Qualitative Data Analysis

(version 0.5, 1/4/05)

Code: analysis-quali

Daniel K. Schneider, TECFA, University of Geneva

Menu

1. Introduction: classify, code and index 2
2. Codes and categories 3
3. Code-book creation and management 6
4. Descriptive matrices and graphics 10
5. Techniques to hunt correlations 17
6. Typology and causality graphs 21

1. Introduction: classify, code and index

☞ Coding and indexing is necessary for systematic data analysis.

☞ Information coding allows to identify variables and values, therefore

- allows for systematic analysis of data (and therefore reliability)
- ensures enhanced construction validity, i.e. that you look at things allowing to measure your concepts

Before we start: Keep your documents and ideas safe !

Write memos (conservation of your thoughts)

- if is useful to write short memos (vignettes) when an interesting idea pops up, when you looked at something and want to remember your thoughts

Write contact sheets to allow remembering and finding things

After each contact (telephone, interviews, observations, etc.), make a short data sheet

- Indexed by a clear filename or tag on paper, e.g. CONTACT_senteni_2005_3_25.doc
- type of contact, date, place, and a link to the interview notes, transcripts.
- principal topics discussed and research variables addressed (or pointer to the interview sheet)
- initial interpretative remarks, new speculations, things to discuss next time

Index your interview notes

- Put your transcription (or tapes) in a safe place
- Assign a code to each "text", e.g. INT-1 or INTERVIEW_senteni_3_28-1
- You also may insert the contact sheet (see above)
- number pages !

2. Codes and categories


A code is a “label” to tag a variable (concept) and/or a value found in a "text"

Basics:

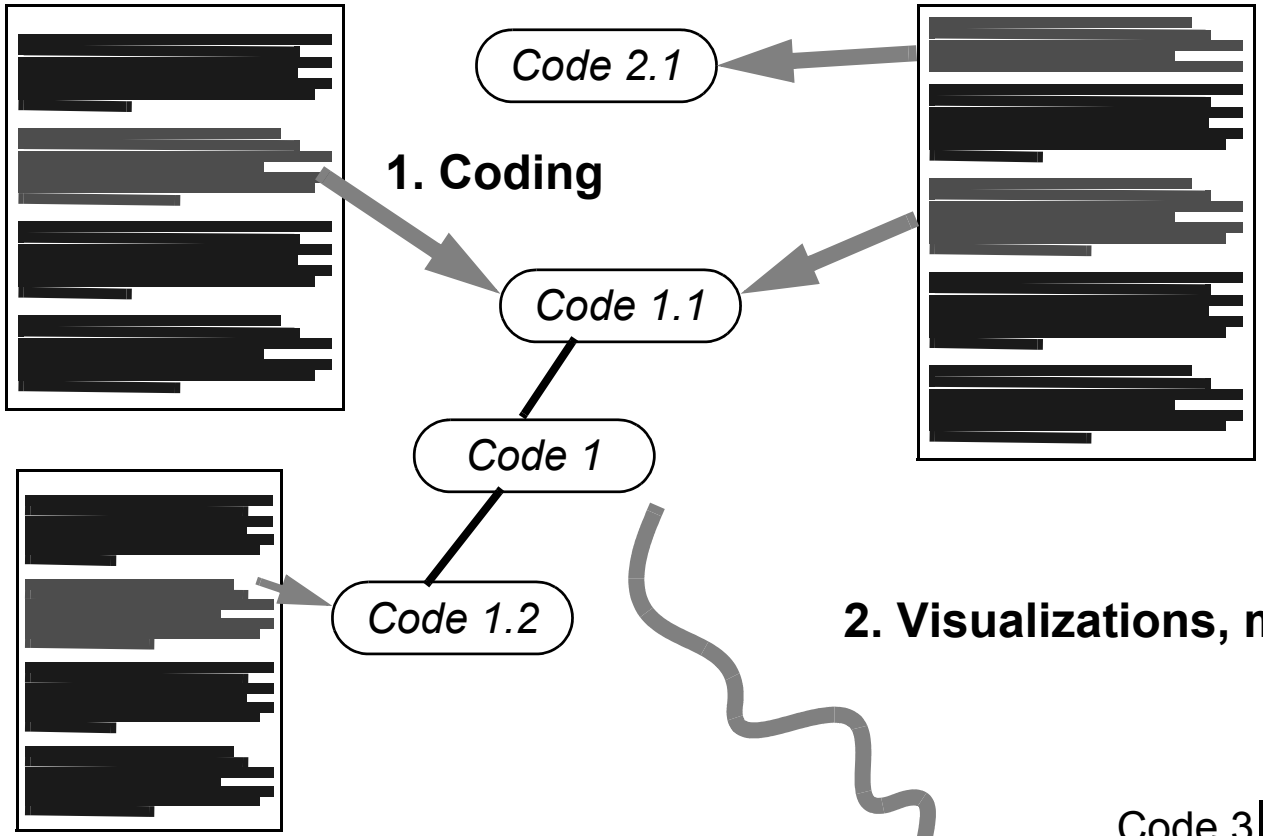
1. A code is assigned to each (sub)category you work on
   - In other words: you must identify ***variable names***
2. In addition, you can for each code assign a set of possible values (e.g.: “positive”/”neutral/ ”negative)
3. You then will systematically scan all your texts (documents, interview transcripts, dialogue captures, etc.) and tag all occurrences of variables.

- Three very different coding strategies exist
  - 3.1 “Code-book creation according to theory” [6]
  - 3.2 “Coding by induction (according to “grounded theory”)” [7]
  - 3.3 “Coding by ontological categories” [8]
-

Benefit

- Coding will allow you to find all informations regarding variables of interest to your research
- Reliability will be improved

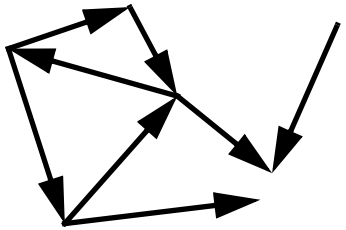
2.1 The procedure with a picture



2. Visualizations, matrices and "grammars"

Code 3	Code 1.1	Code 4
Val 1	Val 2x	Val 3y
	.....	

3. Analysis



2.2 Technical Aspects

- The safest way to code is to use specialized software
  - e.g. Atlas or Nvivo (NuDist),
  - however, this takes a lot of time !
- For a smaller piece (of type master), we suggest to simply tag the text on paper
  - you can make a reduced ***photocopy*** of the texts to gain some space in the margins
  - overline or circle the text elements you can match to a variable
  - make sure to distinguish between codes and other marks you may leave.
- Don’t use "flat" and long code-books, introduce hierarchy (according to dimensions identified)
- Each code should be short but also mnemonic (optimize)
  - e.g. to code according to a schema “principal category” - “sub-category” (“value”):

```
use:         CE-CLIM(+)
instead of: external_context -climate (positive)
```

- Don’t start coding before you have good idea on your coding strategy !
  - either your code book is determined by you research questions and associated theories, frameworks, analysis grids
  - or you really learn how to use an inductive strategy like "grounded theory".

3. Code-book creation and management

3.1 Code-book creation according to theory

☞ The list of variables (and their codes), is defined by theoretical reasoning, e.g.

- analytical frameworks, analysis grids
- concepts found in the list of research questions and/or hypothesis

Example from an innovation study (about 100 codes):

	categories	codes	theoretical references
properties of the innovation		PI	
external context		CE	
	demography	CE-D	
	support for the reform	CE-S	
internal context		CI	.... (fill for your own code book) .....
adoption processes		PA	
	official chronology	PA-CO	
dynamics of the studied site		DS	
external and internal assistance		AEI	
causal links		LC	

3.2 Coding by induction (according to “grounded theory”)

Principle:

- The researcher starts by coding a small data set and then increases the sample in function of emerging theoretical questions
- Categories (codes) can be revised at any time

Starting point = 4 big abstract observation categories:

- conditions (causes of a perceived phenomenon)
- interactions between actors
- strategies and tactics used by actors
- consequences of actions

(... many more details: to use this approach you ***really*** must document yourself)

3.3 Coding by ontological categories

Example:

Types	
Context/Situation	information on the context
Definition of the situation	interpretation of the analyzed situation by people
Perspectives	global views of the situation
Ways to look at people and objects	detailed perceptions of certain elements
Processes	sequences of events, flow, transitions, turning points, etc.
Activities	structures of regular behaviors
Events	specific activities (non regular ones)
Strategies	ways of tackling a problem (strategies, methods, techniques)
Relations and social structure	informal links
Methods	comments (annotations) of the researcher

- This is a compromise between “grounded theory” and “theory driven” approaches

3.4 Pattern codes

- Some researchers also code patterns (relationships)

Simple encoding (above) breaks data down to atoms, categories)

"***pattern coding***" identifies relationships between atoms.

The ultimate goal is to detect (and code) regularities, but also variations and singularities.

Some suggested operations:

1. Detection of ***co-presence*** between two values of two variables
   - E.g. people in favor of a new technology (e.g. ICT in the classroom) have a tendency to use it.
2. Detection of ***exceptions***
   - e.g. technology-friendly teachers who don't use it in the classroom
   - In this case you may introduce new variable to explain the exception, e.g. the attitude of the superior., of the group culture, the administration, etc.
   - Exceptions also may provoke a change of analysis level (e.g. from individual to organization)

Attention: a co-presence does not prove causality

4. Descriptive matrices and graphics

♡ Qualitative analysis attempts to put structure to data
(as exploratory quantitative techniques)
In short: ***Analysis = visualization***

☞ **2 types of analyses:**

1. A ***matrix*** is a tabulation engaging at least one variable, e.g.
   - Tabulations of central variables by case (equivalent to simple descriptive statistics like histograms)
   - Crosstabulations allowing to analyze how 2 variables interact
2. Graphs (***networks***) allow to visualize links:
   - temporal links between events
   - causal links between several variables
   - etc.

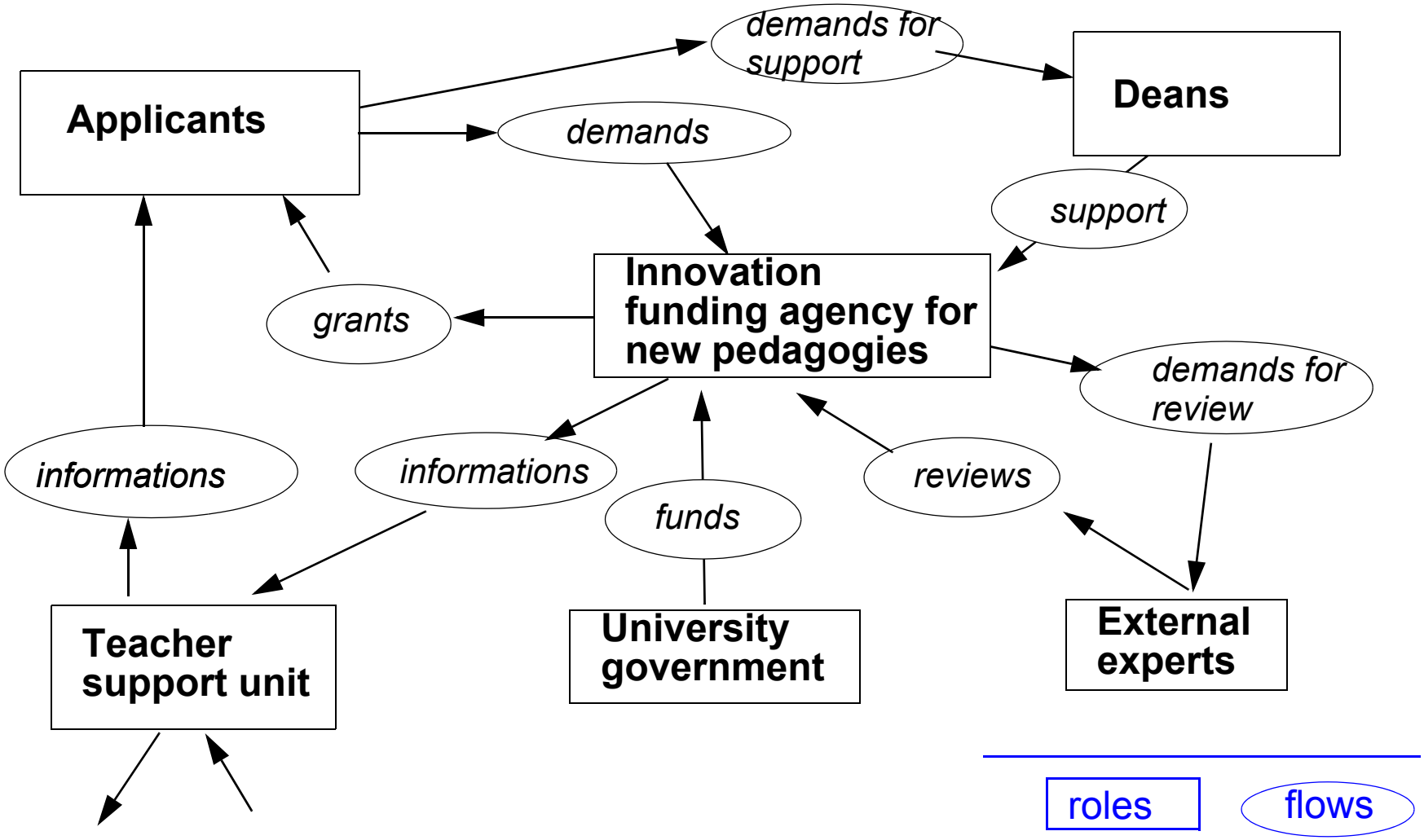
☞ Some advice:

- when use these techniques always keep a link to the source (coded data)
- try to fit each matrix or graph on a ***single page*** (or make sure that you can print things made by computer on a A3 pages)
- you have to favor synthetic vision, but still preserve enough detail to make your artifact interpretable
- Consult specialized manuals e.g. Miles & Huberman, 1994 for recipes or get inspirations from qualitative research in the same domain

4.1 The “context chart”,Miles & Huberman (1994:102)

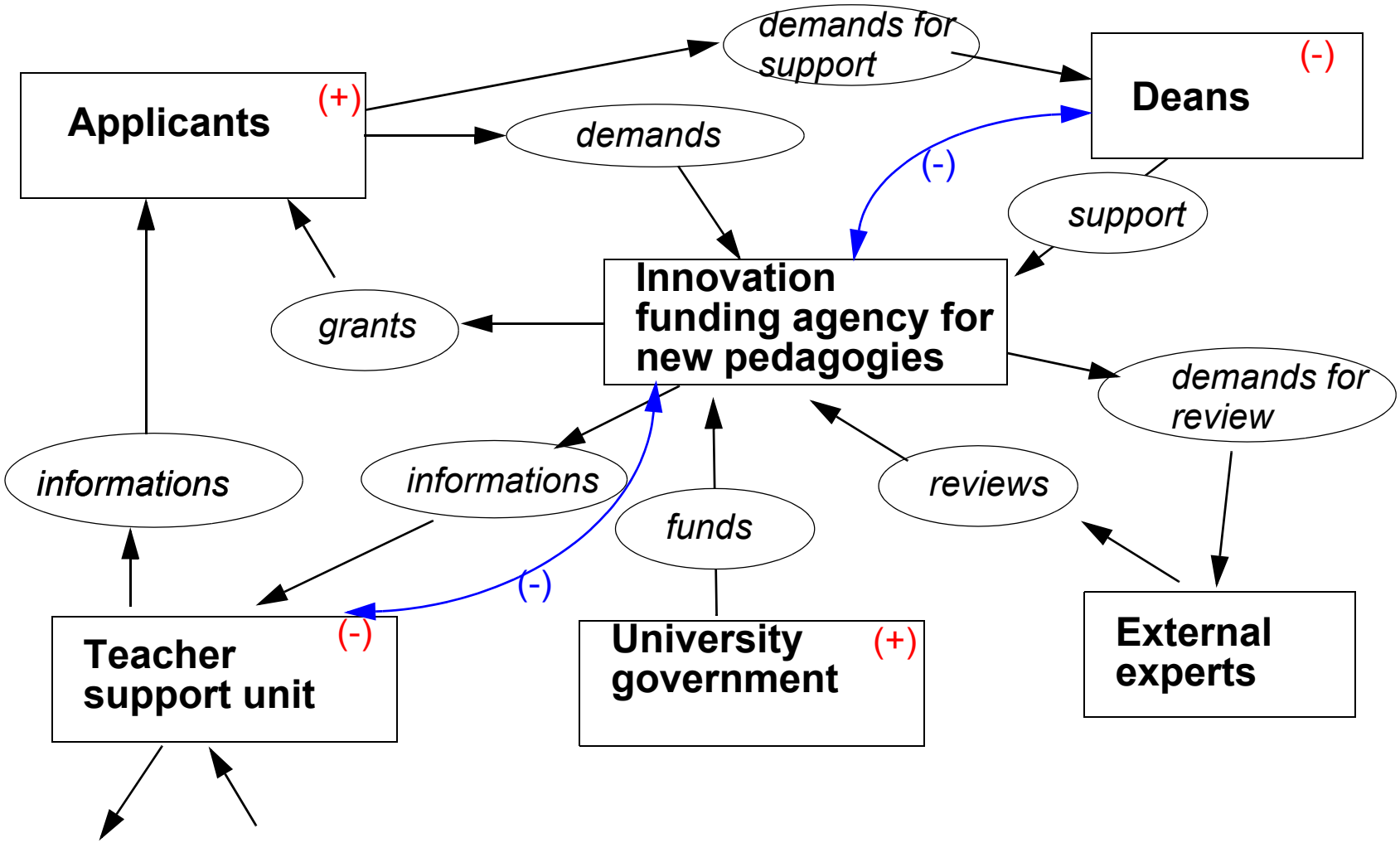
☞ Allows to visualize relations and information flows between rôles and groups

Exemple 4-1: Work flow for a "new pedagogies" program at some university



- There exist codified "languages" for this type of analysis, e.g. UML or OSSAD

Once you have clearly identifed and clarified formal relations, you can use the graph to make annotations (like below)



(+) (-) () positive or negative attitudes towards a legal program

\+ - good or bad relations between authorities (or people)

4.2 Check-lists, Miles & Huberman (1994:105)

☞ Usage: Detailed summary for an analysis of an important variable

Example: “external support is important for succeeding a reform project

Examples for external support	At counselor level	At teacher level
Analysis of deficiencies	**Fill in each cell as below**	
Teaching training		
Change monitoring		
Incentives		
Group dynamics	adequate: “we have met an organizer 3 times and it has helped us” (ENT-12:10)	not adequate: “we just have informed” (ENT-13:20)
etc. ..		

- such a table displays various dimensions of and important variable (external support), e.g. in the example = left column
- in the other columns we insert summarized ***facts*** as reported by different roles.
- Question: Imagine how you would build such a grid to summarize teacher’s, student’s and assistant’s opinion about technical support for an e-learning platform

4.3 Chronological tables Miles & Huberman (1994:110)

- Can summarize a studied object's most important events in time

Exemple 4-2: Task assignments for a blended project-oriented class

	Activity	Date	imposed tools (products)
1	Get familiar with the subject	21-NOV-2002	links, wiki, blog
2	project ideas, Q&R	29-NOV-2002	classroom
3	Students formulate project ideas	02-DEC-2002	news engine, blog
4	Start project definition	05-DEC-2002	ePBL, blog
5	Finish provisional research plan	06-DEC-2002	ePBL, blog
6	Finish research plan	11-DEC-2002	ePBL, blog
7	Sharing	17-DEC-2002	links, blog, annotation
8	audit	20-DEC-2002	ePBL, blog
9	audit	10-JAN-2003	ePBL, blog
10	Finish paper and product	16-JAN-2003	ePBL, blog
11	Presentation of work	16-JAN-2003	classroom

- This type of table is useful to identify important events.
- You can add other information, e.g. tools used in this example

4.4 Matrices for roles (function in an organization or program)

Miles & Huberman (1994:124)

Crossing social roles with one or more variables, abstract example (also see next page):

rôles	persons	variable 1	variable 2	variable 3
rôle 1	person 1	cells are filled in with values (pointing to the source)		
	person 2			
	.....			
rôle 2	person 9			
	person 10			
.....	.....			
rôle n	person n			
	.....			

Crossing roles with roles

s

	rôle 1	...	rôle 3
rôle 1	fill in all sorts of informations about interactions		
...			
rôle 3			

Example: Evaluation of the implementation of a help desk software

Actor	Evalu-ation	assistance given	Assistance received	Immediate effects	Long term effects	Explanation of the researcher
Manager	-	-	-	demotivating	threatened the program	Felt threatened by new procedures
Consult-ant	+	help choosing the right soft. involved himself	-	contributed to the start of the experiment	-	....
"Help-desk worker"	+/-	debugging of machines, little help with software		better job satisfaction because of the tool	slight improvement of throughput	is still overloaded with work
Users	+/-	A few users provided help to peers with the tool	debugging of machines, little help with software	Were made aware of the high amount of unanswered questions	slight improvement of work performance	....

Crossing between roles to visualize relations:

	rôle 1	rôle 1	rôle 3
rôle 1			
trainers		"don't coordinate very much" (1)	doesn't receive all the information (2)
rôle 3			

5. Techniques to hunt correlations

5.1 Matrices ordered according to concepts (variables)

A. Clusters (co-variances of variables, case typologies)

- An idea that certain values should "go together": Hunt co-occurrences in cells
- E.g.: “Can we observe a correlation between expressed ***needs for support*** and expressed ***needs for training*** for a new collaborative platform (data from teachers’s interviews)?

case	var 1	need for support	need for training	need for directives
case 1		important	important	important
case 2		not important	not important	not important
case 3		important	important	important
case 4	yyy	not important	not important	not important
case 5	.....	important	important	important
case 6....		important	not important	not important

- This table shows e.g. that nedd for support and need for training seem to go together, e.g. cases 1,3,5 have association of "important", cases 2 and 4 have association of "not important".
- See next page how we can summarize this sort of information in a crosstab

B. Co-variance expressed in a corresponding crosstab:

training needs * support needs		need for support	
		yes	no
need for training	yes	3	1
	no	1	2

- .... we can observer a correlation here: "blue cells" (symmetry) is stronger than "magenta"!
- check with the data on last slide

C. Example typology with the same data:

	Type 1: "anxious"	Type 2: "dependent"	Type 3: "bureaucrats"	Type 4: "autonomists"
case 1	X			
case 2				X
case 3	X			
case 4				X
case 5	X			
case 6		X		
Total	3	1	0	2

- we can observe emergence of 3 types to which we assign "labels"
- Note: for more than 3 variables use a cluster analysis program

Additional example

The table shows co-occurrence between values of 2 variables. The idea is to find out what effect different types of pressure have on ICT strategies adopted by a school.

	Strategies of a school				
Type of pressure	strategy 1: no reaction	strategy 2: a task force is created	strategy 3: internal training programs are created	strategy 4: resources are reallocated	strat 5: .....
Letters written by parents	(N=4) (p=0.8)	(N=1) (p=0.2)			
Letters written by supervisory boards		(N=2) (p=0.4)	(N=3) (p=0.6)		
newspaper articles				(N=1) (p=100%)	
type ...	.....	....			

D. Recall: Interpretation of crosstabulation

Procedure

- calculate the % for each value of the independent variable
  - Note: this can be either the line or the column depending on how you orient your table
- compute the % in the other direction
- We would like to estimate the probability that a given value of the independent (explaining) variable entails a given value of the dependent (explained) variable

	Variable y to explain = Strategies of action			
Explaining variable x	**do nothing**	**send a mail**	**write a short tutorial**	**Total**
Students making indirect suggestion	4 (80%)	1 (20%)		5 (100 %)
Students explicitly complaining		2 (40%)	3 (60%)	5 (100%)

Interpretation: “... if students explicitly complain, the tutor will react more strongly and engage in more helpful acitities.”

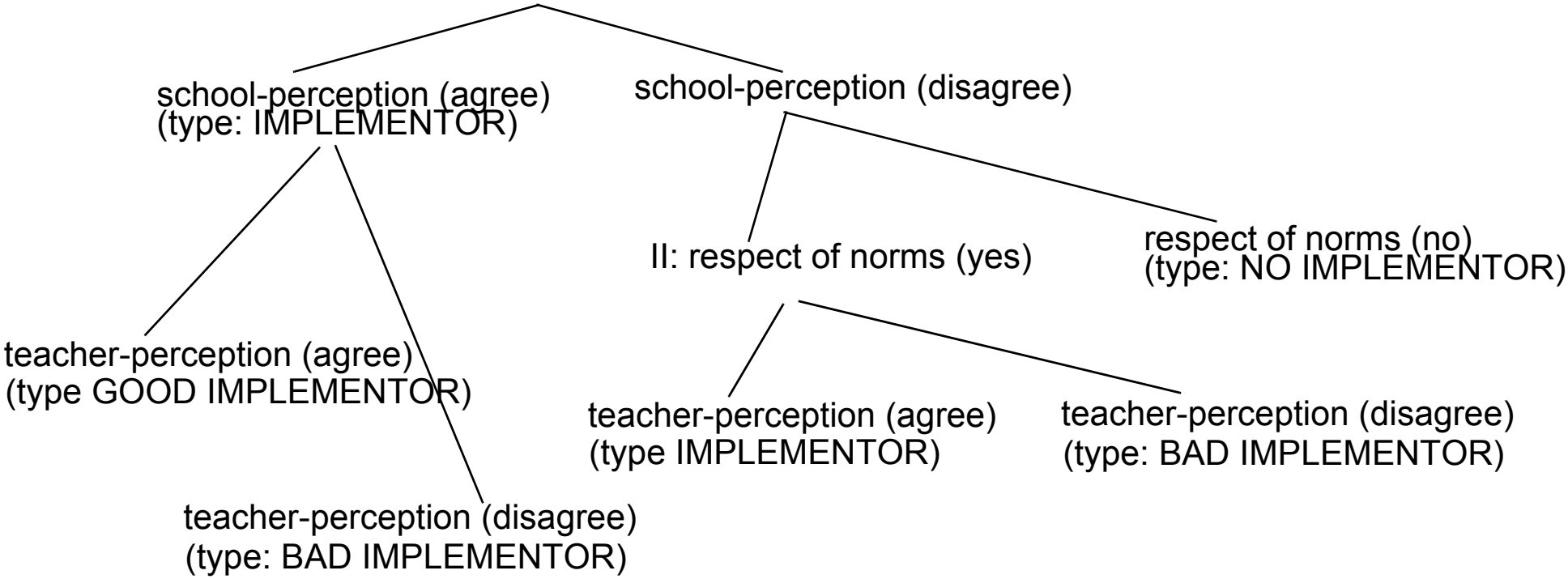
- See also: quantitative data analysis.

6. Typology and causality graphs

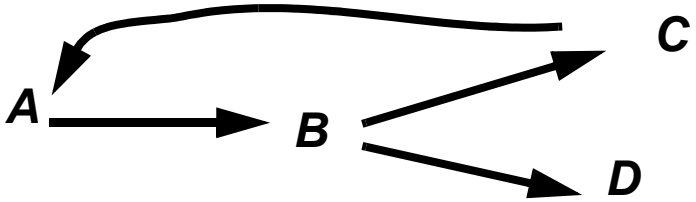
6.1 Typology graphs

- Display attributes of types in a tree-based manner

Exemple 6-1: Perception of a new program by different implementation agencies (e.g. schools) and its actors (e.g. teachers)

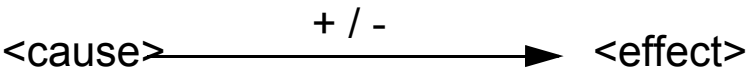


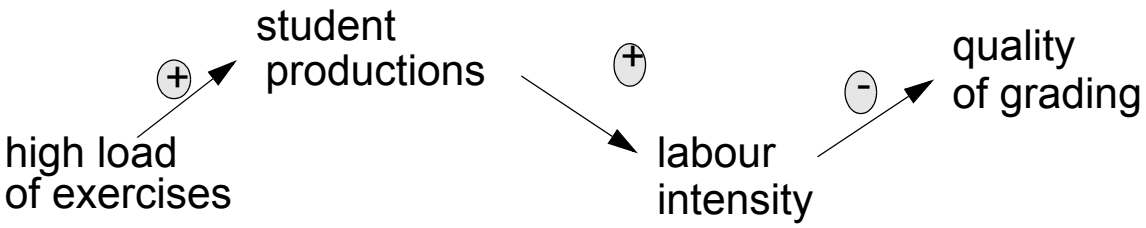
6.2 Subjective causality graphs



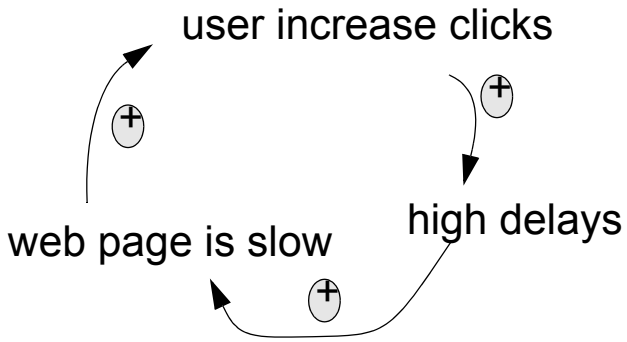
- Cognitive maps à la “operational coding”, AXELROD, 1976
- Allow to compute outcomes of reasoning chains
- Example: Teacher talking about active pedagogies, ICT connections, Forums

About active pedagogies:





About slow ICT connections:



About forum management: