

Coherence relations and DRD identification: theory and analysis

Ted Sanders (Utrecht University)

Wilbert Spooren (Radboud University Nijmegen)

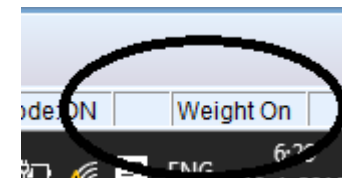
Testing patterns in the coprus analysis of crs and drd's:

Statistical methods: exercise loglinear analysis

Open the file “Omdat want written spoken.sav”

Analysis omdat want written spoken.sav [DataSet1] - IBM SPSS Statistics Data Editor

	Genre	DRD	Relation	Freq	var
1	written	want	non-volitional	1	
2	written	want	volitional	7	
3	written	want	epistemic	38	
4	written	want	speech act	4	
5	written	omdat	non-volitional	10	
6	written	omdat	volitional	16	
7	written	omdat	epistemic	24	
8	written	omdat	speech act	0	
9	spoken	want	non-volitional	4	
10	spoken	want	volitional	57	
11	spoken	want	epistemic	56	
12	spoken	want	speech act	26	
13	spoken	omdat	non-volitional	19	
14	spoken	omdat	volitional	87	
15	spoken	omdat	epistemic	8	
16	spoken	omdat	speech act	3	
17					



Make crosstabs of Relation * Marking, separate layers of Genre; ask for observed and expected frequencies

Analysis omdat want spoken.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction

Frequencies...
Descriptives...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...
Q-Q Plots...

	Genre	DRD
1	written	want
2	written	want
3	written	want
4	written	want
5	written	omdat
6	written	omdat
7	written	omdat
8	written	omdat
9	spoken	want

Crosstabs: Cell Display

Counts
☒ Observed
☒ Expected
☐ Hide small counts
Less than 5

z-test
☐ Compare column proportions
☐ Adjust p-values (Bonferroni method)

Percentages
☐ Row
☐ Column
☐ Total

Residuals
☐ Unstandardized
☐ Standardized
☐ Adjusted standardized

Noninteger Weights
☒ Round cell counts
☐ Round case weights
☐ Truncate cell counts
☐ Truncate case weights
☐ No adjustments

Continue Cancel Help

Crosstabs

Row(s):
Relation

Column(s):
DRD

Layer 1 of 1
Previous Next
Genre

☐ Display clustered bar charts
☐ Suppress tables

Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...

OK Paste Reset Cancel Help

Check the number and size of expected frequencies

Relation * DRD * Genre Crosstabulation

Genre				DRD		Total
				want	omdat	
written	Relation non-volitional	Count		1	10	11
		Expected Count		5,5	5,5	11,0
	volitional	Count		7	16	23
		Expected Count		11,5	11,5	23,0
	epistemic	Count		38	24	62
		Expected Count		31,0	31,0	62,0
	speech act	Count		4	0	4
		Expected Count		2,0	2,0	4,0
	Total	Count		50	50	100
		Expected Count		50,0	50,0	100,0
spoken	Relation non-volitional	Count		4	19	23
		Expected Count		12,7	10,4	23,0
	volitional	Count		57	87	144
		Expected Count		79,2	64,8	144,0
	epistemic	Count		56	8	64
		Expected Count		35,2	28,8	64,0
	speech act	Count		26	3	29
		Expected Count		16,0	13,1	29,0
	Total	Count		143	117	260
		Expected Count		143,0	117,0	260,0
Total	Relation non-volitional	Count		5	29	34

Check the number and size of expected frequencies

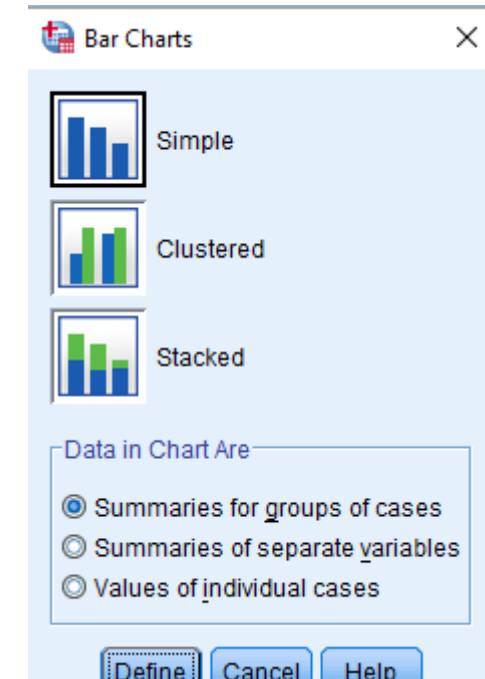
Relation * DRD * Genre Crosstabulation

Genre				DRD		Total
				want	omdat	
written	Relation non-volitional	Count		1	10	11
		Expected Count		5,5	5,5	11,0
	volitional	Count		7	16	23
		Expected Count		11,5	11,5	23,0
	epistemic	Count		38	24	62
		Expected Count		31,0	31,0	62,0
	speech act	Count		4	0	4
		Expected Count		2,0	2,0	4,0
	Total	Count		50	50	100
		Expected Count		50,0	50,0	100,0
spoken	Relation non-volitional	Count		4	19	23
		Expected Count		12,7	10,4	23,0
	volitional	Count		57	87	144
		Expected Count		79,2	64,8	144,0
	epistemic	Count		56	8	64
		Expected Count		35,2	28,8	64,0
	speech act	Count		26	3	29
		Expected Count		16,0	13,1	29,0
	Total	Count		143	117	260
		Expected Count		143,0	117,0	260,0
Total	Relation non-volitional	Count		5	20	25

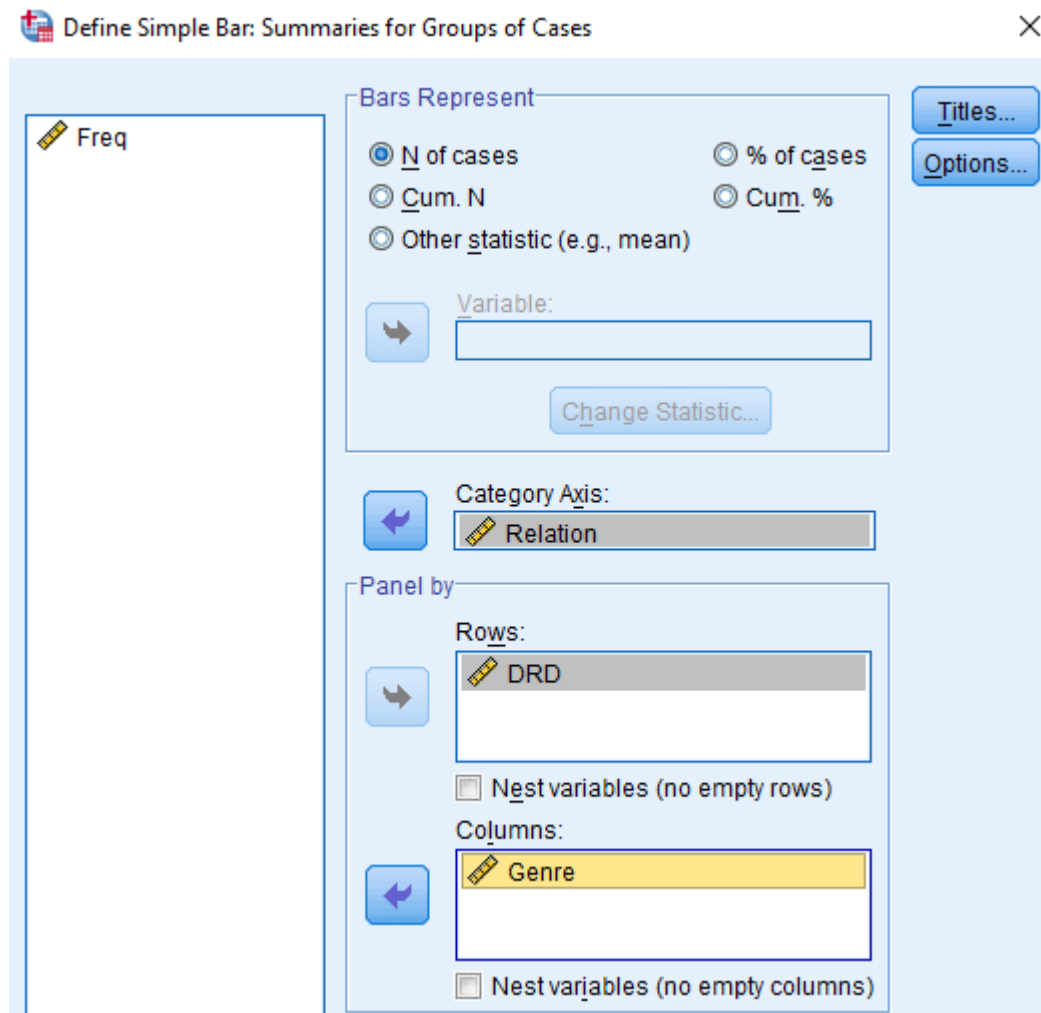
Make graphs of Relation (category axis), DRD (row panel), Genre (column panel)

Analysis omdat want spoken.sav [DataSet1] - IBM SPSS Statistics Data Editor

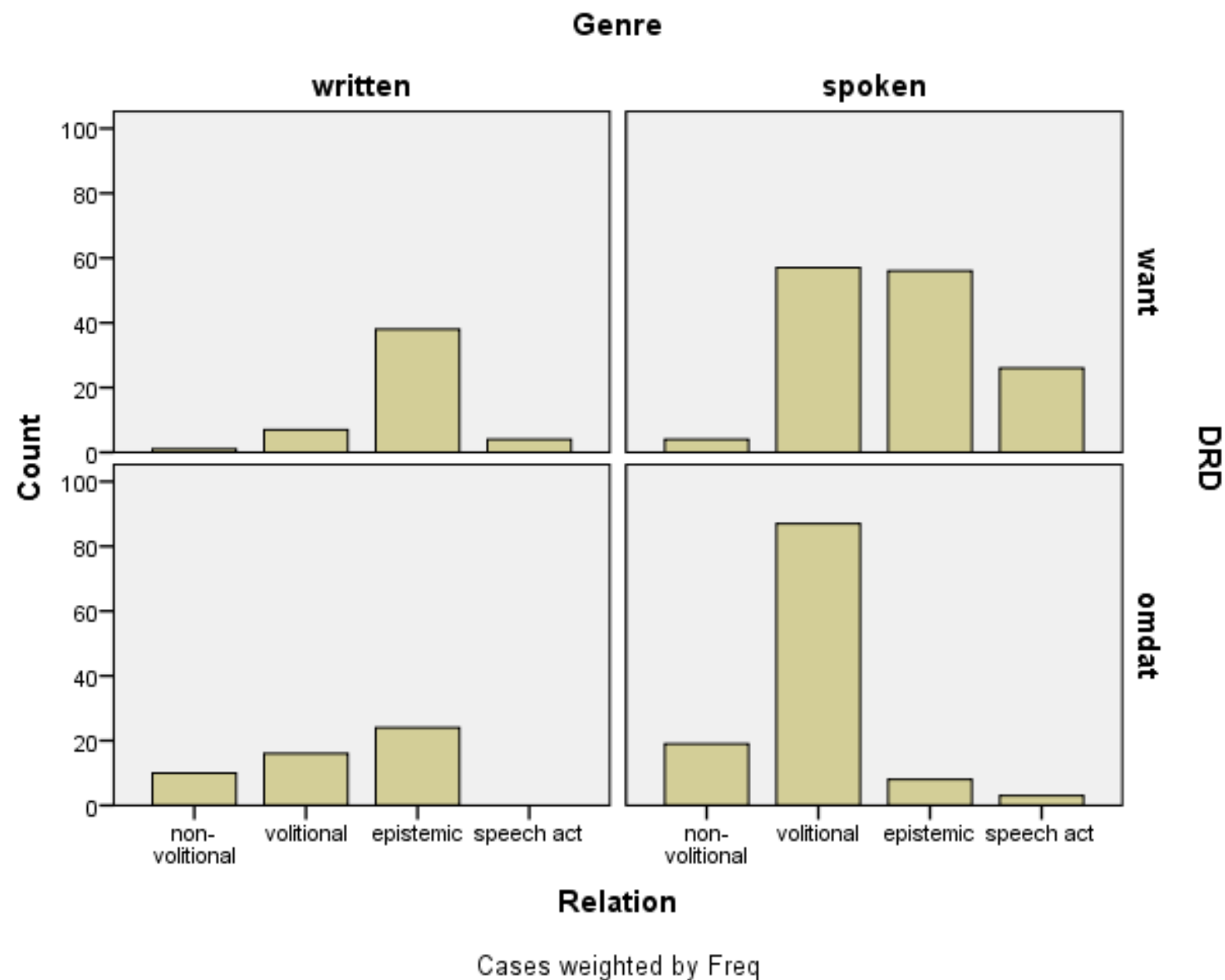
	Genre	DRD	Relation	Freq			
1	written	want	non-volitional				
2	written	want	volitional				
3	written	want	epistemic	38			
4	written	want	speech act	4			
5	written	omdat	non-volitional	10			
6	written	omdat	volitional	16			
7	written	omdat	epistemic	24			
8	written	omdat	speech act	0			
9	spoken	want	non-volitional	4			
10	spoken	want	volitional	57			
11	spoken	want	epistemic	56			
12	spoken	want	speech act	26			
13	spoken	omdat	non-volitional	19			
14	spoken	omdat	volitional	87			
15	spoken	omdat	epistemic	8			



Make graphs of Relation (category axis), DRD (row panel), Genre (column panel)



Resulting graph: answer to research question?



Omdat and *want* seem to express a different relation (*want* more epist/speech act, *omdat* more volitional), but there is also a relation with genre (difference bigger in spoken)

Loglinear analysis

Analysis omdat want written spoken.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window

	Genre	DRD
1	written	want
2	written	want
3	written	want
4	written	want
5	written	omdat
6	written	omdat
7	written	omdat
8	written	omdat
9	spoken	want
10	spoken	want
11	spoken	want
12	spoken	want

Analyze > Loglinear > Model Selection...

Model Selection Loglinear Analysis

Factor(s):

- Genre(? ?)
- DRD(? ?)
- Relation(? ?)

Define Range...
Number of cells: 0

Cell Weights:

Model Building

☒ Use backward elimination: Maximum steps: 10
Probability for removal: ,05

☐ Enter in single step

OK Paste Reset Cancel Help

Factor(s):

- Genre(1 2)
- DRD(1 2)
- Relation(1 4)

Define Range...
Number of cells: 16

Loglinear Analysis: Options

Display

☒ Frequencies
☒ Residuals

Plot

☐ Residuals
☐ Normal Probability

Display for Saturated Model

☐ Parameter estimates ☒ Association table

Model Criteria

Maximum iterations: 20
Convergence: Default
Delta: ,5

Loglinear analysis: output

Cell Counts and Residuals

Genre	DRD	Relation	Observed		Expected		Residuals	Std. Residuals
			Count ^a	%	Count	%		
written	want	non-volitional	1,500	0,4%	1,500	0,4%	,000	,000
		volitional	7,500	2,1%	7,500	2,1%	,000	,000
		epistemic	38,500	10,7%	38,500	10,7%	,000	,000
		speech act	4,500	1,3%	4,500	1,3%	,000	,000
	omdat	non-volitional	10,500	2,9%	10,500	2,9%	,000	,000
		volitional	16,500	4,6%	16,500	4,6%	,000	,000
		epistemic	24,500	6,8%	24,500	6,8%	,000	,000
		speech act	,500	0,1%	,500	0,1%	,000	,000
spoken	want	non-volitional	4,500	1,3%	4,500	1,3%	,000	,000
		volitional	57,500	16,0%	57,500	16,0%	,000	,000
		epistemic	56,500	15,7%	56,500	15,7%	,000	,000
		speech act	26,500	7,4%	26,500	7,4%	,000	,000
	omdat	non-volitional	19,500	5,4%	19,500	5,4%	,000	,000
		volitional	87,500	24,3%	87,500	24,3%	,000	,000
		epistemic	8,500	2,4%	8,500	2,4%	,000	,000
		speech act	3,500	1,0%	3,500	1,0%	,000	,000

a. For saturated models, ,500 has been added to all observed cells.

All expected freq's are equal to observed freq's; all differences between obs. and exp. freq's are 0; $\chi^2 = 0$

Perfect fit between model and observations: all factors are in the model (saturated model)

Loglinear analysis: output

Goodness-of-Fit Tests

	Chi-Square	df	Sig.
Likelihood Ratio	,000	0	.
Pearson	,000	0	.

K-Way and Higher-Order Effects

	K	df	Likelihood Ratio		Pearson		Number of Iterations
			Chi-Square	Sig.	Chi-Square	Sig.	
K-way and Higher Order Effects ^a	1	15	381,458	,000	418,756	,000	0
	2	10	147,067	,000	130,529	,000	2
	3	3	4,481	,214	3,819	,282	7
K-way Effects ^b	1	5	234,391	,000	288,226	,000	0
	2	7	142,586	,000	126,710	,000	0
	3	3	4,481	,214	3,819	,282	0

a. Tests that k-way and higher order effects are zero.

b. Tests that k-way effects are zero.

Loglinear analysis: output

Partial Associations

Effect	df	Partial Chi-Square	Sig.	Number of Iterations
Genre*DRD	1	9,303	,002	2
Genre*Relation	3	58,291	,000	2
DRD*Relation	3	92,149	,000	2
Genre	1	73,660	,000	2
DRD	1	1,879	,170	2
Relation	3	158,852	,000	2

Loglinear analysis: output

Backward Elimination Statistics

Step Summary						
Step ^a		Effects	Chi-Square ^c	df	Sig.	Number of Iterations
0	Generating Class ^b	Genre*DRD*Relation	,000	0	.	
	Deleted Effect 1	Genre*DRD*Relation	4,481	3	,214	7
1	Generating Class ^b	Genre*DRD, Genre*Relation, DRD*Relation	4,481	3	,214	
	Deleted Effect 1	Genre*DRD	9,303	1	,002	2
	2	Genre*Relation	58,291	3	,000	2
	3	DRD*Relation	92,149	3	,000	2
2	Generating Class ^b	Genre*DRD, Genre*Relation, DRD*Relation	4,481	3	,214	

a. At each step, the effect with the largest significance level for the Likelihood Ratio Change is deleted, provided the significance level is larger than ,050.

b. Statistics are displayed for the best model at each step after step 0.

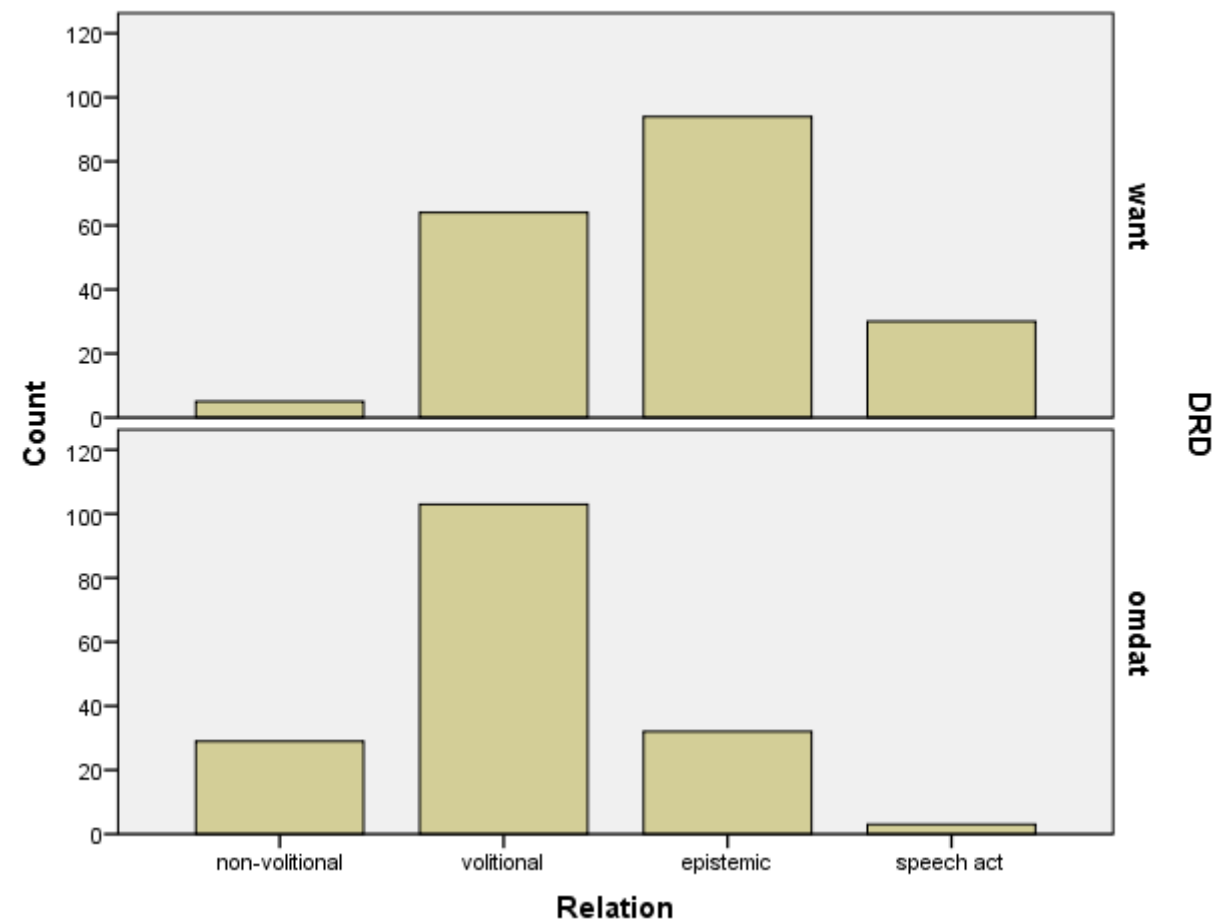
Only the 3-way interaction can be deleted.

Deletion of every 2-way interaction leads to significant reduction of the fit of the model

The analysis stops with a model containing three 2-way interactions, with a fit of 4.48

Graphs of the interactions: Relation * Genre

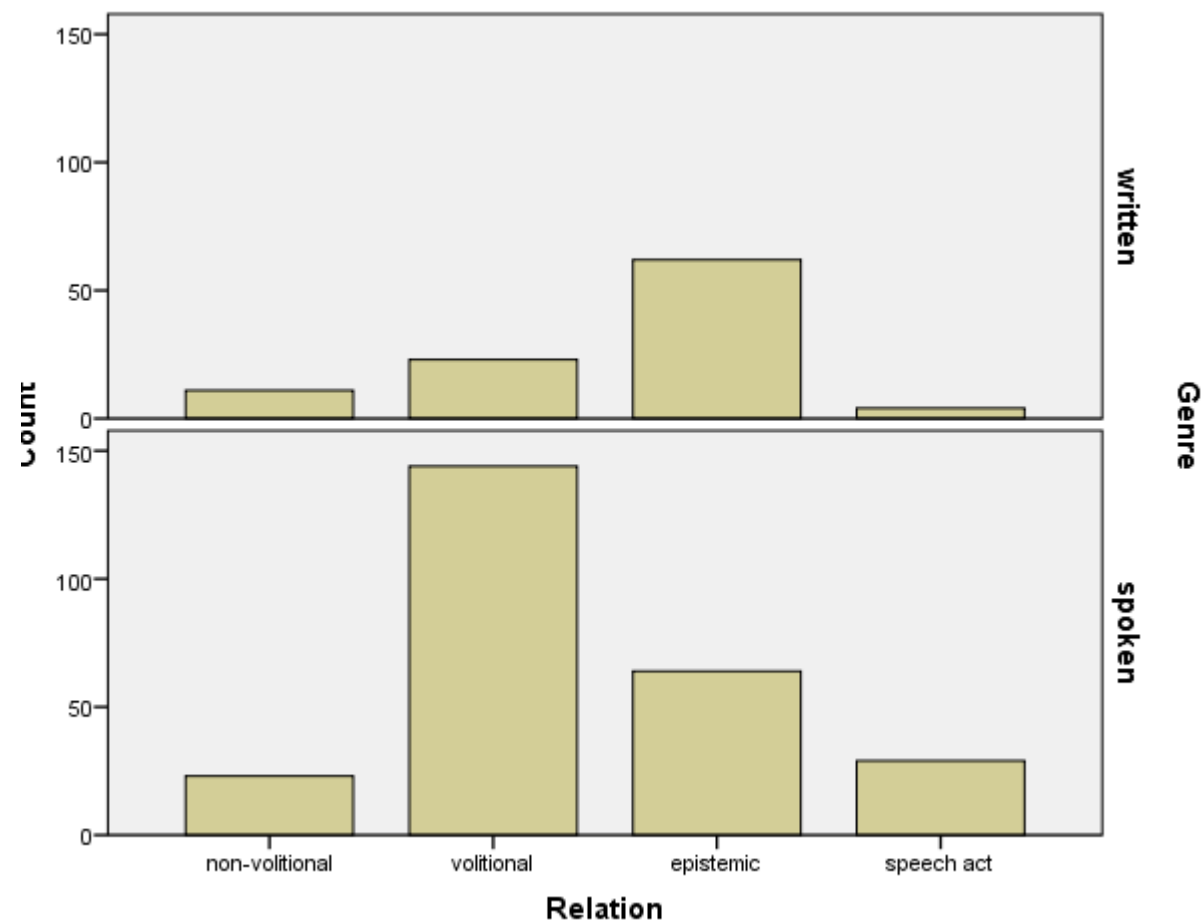
Graph



Want: relative many epistemic speech act

Omdat: relative many non-volitional and volitional

Graphs of the interactions: Relation * Genre



Written: less relations across the board, especially volitional and speech act

Size of the effect for 2-way interactions: odds-ratio for 2x2 tables, Cramer's V for larger tables

Analysis omdat want written spoken.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window

Genre DRD

1	written	want
2	written	want
3	written	want
4	written	want
5	written	omdat
6	written	omdat
7	written	omdat
8	written	omdat
9	spoken	want

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction

Frequencies...
Descriptives...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...

Crosstabs

Genre
Freq

Row(s):
Relation

Column(s):
DRD

Layer 1 of 1

Previous Next

Display clustered bar charts
Suppress tables

Display layer variables in table layers

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,463	,000
	Cramer's V	,463	,000
N of Valid Cases		360	

Crosstabs: Statistics

☒ Chi-square ☐ Correlations

Nominal

☐ Contingency coefficient
☒ Phi and Cramer's V
☐ Lambda
☐ Uncertainty coefficient

Ordinal

☐ Gamma
☐ Somers' d
☐ Kendall's tau-b
☐ Kendall's tau-c

Nominal by Interval

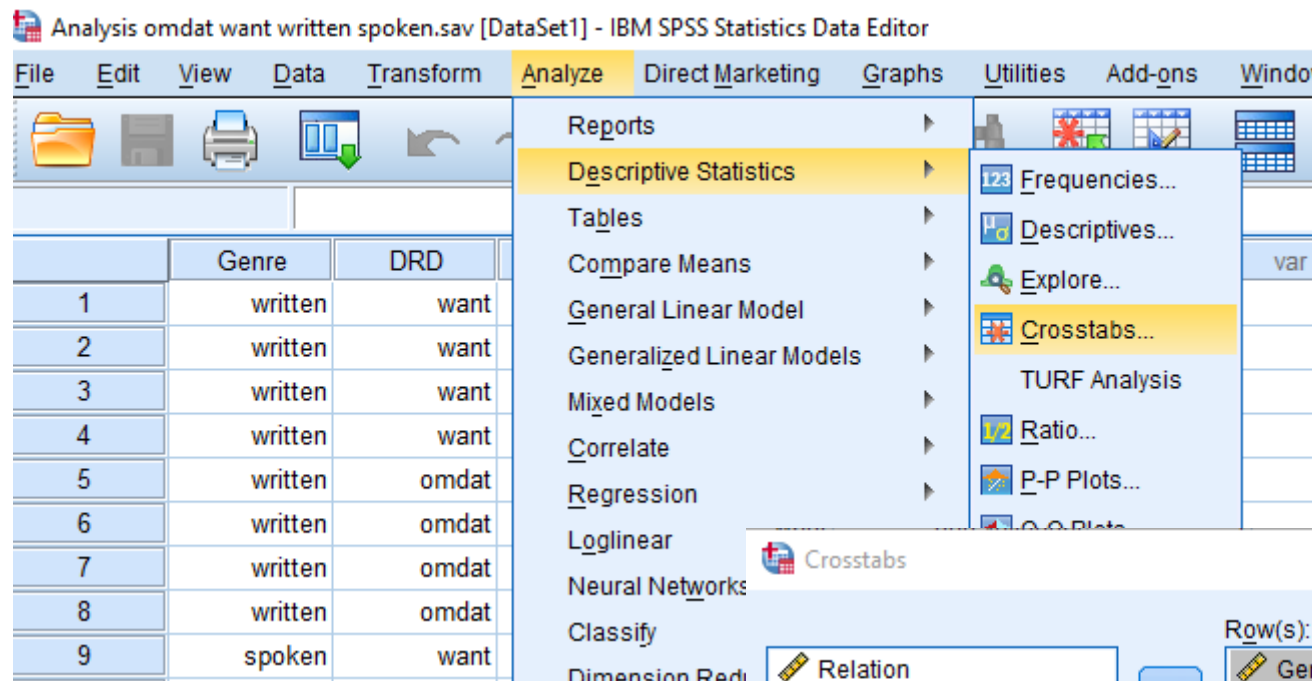
☐ Eta

☐ Kappa
☐ Risk
☐ McNemar

☐ Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals: 1

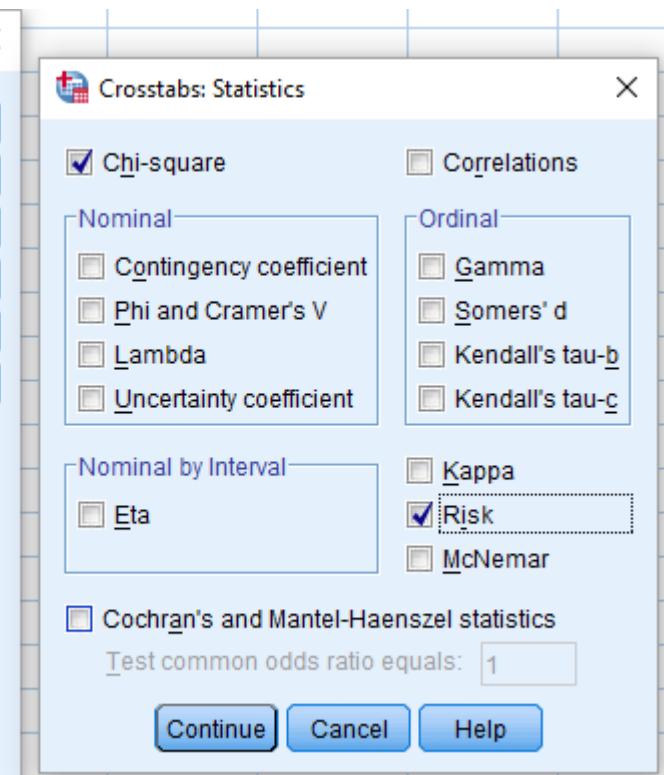
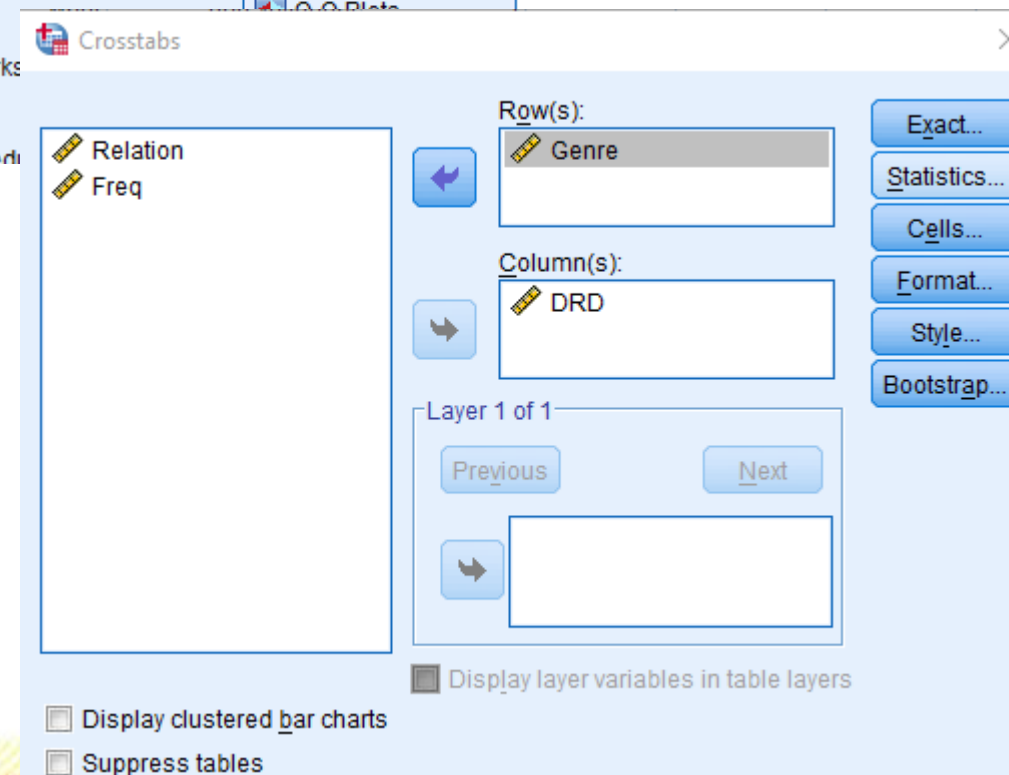
Continue Cancel Help

Size of the effect for 2-way interactions: odds-ratio for 2x2 tables, Cramer's V for larger tables



Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Genre (written / spoken)	,818	,516	1,299
For cohort DRD = want	,909	,726	1,138
For cohort DRD = omdat	1,111	,876	1,409
N of Valid Cases	360		



Reporting the result

- The 3-way loglinear analysis produced a model containing three effects: a 2-way interaction between genre and DRD ($\chi^2(1) = 9.30, p < .05$), a 2-way interaction between DRD and relation ($\chi^2(3) = 58.29, p < .01$), and a 2-way interaction between genre and relation ($\chi^2(3) = 92.15, p < .01$). The goodness-of-fit of the resulting model was $\chi^2(3) = 4.48, p = .21$. The 2-way interaction between genre and DRD can be expressed in terms of an odds ratio: It is 0.82 more likely that a DRD is *want* if the genre is written than if the genre is spoken. The 2-way interaction between genre and relation is a medium effect (Cramer's V: .37) reflecting that written texts have less relations across the board, especially volitional and speech act relations. The 2-way interaction between DRD and relation is a medium effect (Cramer's V: .47) reflecting that *want* is often used to express epistemic and speech act relations, whereas *omdat* is often used to express non-volitional and volitional relations.”