

Doing quantitative content analysis

Unitisation
Coding Reliability Validity

Jan-Willem Strijbos

JURE, Leuven, 10 July 2008



In quantitative content analysis the communication is segmented (if this is applicable), coded and summarised, and frequencies/ percentages or patterns are used for comparisons and/ or statistical testing.

Research goal

Collaborative setting

Unit is influenced by ...

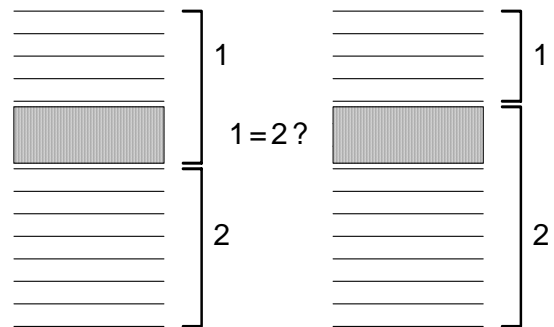
Tool

Unit boundary overlap

Nature of communication

Coder A

Coder B



Split compound sentences

'Meaning' of components

Alternative unit

Punctuation
(...) or ...
Colon
Semi-colon
'and'
Comma
Dash
Brackets

Segmentation examples

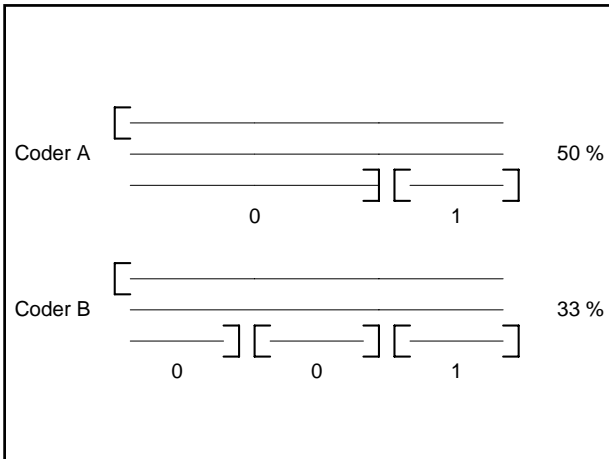
[I agree that you will start with data collection and commitment,]

[(however) we have to decide first whether we will use the PERS method or the method by Hoppe.]

This sentence is segmented after the comma. Both parts can be regarded as a 'meaningful' sentence in itself.

[If John has completed the PERS analysis, I can start with the analysis of public support.]

The first part of this sentence before the comma cannot be regarded as a 'meaningful' sentence in itself. This is an example of a 'conditional clause'. Examples of indicators are 'if', 'since', 'before', 'as long as', 'in case', 'because' and 'given'. This sentence is NOT segmented.



Segmentation reliability

Cross-validation

Segmentation **.79 - .85 [.82]** **.87 - .91 [.89]**

.85 - .88 [.87]

Multiple trials

Proportion agreement

0 < .70 < .80 < 100



Coding categories

Number of codes **Data-driven?**

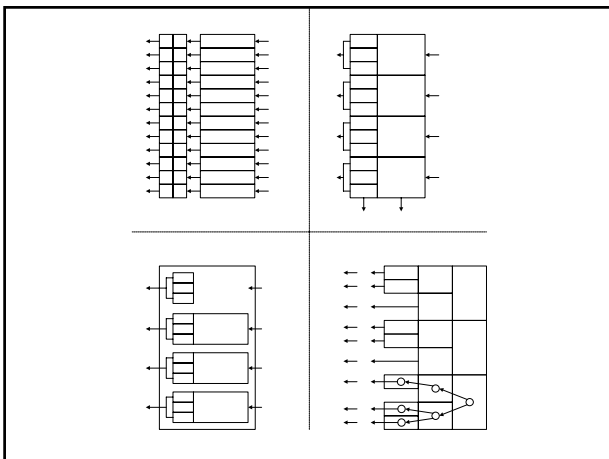
Rules **Theory-driven?**

Mutually exclusive

Iterative development cycles

10% of all data

Include all research conditions



Single construct scheme

Type	Example
Task coordination	"Why is Ron not responding?" "Please give your ideas." "Who will make an inventory of pressure groups involved?"
Task content	"The assignment is about the public transport in Amsterdam." "We should delete section two and check for errors in three."
Task social	"I think we as group did a great job in a virtual project team." "John Doe, my compliments for your PERS analysis."
Non-task	"I am still struggling to find out how I am to operate Edubox." "How was your holiday in Lebanon, Syria and Jordania?"
Non-codable	"Attached a new schedule with the latest deadlines and tasks."

Multiple trials **Kappa**

$0 < .45 < .60 < .75 < 1$

Alpha

Coding reliability

Coding of statements

.69, .70, .68 [Kappa = .70]

Chat

Small groups

PowWow!

Dimensions

Finding the Perimeter of an Octagon

Given the regular octagon below, answer the following questions:

1. What's wrong with this picture?
2. If you fix what's wrong, what's the perimeter of the octagon?

Extra: Assume that the thing you found to be wrong is actually right. What else could you change to make things right? What's the resulting perimeter of the octagon?

Threading of ... **Conversation?**

Problem solving?

Reliability

Unit = chat line **Valid units?**

Coding

Dimensions

Conversational thread
Conversational dimension
Social dimension
Problem solving thread
Problem solving

	Coder B	
Coder A	Thread	No thread
Thread	Same	Different
No thread		

	Coder A	Coder B
1	A	?
2	B	A
3	?	A
4	C	C
5	B	C
6	D	D
7	?	?
8	?	?
9	A	A
10	C	?

	Coder A	Coder B
1	A	?
2	B	A
3	?	A
4	C	C
5	B	C
6	D	D
7	?	?
8	?	?
9	A	A
10	C	?

	Coder A	Coder B
1	A	?
2	B	A
3	?	A
4	C	C
5	B	C
6	D	D
7	?	?
8	?	?
9	A	A
10	C	?

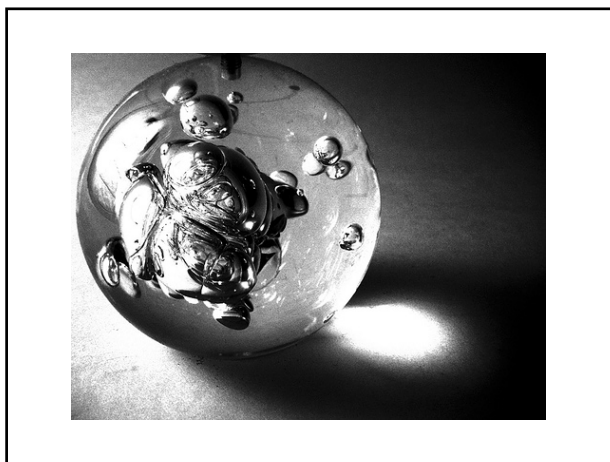
	Coder A	Coder B
1	A	?
2	B	A
3	?	A
4	C	C
5	B	C
6	D	D
7	?	?
8	?	?
9	A	A
10	C	?

	Coder A	Coder B
1	A	X
2	B	A
3	X	A
4	C	C
5	B	C
6	D	D
9	A	A
10	C	X

	Coder A	Coder B
1	A	X
2	B	A
3	X	A
4	C	C
5	B	C
6	D	D
7	X	X
8	X	X
9	A	A
10	C	X

32								
32	AME	I have an idea that might help us find whats wrong with the pic.	7:06:19					
33	MCP	We could use good of Pythag thm to see what BV is	7:06:30	0:00:49	s	is		
34	AME	Lets not	7:06:40	0:00:11	o	cg	s	rf
35	MCP	Whats your idea?	7:06:46	0:00:06	33	d	cg	33
36	AME	It states that something is wrong with the pic.	7:07:01		32	rq	ci	32
37	AME	so we can't find what BV is	7:07:08	0:00:15	35	e	cg	35
38	MCP	Yeah, and I think if we 'found' BV, it would be something not possible.	7:07:31	0:00:07	36	el	cg	36
39	MCP	16 + BV^2 = 21.16	7:08:10	0:00:23	37	o	cg	37
40	MCP	BV^2 = 5.16	7:08:20	0:00:39	38	o	cg	38
41	AME	I got it	7:08:23	0:00:10	39	el	cg	39
42	AME	I know whats wrong with the pic	7:08:29	0:00:03	41	se		
43	MCP	BV = 2.27	7:08:31	0:00:06	41	s	is	
44	FIR	ok. now i'm following!	7:08:44	0:00:02	39	el	cg	39
				0:00:13	39	f	ci	39

59	AME	What did you say BV was?	7:10:51	0:00:08	58	rq	ci	43	
60	FIR	2.27	7:11:05	0:00:14	59	rp		59,43	re
61	MCP	With the numbers given, BV would be	7:11:10	0:00:05	59	rp		59	
62	MCP	yeah	7:11:11	0:00:01	60	f		60	
63	AME	I think thats wrong	7:11:14	0:00:03	60	d	ci	60	ch
64	FIR	how so?	7:11:19	0:00:05	63	rq		63	
65	AME	I know whats wrong with the pic	7:11:28	0:00:09		s	is		
66	MCP	base would be twice that	7:11:31	0:00:03	61	o		61	r
67	FIR	what	7:11:33	0:00:02	65	rq			
68	MCP	4.54 ish	7:11:41	0:00:08	66	et		66	
69	AME	The diagonl is not 4.6	7:11:45	0:00:04	67	e		67	t
70	MCP	Right	7:11:51	0:00:06	69	f		69	
71	FIR	exactly	7:12:02	0:00:11	69	a		69	ch
72	MCP	Otherwise, the red lines and the base are almost an equilateral triangle	7:12:14	0:00:12	70	o		69	rf



Number of codes **Data-driven?**

Rules **Theory-driven?**

Mutually exclusive

Validity

Iterative development cycles

10% of all data

Include all research conditions

Messick (1989, 1995)	Rourke and Anderson (2004)
Content aspect	Purpose of the coded data
Substantial aspect	Behaviors representing the construct
Structural aspect	Reviewing the codes and indicators
	Holding preliminary try-outs
	Guidelines for administration/ scoring
Consequential aspect	
Generalisability aspect	
External aspect	

Re-use **De Wever et al., 2006**

Andriessen et al., 2007

Generalisability

Same data, two similar coding schemes
(external aspect)

Schellens & Valcke, 2005

1. Determine unit of analysis
2. Develop segmentation procedure
3. Reliability of segmentation

Doing quantitative CA

4. Re-use, adapt or develop codes
5. Assess validity & establish rules
6. Reliability and validity of coding



JWstrijbos@FSW.leidenuniv.nl