrict its popularity among the public”. 19,4 percent of the respondents have disagreed with the statement.

When computing total distribution of the data Lacord computes also an index that indicates how good a description of data it is in the sense of homogeneity explained above. This index is LOG-Like and describes the probability of the whole data set under this („1-class“) solution. LOG-Like, for its part, is computed from the probabilities of so called *coding patterns*. A coding pattern is a combination of categories of several variables that ‘belong’ to a single coding unit. For example a row in a data matrix comprises a coding pattern. In table 3 the categories can combine 27 different coding patterns. One example of a coding pattern is 000 (agree with all the statements), another example is 020 (agrees with the first and the third statement but disagrees with the second statement).

Assuming local independence the probabilities are computed for each coding pattern by multiplying the probabilities of single categories with each other. The probability of a coding pattern is the greater, the more coding units have its categories. For example the probability of coding pattern 000 is 0,861 x 0,750 x 0,556 = 0,359037, and the probability of the coding pattern 020 is 0,861 x 0,194 x 0,556 = 0,09287. The probability of the whole solution could be computed by multiplying the probabilities of each coding unit with each other. In practice, the probability is not computed this way because the product would be so small that it would be hard to operate with. A corresponding outcome will be reached by adding up the logarithms (ln) of each coding unit’s probabilities. This is how the LOG-Like -index is computed. The greater the probabilities of coding units are, the ‘better’ is the LOG-Like -index, and the more homogeneous is the group. This is because in a heterogeneous data the mathematical probabilities of coding patterns are typically smaller than their real shares in the data (in some cases, though, the mathematical probability may overcome the real share). If the data is changed more homogeneous, the mathematical probabilities will approach the real shares and also LOG-Like -index will grow.

The computations of LOG-Like are demonstrated in table 4. The table presents the coding patterns of our example data, the number of coding patterns (N), their share in the data (N/72), their mathematical probabilities (p), logarithm of these probabilities ln(p), and, lastly, the LOG-Like -index.

TABLE 4. *The probabilities of the coding patterns and the LOG-Like -index computed from them: one class -solution.*

	Coding pattern	**N**	**N/72**	**p**	**ln(p)**	**N*ln(p)**
	0 0 0	27	0,375	0,359	-1,024	-27,657
Total **data** g=1,000	0 0 1	1	0,014	0,036	-3,320	-3,320
	0 0 2	18	0,250	0,251	-1,382	-24,867
	0 1 2	3	0,042	0,019	-3,976	-11,929
	0 2 0	4	0,056	0,093	-2,377	-9,506
	0 2 1	2	0,028	0,009	-4,672	-9,344
	0 2 2	7	0,097	0,065	-2,734	-19,136
	1 0 0	3	0,042	0,023	-3,757	-11,271
	1 1 1	1	0,014	0,000	-8,647	-8,647
	2 0 0	5	0,069	0,035	-3,364	-16,818
	2 2 0	1	0,014	0,009	-4,716	-4,716
	Total	72	1,000	0,899	-39,968	-147,211
					LOG-Like:	**-147,211**

The sum of the coding pattern -probabilities (0,899) is clearly smaller than one. This means that the variables are statistically related. LOG-Like index is another way to express this degree of relatedness, and it too is computed in the table 4.

The idea of LCA is to divide the data into subgroups so, that the variables in each group are as unrelated as possible. In practice, Lacord first divides the data into two accidental groups, and iterate grouping until the LOG-Like for two-class solution is as good as possible.[3] After this, the program computes in the similar way the best possible three, four, five and six class -solutions. Increasing the number of classes improves the LOG-Like -index. In other words, increasing the number of classes makes the classes more homogeneous. We will discuss later how the right number of classes can be chosen.

TABLE 5. *The probabilities of the coding patterns and the LOG-Like -index computed from them: two class -solution.*

	Class 1: g_1=0,555				Class 2: g_2=0,445				Total data, two class -solution: g=1				
Coding												ln	N*ln
pattern	n_1	n_1/40	p_1	g_1*p_1	n_2	n_2/32	p_2	g_2*p_2	N	N/72	$g_1p_1+g_2p_2$	($g_1p_1+g_2p_2$)	($g_1p_1+g_2p_2$)
0 0 0	27	0,675	0,678	0,376	0	0	0,001	0,001	27	0,375	0,377	-0,976	-26,348
0 0 1	0	0	0,000	0,000	1	0,031	0,072	0,032	1	0,014	0,032	-3,441	-3,441
0 0 2	0	0	0,000	0,000	18	0,563	0,503	0,224	18	0,250	0,224	-1,497	-26,941
0 1 2	0	0	0,000	0,000	3	0,094	0,106	0,047	3	0,042	0,047	-3,055	-9,166
0 2 0	4	0,100	0,097	0,054	0	0	0,001	0,000	4	0,056	0,054	-2,919	-11,674
0 2 1	0	0	0,000	0,000	2	0,063	0,034	0,015	2	0,028	0,015	-4,190	-8,380
0 2 2	0	0	0,000	0,000	7	0,219	0,238	0,106	7	0,097	0,106	-2,245	-15,717
1 0 0	3	0,075	0,066	0,036	0	0	0,000	0,000	3	0,042	0,036	-3,312	-9,936
1 1 1	0	0	0,000	0,000	1	0,031	0,000	0,000	1	0,014	0,000	-8,442	-8,442
2 0 0	5	0,125	0,131	0,073	0	0	0,000	0,000	5	0,069	0,073	-2,619	-13,097
2 2 0	1	0,025	0,019	0,010	0	0	0,000	0,000	1	0,014	0,010	-4,565	-4,565
Total	40	1,000	0,991	0,550	32	1,000	0,954	0,425	72	1,000	0,975	-37,262	-137,709
												LOG-Like:	**-137,709**

In table 5 it is presented how Lacord-program divides the example data into two LCA-classes.[4] The table shows clearly, how the mathematical probabilities of the coding patterns ($g_1p_1+g_2p_2$) correspond now better to the real share of the coding patterns (N/72). Also the LOG-Like -index is now better than in the one class -solution. We have made tables 4 and 5 only to demonstrate how the computations of LCA are made. In practical LCA-work this kind of tables are not needed. Lacord-program gives only shares of the categories in each LCA-class, the LOG-Like -index and several other, not so essential indexes. Let us now examine the two class -solution in the form that Lacord-program presents it (table 6):

TABLE 6. *Opinions on the development of Internet and its significance to the newspapers: two class -solution.*

```
                                            Newspapers    Technicality  Net-publications
                                            must prepare  will restrict don't interest
                                                          already now   still long
              general public


1.CLASS       0.555 *0* agree               0.775         0.875         1.000
                    *1* can't say           0.075         0.000         0.000
                    *2* disagree            0.150         0.125         0.000

2.CLASS       0.445 *0* agree               0.969         0.594         0.002
                    *1* can't say           0.031         0.125         0.125
                    *2* disagree            0.000         0.281         0.874
```

The homogeneity within and heterogeneity between the classes can be seen clearly also in the table 6. The classes differ most obviously when observing the statement about the interest in net-publications. All respondents in class one agree with the statement, whereas most respondents in class two disagree with it. This could be interpreted for example by naming the latent continuum that explains the grouping as ‘technology optimism’. The respondents in class 2 would then be clearly more technology optimistic than respondents in class 1. Especially answers to the statement about interest in net-publications support this interpretation. Nobody of the pessimists believe that net-publications interest general public whereas the optimists are of the opposite standpoint. Also the other two statements support this interpretation, although the differences between classes are not that remarkable in these.

In the three class -solution the classes are still a bit more homogeneous (LOG-Like = -135,415) than in the two class -solution. If we still hold on our assumption that the latent property is technology optimism - technology pessimism -continuum, it seems obvious, that most pessimist respondents are now gathered in second class. The class is freed from respondents who think that newspapers should prepare themselves for net-publications. The most optimist respondents, for their part, are gathered in the first class. In this class are no more anyone who thinks that Internet’s technicality will still long restrict popularity among the public. Both of these classes are quite small, both contain about 13 percent of respondents. The biggest class (class 3) consists of respondents who’s answers seems at first glance quite contradictory.

TABLE 7. *Opinions on the development of Internet and its significance to the newspapers: three class -solution.*

	general public		Newspapers must prepare	Technicality will restrict already now	Net-publications don’t interest still long
1.CLASS	0.134	*0* agree	**0.896**	0.001	0.000
		1 can’t say	0.104	0.416	0.290
		2 disagree	0.000	**0.584**	**0.710**
2.CLASS	0.134	*0* agree	0.067	**0.889**	**1.000**
		1 can’t say	0.311	0.000	0.000
		2 disagree	**0.622**	0.111	0.000
3.CLASS	0.732	*0* agree	**1.000**	**0.861**	**0.576**
		1 can’t say	0.000	0.000	0.023
		2 disagree	0.000	0.139	**0.402**

On the optimism-pessimism -continuum the third class could maybe consist of respondents who are neither very optimistic or pessimistic in their attitude. On the one hand they think it is important to prepare oneself for net-publications, but on the other hand they think that the technicality will still a long time restrict its popularity among the public. Another, and maybe better interpretation is, that in addition to optimism - pessimism -continuum, the answers are based also on another latent continuum. This could be a kind of ‘be prepared for anything’ -way of thinking. The respondents in the third class want secure their rear even if they don’t believe that Internet would soon spread among the public. The third class would then represent ‘securers’, when the first and second class would represent respondents who think that only those actions are reasonable that seem necessary.

It is still necessary to note, that LCA primarily finds out the structure of the data, not which coding units belong to which classes. It is possible only to say the probability with which each coding unit belongs to each class. One and the same coding unit can also at the same time belong - with different probabilities - to several classes.

How to choose the most informative number of classes?

The homogeneity of classes is most perfect when the variables in each class are perfectly unrelated. Homogeneity can be improved by adding the number of classes, and the classes are perfectly homogeneous at latest when each different coding pattern have a class of its own. This is called ‘saturated model’. In our example there were 11 different coding patterns, and thereby the saturated model would consist of 11 classes. In practice, it is not advisable to increase the number of classes very much, because this would make the solution too complex and unillustrative. The problem is to find the most ‘efficient’ solutions in between the one class solution and the saturated model.

There are several indexes to find the best number of classes. They all are based on some information theoretical principles with which the benefit gained by the homogeneity of classes is made comparable with the drawbacks of the complexity. The indexes are, though, theoretically justified, but they still can suggest different solutions and are therefore not universal. It should also be noted that it may sometimes be reasonable to favour simplicity or accuracy at the expense of the theoretical informativeness. Lacord-program computes automatically an index that is called AIC by its inventor (Akaike’s Information Criterion). The index can be expressed with next equation (Kempf 1994, 14):

$$AIC = -2\ln(L(x)) + 2n(P)$$

Part ‘ln(Lx))’ of the equation denotes the LOG-Like -index presented above. Part ‘n(P)’ of the equation denotes the number of parameters that have to be estimated for the solution. For example sizes of LCA-classes and shares of categories in each LCA-class are this kind of parameters to be estimated. When increasing the number of classes, ln(Lx)) will approach zero, which decreases AIC, because ln(L(x)) is always a negative figure. At the same time, the number of parameters to be estimated will increase sharply, which, for its part, makes AIC increase. The best solution is the one, that gives the smallest AIC-index. The idea of AIC is, so, to find the number of classes after which the benefit gained with adding classes doesn’t cover the drawback caused by the increased complexity of the solution.

Another well known index is called BIC (Best Information Criterion). Compared to AIC it weights more heavily the drawbacks of complexity. Lacord doesn’t compute BIC, but it is easy to compute of its equation (Kempf 1994, 14):

$$BIC = -2\ln(L(x)) + \ln(n)n(P)$$

The only difference in the equations of AIC and BIC is that the multiplier of the number of parameters is ‘2’ in AIC, whereas it is the logarithm of the number of coding units ln(n) in BIC. The logarithm is the bigger the more there are coding units. This is why especially with large datas BIC tends to suggest fewer classes than AIC.

Stanley S. Sclove (1987, 337) states, that the indexes differ from each other in how sharply the multiplier penalising from complexity increases when increasing coding units. The most sharply increasing multiplier is BIC’s ln(n), while the most flatly increasing multiplier is ln(ln(n)). This ‘flattest’ multiplier approaches the constant multiplier 2 that is used in AIC when there are plenty of coding units. We name this ‘flattest’ multiplier here CIC:

$$CIC = -2\ln(L(x)) + \ln(\ln(n))n(P)$$

It seems now evident, that the smallest theoretically justified number of classes is obtained by using BIC-index and the greatest number is obtained by CIC-index. If both indexes suggest the same number of classes, this number is globally best. If the suggestions differ from each other, any one of the suggested numbers or some number in between can be chosen.

In our example analysis there are 72 coding units, and logarithm of 72 is 4,28. Logarithm of 4,28 is correspondingly 1,45. Consequently, when using BIC, the number of parameters must be multiplied with 4,28 and when using CIC it must be multiplied with 1,45. Thus it is easy to see how BIC emphasises most strongly the drawbacks caused by complexity of many-class solutions.

FIGURE 1. *The multipliers penalising for complexity in relation to the number of coding units.*

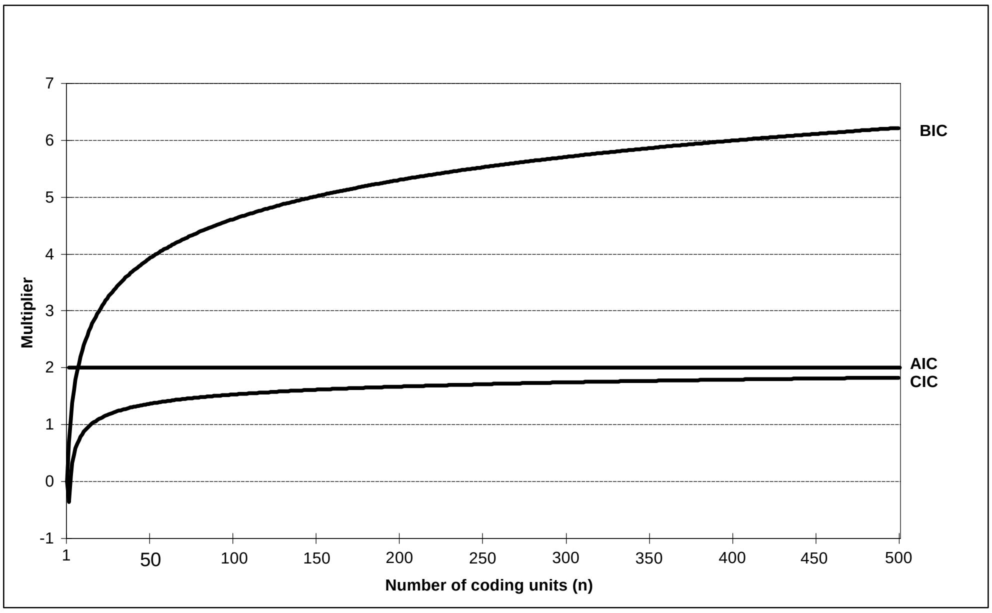


TABLE 8. *Some indexes of example analysis.*

Number of classes	LOG-Like[5]	Npar	LIK.ratio	DF	AIC	BIC	CIC
1.	-147,283	6	31,137	20	306,565	**320,226**	303,285
2.	-137,739	13	12,051	13	**301,479**	331,075	**294,369**
3.	-135,415	20	7,403	6	310,831	356,363	299,893
4.	-134,890	27	6,351	-1	323,779	385,250	309,016
5.	-132,154	34	0,88	-8	332,308	409,715	313,716
6.	-131,719	41	0,01	-15	345,438	438,781	323,018

In addition to LOG-Like and AIC-indexes Lacord prints for every solution the number of parameters estimated (Npar), the degree of freedom (DF), and a chi-square distributed test statistic (LIK.ratio) which compares the LOG-Like index of the respective solution (H_0) with that of the saturated model (H_1). Lacord doesn't print BIC or CIC, but they can easily be computed of the equations presented above. If the degree of freedom (DF) is negative, the solution is not uniquely defined and it must be omitted. So, in our example the choice must be done between one-, two- or three class -solutions. Often different indexes suggest the same number of classes (Kempf 1994, 14-15), but in our example BIC suggested one class -solution, but AIC and CIC suggested two class -solution. We think, that all these indexes are just supporting devices, and they must not be obeyed blindly. Even in our example one can quite well choose either one-, two-, or three class -solution or use them all when interpreting the results.

LCA compared with other methods of multivariate analysis

Lazarsfeld developed LCA in order to create a multivariate analysis method that would work also with nominal variables. From this emerges one essential difference between LCA and factor analysis: For LCA the phenomena observed are qualitative, for factor analysis they are quantitative. In other words, for LCA the categories 1 and 4 are different exactly in the same manner as the categories 1 and 2. Factor analysis, for its part, interprets the category 4 to be exactly three units greater than the category 1. This has practical implications in which kind of tasks each method is suitable. We try to demonstrate this later.

Both LCA and factor analysis try to bring out something latent. Factor analysis finds out the latent continuums that structure the data, and it tells also how strongly each continuum structure the data. The interpretations of factors are based on those variables that have great loadings (positive or negative) on each factor. These loadings can be interpreted to be correlations between the factor and each variable (Alkula et al. 1994, 270). LCA, for its part, shows particularly the configuration of the data on the latent variable that can be - unlike in factor analysis - either unidimensional or multidimensional (McCutcheon 1987, 8). In the case of multidimensional latent variable a single latent class is quite much like a single factor in factor analysis. The category probabilities of each variable in LCA-classes are comparable to factor loadings (McCutcheon 1987, 19). In the case of unidimensional interpretation, it is better to say that factors are like latent continuums in relation to which LCA-classes are as homogeneous as possible.

LCA's ability to find out multidimensional latent variables makes it useful in social sciences, where typologies have a prominent part. It makes it possible to find out latent styles or types that are not based on any continuum but that anyhow have some characteristics peculiar to them.

Factor analysis can be continued by computing so called factor scores for each coding unit and locating coding units to factors according to the scores. This procedure reminds LCA quite a lot, and it is interesting to test in practice how similar results the methods will give.

Table 9 presents a fictious data that consists of grades given to ten students in Swedish, English, and German languages. This data is possible to analyse both with

LCA and factor analysis, but the traits of each method must be taken account when interpreting the results.

TABLE 9. *Grades in languages in an example data.*

	Swedish	English	German
Aki	3	2	0
Arja	0	3	1
Auli	3	2	0
Axel	1	1	2
Asko	0	2	1
Adolf	1	1	2
Atte	1	0	3
Anna	1	0	2
Arto	1	0	3
Anne	1	1	2

It can be noted that the grades in Swedish and English correlate with each other slightly positively, but the grades in German correlate clearly negatively with both Swedish and English. Consequently, it seems probable that the variables could be condensed in one or two factors in which the correlations between variables would come out. Because of the nature of our little data we use principal component analysis instead of proper factor analysis. Principal component analysis is quite similar with factor analysis, and its aim is to ”include maximal amount of variance included in original variables in only a few - mutually unrelated - principal components” (Ranta & Rita & Kouki 1991, 459).[6]

The analysis produces first one principal component (factor), that explains the variance as well as possible. After this, it produces more principal components, that explain the reminder of the variance. In our example the analysis produced the two principal components presented in table 10. Together they explain 97 percent of the variance.

TABLE 10. *The principal components of the example data.*

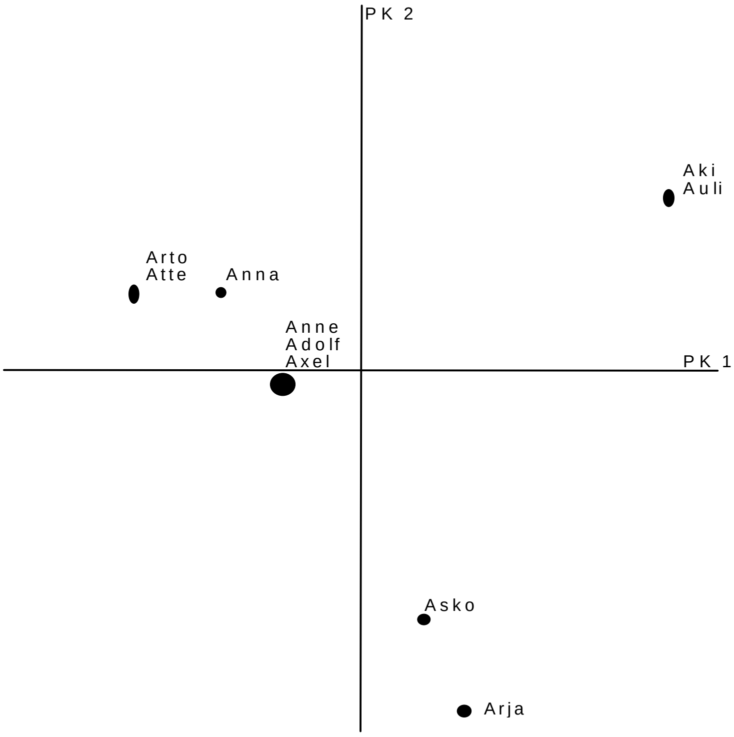
	PC 1	PC 2
Swedish	0,56	0,82
English	0,84	-0,51
German	-0,99	0,03
Degree of explanation	66 %	31 %

The first principal component is most important, and it expresses the factor that makes people succeed better in Swedish and English than in German, or correspondingly, better in German than in Swedish or English. The other principal component, then, explains why some people succeed better in Swedish than English or the other way round.

The analysis seems to condense very well the information in the table 9. By using a little bit imagination it would be possible to interpret the first principal component to express the influence of popular culture. The people who watch or listen a lot of programs in English will get good grades in English because they learn the language also on their leisure time. Instead, there is not enough time to grind away at German irregular verbs. The other principle component could be interpreted as expressing the influence of Swedish-speaking environment.

When the factor scores are computed for coding units, it is possible to observe how the students locate themselves on the continuums expressed by principal components. The horizontal axis in the figure 2 expresses the 'popular culture' -component and the vertical axis expresses the 'Swedish environment' -component. We have placed the coding units (students) on to the 'gravitational field' composed by these components with the aid of regression factor scores computed with the SPSS-program.

FIGURE 2. *Example data located in latent space by the factor scores.*



The persons seems to locate themselves quite clearly in three or four separate groups. Following our hypothesis, Aki and Auli would live in Swedish speaking environment and they would be quite heavy popular culture consumers. Asko and Arja, for their part, would live in Finnish speaking environment, and also they would also be slightly interested in popular culture. Atte, Arto and Anna would concentrate more in German language than in popular culture as also Anne, Adolf and Axel.

If we now present this grouping based on factor scores in the form typical for LCA, it looks like this:

TABLE 11. *Example data grouped by factor scores: a three class -solution.*

	Class size	**Grade**	**Swedish**	**English**	**German**
1.CLASS	0,200	0	**1,000**	0,000	0,000
Asko, Arja		1	0,000	0,000	**1,000**
		2	0,000	**0,500**	0,000
		3	0,000	**0,500**	0,000
2.CLASS	0,600	0	0,000	**0,500**	0,000
Arto, Atte		1	**1,000**	**0,500**	0,000
Anna, Anne		2	0,000	0,000	**0.667**
Adolf, Axel		3	0,000	0,000	**0,333**
3.CLASS	0,200	0	0,000	0,000	**1,000**
Aki, Auli		1	0,000	0,000	0,000
		2	0,000	**1,000**	0,000
		3	**1,000**	0,000	0,000

The aim of LCA is to depict how the data is arranged on the latent continuums that organise the data. Therefore it is reasonable to assume, that LCA classes would resemble quite a lot the grouping based on factor scores. The essential difference between the methods is, though, that LCA observes instead of quantitative variance only similarities and differences.[7] This can lead to differences in the groupings.

The three class -solution of LCA was exactly the same as the factor score -solution presented in table 11. However, the AIC-index suggested to choose the two class -solution in which classes 1 and 3 are merged:[8]

TABLE 12. *The example data analysed with LCA: a two class -solution.*

	Class size	**Grade**	**Swedish**	**English**	**German**
1,CLASS	0,400	0	**0,500**	0,000	**0,500**
		1	0,000	0,000	**0,500**
		2	0,000	**0,750**	0,000
		3	**0,500**	**0,250**	0,000
2,CLASS	0,600	0	0,000	**0,500**	0,000
		1	**1,000**	**0,500**	0,000
		2	0,000	0,000	**0,667**
		3	0,000	0,000	**0,333**

Also the two class -solution can be demonstrated with the figure 2. The first LCA-class is composed of the figure's right side, the second LCA-class of its left side. The 'vertical' continuum in the picture doesn't influence in the two class -solution. This can be seen in table 12 in the fact that the class 1 consists of those who have succeeded very well or very badly in Swedish. The ones with 'average grades' are in class 2. LCA's 'insensitiveness' to quantitative relatedness can be clearly seen here. For LCA it is just the same if the number is great or small, because LCA's classifications are based only on similarities and differences of the numbers.

Let us next imagine, that the figures in our example were not grades in languages but answers to a query by the Finnish Tourist Board, and the respondents were asked the thing that attracts them most in Sweden, England and Germany. The categories would have been 0) nothing, 1) nature, 2) culture, 3) standard of living. The variables would be Sweden, England and Germany. Principle component analysis and factor analysis give hardly any sensible results with this kind of nominal variables,[9] but LCA's results are still reasonably interpreted. The first LCA-class likes specially English culture and standard of living, but doesn't necessarily see anything charming in Sweden or Germany. If something in Sweden attracts, it is standard of living. If something in Germany attracts, it is nature. The second class, then, is unanimously attracted by Swedish nature and most of the admire also German culture. Half of them sees nothing pleasant in England, another half consider nature as the most charming feature in England.

This LCA-solution could be interpreted so that the second LCA-class would express a factor that explained both attraction to Swedish nature and German culture. Maybe this factor could be German culture itself. Perhaps persons in the second class are from Germany, and have learned there to appreciate Scandinavian nature and at the same time gained a little bit suspicious attitude towards England. Own cultural heritage with Wagner, Hegel etc. is of course admired too. This kind of hypothesis can be tested by checking demographical facts of the persons in second class. This is a procedure commonly applied in LCA, and there is a special program for doing the computations. Here it is enough if we check from the original data the names of the persons who have answered 1 (nature) for Sweden and 2 (culture) for Germany: Axel, Adolf, Anna and Anne. Our hypothesis seems to gain some verification...

Both LCA and factor analysis are clearly quantitative methods, but LCA observes qualitative phenomena while factor analysis observes quantitative phenomena (cf. Suhonen 1994, 74-76). In content analysis the variables are often - although not always - nominal, and this is why LCA seems to be specially suitable for it.

LCA and distance measures

LCA and factor analysis resemble each other because they both are based on dependencies between variables. They don't simply put similar observations to same class. Instead they find factors that would explain dependency (factor analysis) or classify most similar coding units to same classes (LCA). Here we make a difference between an observation and a coding unit. Observation mean for example a certain coding pattern, but a coding unit means for example a respondent who has a certain attitude. Similar coding units don't necessarily produce similar observations.

It is possible to classify observations also less statistically, straight on the basis of their manifest differences and similarities. There are several distance measures for this purpose. One of the most popular is so called general distance measure (Pietilä 1970, 112) that is also called Euclidean distance (Everitt 1995, 46). We classify here our school-grade data using SPSS-program's procedure that creates the classes so that in each class the distance between observations is as small as possible. Classification proceeds hierarchically so, that exactly similar observations are grouped first. In the example data there are six classes like this.

The next step is to compute the distances of these 6 classes with the distance measure $D = \sqrt{\sum d^2}$. For example the distance of coding pattern 112 from the coding pattern 320 can be obtained by computing first how much each category differs between the coding patterns (d), taking a square of this, counting these squared differences together, and taking a square root of the sum. For the first variable the difference of categories is 2, for the second it is 1 and for the third it is 2. This figures are squared and summed which means 4+1+4=9. The distance of coding patterns is the square root of this sum, i.e. 3. The distances between every different coding pattern are computed similarly, and the patterns with the smallest distance are merged with each other. This procedure is continued until finally all the observations belong to the same class.

TABLE 13. *Example data classified by coding patterns.*

	Swedish	**English**	**German**
Adolf, Axel, Anne	1	1	2
Aki, Auli	3	2	0
Atte, Arto	1	0	3
Arja	0	3	1
Asko	0	2	1
Anna	1	0	2

The problem with this kind of cluster analysis is to find out the ‘right’ number of classes, and it seems that there is no general rule for this. We chosen as examples the two and three class solutions, and to make the comparisons easier, we present them in the form typical for LCA.

TABLE 14. *Example data classified by the distance measure: a three class -solution.*

	Class size	**Grade**	**Swedish**	**English**	**German**
1.CLASS	0,600	0	0,000	**0,500**	0,000
Arto, Atte		1	**1,000**	**0,500**	0,000
Anna, Anne		2	0,000	0,000	**0.667**
Adolf, Axel		3	0,000	0,000	**0,333**
2.CLASS	0,200	0	0,000	0,000	**1,000**
Aki, Auli		1	0,000	0,000	0,000
		2	0,000	**1,000**	0,000
		3	**1,000**	0,000	0,000
3.CLASS	0,200	0	**1,000**	0,000	0,000
Asko, Arja		1	0,000	0,000	**1,000**
		2	0,000	**0,500**	0,000
		3	0,000	**0,500**	0,000

The three class solution produced by distance measure is exactly the same as the solution produced by factor scores. LCA, for its part, suggested a two class solution, in which Aki, Auli, Asko and Arja were grouped into the same class. In the two class solution produced by a distance measure also Asko and Arja were put into the biggest group as is presented in table 15.

TABLE 15. *Example data classified by the distance measure: a two class -solution.*

	Class size	**Grade**	**Swedish**	**English**	**German**
1,CLASS	0,200	0	0,000	0,000	**1,000**
Aki, Auli		1	0,000	0,000	0,000
		2	0,000	**1,000**	0,000
		3	**1,000**	0,000	0,000
2,CLASS	0,800	0	**0,250**	**0,375**	0,000
Arto, Atte		1	**0,750**	**0,375**	**0,250**
Anna, Anne		2	0,000	**0,125**	**0,500**
Adolf, Axel		3	0,000	**0,125**	**0,250**
Asko, Arja					

The use of a distance measure seems to lead, at least in our example, to the situation that step by step the observations are accumulated into the biggest, ‘average’ -class. The classification by distance measures is easy to understand because it is based on subtractions between categories of different coding patterns. Aki and Auli are in their own class still in the two class -solution because their grades both in English and in Swedish differ from the corresponding grades of the others.

In the example the distance measure was used for quantitative variables. The difference with LCA is probably due to this. It seems that the interpretation of grades as quantitatively different ‘stretches’ the distances between observations, and in this way it removes particularly Aki and Auli into their own class. If the grades were interpreted only qualitatively different, as is the case with LCA, Aki and Auli would be in a way ‘nearer’ Asko and Arja. If the data were a query by the Finnish Tourist Board, the stretching of the distances would have been absurd, and in that case LCA or some qualitative distance measure (cf. Everitt 1995) would have been the most relevant method.

Some concluding remarks

Kracauer based his critique of quantitative methods on a conception that condensing textual features to numerical code may break the essential structural features of the object studied. Not even LCA can escape this critique. In this sense LCA is like all the other methods that transform textual properties into numerical code. Presumably there is no other solution to this problem than developing the variables so that they as well as possible can tackle the text and its “texture of intimations and implications”. There has been some clear progress in this field after Kracauer’s days. It is also possible to think that recent text theoretical research could contribute to the practical questions concerning for example content analytical variables and their use.

The strengths of LCA are more evident on the level of the whole data than on the level of a single coding unit. LCA tries specially to find a structure of the data, and the classification of coding units is only a secondary task for it. LCA doesn’t place single coding units unambiguously to special places on latent continuums. It gives only probabilities of coding units to situate on certain locations. Here LCA differs for example classifications made with factor scores because factor scores give unambiguous latent co-ordinates for each coding unit. LCA and factor analysis/factor scores approach the structuration of data from different viewpoints, and they can give radically different results as for example in a study about teaching styles in 1970's and 1980's in England (Aitkin, Andersson & Hinde 1981). In this sense LCA seems to contribute something of its own.

Because LCA does not make computations with correlation coefficients, also nominal variables can be used. This is a very useful characteristics especially in content analysis, because content analytical variables are often nominal. LCA can be used also in surveys, although it is true that for example variables testing attitudes may be interpreted as quantitative, and this may argue for the use of methods based on correlation coefficients. LCA in its original form is indifferent for the amount of positiveness or negativeness in a reply, but interprets each category only qualitatively different.

With LCA it is possible to combine quantitative and qualitative analysis. When the latent classes have first been found out quantitatively, the most typical coding units in these classes can then be analysed qualitatively. In a way, LCA takes care that the items analysed qualitatively really are typical in the data. In other words, LCA gives information of how far the results of qualitative analysis can be generalised. When combining LCA with qualitative analysis, it is possible to benefit both from the generalizability of quantitative analysis and sensitivity of qualitative analysis. It is also possible, that the qualitative analyses of 'typical cases' can give ideas for new variables and, why not, for new LCA based on these variables. It might even be said that if the criteria of variable apparatus are met, LCA makes it possible to 'see a forest for the trees'.

All that has been said this far is true, beautiful and right. Still we seem to be ending to a slightly confusing conclusion about the relations between different classification methods. It seems now, that statistically rooted methods, like LCA and factor analysis, differ fundamentally from classification methods that are based more straightforwardly on manifest differences and similarities of observations.

Let us first consider a method based on manifest differences and similarities. Newspaper stories, for example, might be classified neatly into classes that consist of stories relatively similar to each other. It would also be easy to show which stories belong to which class. All the time it would be question of similarities of observations and of classes based on them (this is how we proceeded in our example of the use of distance measures). After this it is, then, possible to proceed with any kind of (quantitative or qualitative) methods.

Statistical methods like LCA, for their part, are based on axiom (the principle of local independence) that states how latent properties 'behave' when they are (for example by tests) forced into the 'observable world'. In this case, classification does not mean simply that similar observations are placed into the same group. Instead, classification is based on the assumption how coding units that share same latent properties will behave. In this case, different observations can easily be placed to same class, although it can't be said surely which coding unit belong to which class. Of course, the most typical cases of a class can be found if the most observations in the class are similar. These typical cases can then be analysed qualitatively. Qualitative analysis is more cautious if the class consists of different observations and there is about same amount of each type of observation.

From the point of view of non-statistical classification the statistically justified classes may look senseless or purely theoretical (as is the second class in table 2). From the statistical point of view, then, classifications based on pure observations may seem erroneous or at least superficial, because the latent structures explaining dependencies are not taken account of. The choice with the statistical or not so statistical starting point seem to be quite a philosophical dilemma and neither choice probably have an unambiguous justification.

References

Aitkin, Murray & Anderson, Dorothy & Hinde, John (1981) Statistical Modelling of Data on Teaching Styles. Journal of the Royal Statistical Society 144(1981):4.

Alkula, Tapani & Pöntinen, Seppo & Ylöstalo, Pekka (1994) Sosiaalitutkimuksen kvantitatiiviset menetelmät (Quantitative Methods in Social Research). Helsinki ja Juva: WSOY.

Berelson, Bernard (1952) Content Analysis in Communication Research.

Everitt, Brian S. (1995) Cluster Analysis. London, Sidney, Auckland: Arnold.

Everitt, Brian & Hay, Dale (1992) Talking about Statistics. A Psychologist's Guide to Data Analysis. London, Melbourne, Auckland: Edward Arnold.

Heinonen, Ari (1997) Sanomalehdistö ja Internet - toiveita, huolia, epätiotoisuutta (Internet and the Press - Hopes, Worries, Uncertainty). Journalismin tutkimuksen ja kehitystyön yksikön raportti. Tampereen yliopiston tiedotusopin laitoksen julkaisuja C 21.

Kempf, Wilhelm (1994) Towards an Integration of Quantitative and Qualitative Content Analysis in Propaganda Research. Diskussionsbeiträge Nr 27/1994 der Projektgruppe Friedensforschung Projekt 13/85, Universität Konstanz.

Kempf, Wilhelm & Luostarinen, Heikki (1997) New World Order Rhetorics. A Comparative Study of American and European Media During the Gulf War. Diskussionsbeiträge Nr 35/1997 der Projektgruppe Friedensforschung Projekt 13/85 & 590/95, Universität Konstanz.

Kempf, Wilhelm (in print): Escalating and deescalating aspects in the coverage of the Bosnian conflict - a comparative study. In: Kempf Wilhelm & Heikki Luostarinen (eds): Journalism in the New World Order. Volume II: Studying War and the Media. London: Sage.

Kracauer, S. (1990) Für eine qualitative Inhaltsanalyse. Teoksessa. Kracauer, Siegfried. Schriften. (Band 5). Frankfurt am Main: Suhrkamp.

Lazarsfeld, Paul F. (1950) Logical and Mathematical Foundations of Latent Structure Analysis. In Stouffer, Samuel A. & al (eds.) Measurement and Prediction. Princeton & New Jersey: Princeton University Press.

Lazarsfeld, Paul F. (1959) Latent Structure Analysis. In Koch, Sigmund (ed.) Psychology: A Study of a Science. Vol 3. Formulations of the Person and the Social Context. New York & Toronto & London: McGraw-Hill Book Company.

Luostarinen, Heikki (1986) Perivihollinen (The Ancient Foe). Tampere: Vastapaino

McCutcheon, Allan L. (1987) Latent Class Analysis. Sage University Paper series on Quantitative Applications in Social Sciences 64. Newbury Park: Sage.

Nohrstedt, S & Ottosen, R. (eds) (in print) Journalism in the New World Order. Volume I: Gulf War, National News Discourses and Globalization. London: Sage.

Pietilä, Veikko (1997) Joukkoviestinnän valtateillä (Mainstreams in Masscommunication Research). Tampere: Vastapaino.

Pietilä, Veikko (1970) Johdaltusta sisällönerittelyyn II (Introduction to Content Analysis). Tiedotusopin laitoksen opetusmoniste. Tampereen yliopisto

Ranta, Esa & Rita, Hannu & Kouki, Jari (1991) Biometria. Tilastotiedettä ekologeille. (Biometrics. Statistics for Ecologists). Helsinki: Yliopistopaino.

Sclove, Stanely L. (1987) Application of Model-Selection Criteria to Some Problems in Multivariate Analysis. Psychometrica 52(1987):3.

Suhonen, Pertti (1994) Mediat, me ja ympäristö (Media, Us and Environment). Helsinki: Hanki ja jää.

Notes

[1] Whatever Lazarsfeld said, his desire has been to make classifications and tables. In describing the differences between quantitative and ”qualitative” approaches Berelson uses carefully produced tables and classifications. When he describes the history of content analytical reseach, the presentation is based on a table which shows the quantitative distribution of content analytical studies in periods of five years. When he comments the field of application of content analysis Lazarsfeld states at first that it is not easy to make a classification of the different forms of content analysis. ”Seventeen different forms (or applications or functions) of content analysis have been, however, specified for description and reflection” (Berelson 1952, 26).

[2] By giving examples how content analysis have been used, Berelson however gives following examles: What are the most important metaphors in Shakespear’s dramas? How the personality of an author leaves its’ trace on what he or she writes. How the ethnic minorities are described in the popular press fiction? One could continue the list..

[3] The groupings are based on mathematical probabilites of coding patterns, not on groupings of real coding units. This is why real coding units can’t always be grouped in such classes as Lacord proposes. However, the distributions of categories match with the data in all solutions proposed by Lacord.

[4] In solutions with more than one classes it must be taken account of the number and size of classes. For example in a two class -solution the probability of coding pattern100 can be obtained by computing first the probabilities in both classes separately:

[5] The values of LOG-Like -index are here slightly different from corresponding values in tables 4 and 5 probabli because differences in the number of decimals used in computations.

[6] Sometimes principal component analysis is thought to be an indepenent method that resembles factor analysis but still differs clearly from it (for example Ranta & Rita & Kouki 1991). Some other times principal component analysis is thought to be more like one form of factor analysis (for example Everitt & Hay 1992, 112-114). The difference between the methods is, that the proper factor analysis observes the covariance structure of the variables while principal component analysis tries to explain the variance generally (Ranta & Rita & Kouki 1991, 459).

[7] Nowadays there are also applications of LCA for ordinate variables. In fact, Lacord-program have an optional models for this kind of analysis. Here we restrict our consideration on models that are intended for analysis of nominal data.

[8] The indexes for LCA-solutions:

Numb. of cl.	LOG-Like	Npar	DF	AIC	CIC
1	-35.963	9	54	89.926	79.432
2	-22.503	19	44	**83.005**	**60.853**
3	-18.867	29	34	95.734	61.921
4	-16.958	39	24	111.915	66.443
5	-16.958	49	14	131.915	74.784
6	-16.958	59	4	151.915	83.124
Satur. model	-16.957				

[9] Factor analysis is, though. possible with nominal variables, but in this cas the variables must first be dichotmized. In other words, each category must be transformed to a separate variable.