

ANOVA (GLM- Repeated measures)

Jantien Donkers

When to use..

- o T-test: 2 conditions testing 1 independent variable (e.g. text complexity – simple/complex in relation to number of recalled words)
- o ANOVA: 3 or more conditions testing 1 (one way ANOVA) or more (two way ANOVA) variables

ANOVA's

- o F-ratio

- o Size of the variance due to the experimental conditions in relation to the error (unexpected) variance

- o Degrees of freedom (df)

ANOVA's (2)

- o H_0 : all means are equal
- o Alternative: all of means are different, or just one of them

- o Variation among groups is compared to variation within groups: a relatively large difference is evidence against H_0

Parametric assumptions

- o Experimental scores are measured on an *interval scale*
- o Scores are *normally distributed*
- o Variability of scores for each conditions should be roughly the same (*homogeneity of variance*)

This experiment

- o 6 conditions testing 2 independent variables
- o Dependent variable: reading time

This experiment (2)

- o a self-paced reading study in Dutch
- o “who” and “which” questions
- o *specificity* and *structure*

Why?

- o In various research it is observed that “which” questions are more difficult than “who” questions, but:
- o It has never been addressed *why* this is the case

Variable 1: specificity

“Who”

Wie heeft de keizer gezocht in de kelder?

“Which” generic

Welke persoon heeft de keizer gezocht in de kelder?

“Which” specific

Welke bediende heeft de keizer gezocht in de kelder?

Variable 2: structure

o **Sentence structure** is manipulated by context:

SO

Terwijl de dronken bediende een dutje deed, zocht de nuchtere bediende de keizer in de kelder.

OS

Terwijl de dronken bediende een dutje deed zocht de keizer de nuchtere bediende in de kelder.

Overview conditions

o 3x2 design: 6 conditions

Who	SO	OS
WhichGen	SO	OS
WhichSpec	SO	OS

o 8 scenario/question combinations per conditions
(total of 48)

Participants

- o 48 (14 male, 34 female)
- o Mean age 22.1 (sd 2.34)
- o Normal or corrected to normal vision
- o Paid for participation

Procedure & analysis

- o Phrase-by-phrase self-paced reading (using E-prime software package)
- o “Moving window”
- o Accuracy: participants had to judge a provided answer (correct/incorrect) by pressing the corresponding button
- o Reading times and accuracy analyzed

Terwijl de dronken bediende een hapje at, zocht de nuchtere bediende de keizer in de kelder.

+

March 14, 2005

Methodology & Statistics

Wie

March 14, 2005

Methodology & Statistics

heeft

March 14, 2005

Methodology & Statistics

de keizer

March 14, 2005

Methodology & Statistics

gezocht

March 14, 2005

Methodology & Statistics

in de kelder?

March 14, 2005

Methodology & Statistics

de nuchtere bediende

March 14, 2005

Methodology & Statistics

Hypotheses

- o Set-restriction (“specificity”) is a (the?) complicating factor during wh-question processing
 - o Also when the questions are presented within an appropriate context
- o **BUT**
- o Processing difficulties in set-restricted wh-questions interact with difficulties in sentence structure

Hypothesis 1

o Set-restriction is a complicating factor during wh-question processing

➤ WhichSpec > WhichGen = Who

Alternatives:

➤ WhichSpec = WhichGen > Who

➤ WhichSpec = WhichGen = Who

(in reading times (RTs))

Hypothesis 2

- o Processing difficulties in set-restricted wh-questions interact with difficulties in sentence structure
 - WhichSpec OS > WhichGen OS = Who OS
 - WhichSpec SO = WhichGen SO = Who SO

Possible alternative:

- All wh-types OS > SO

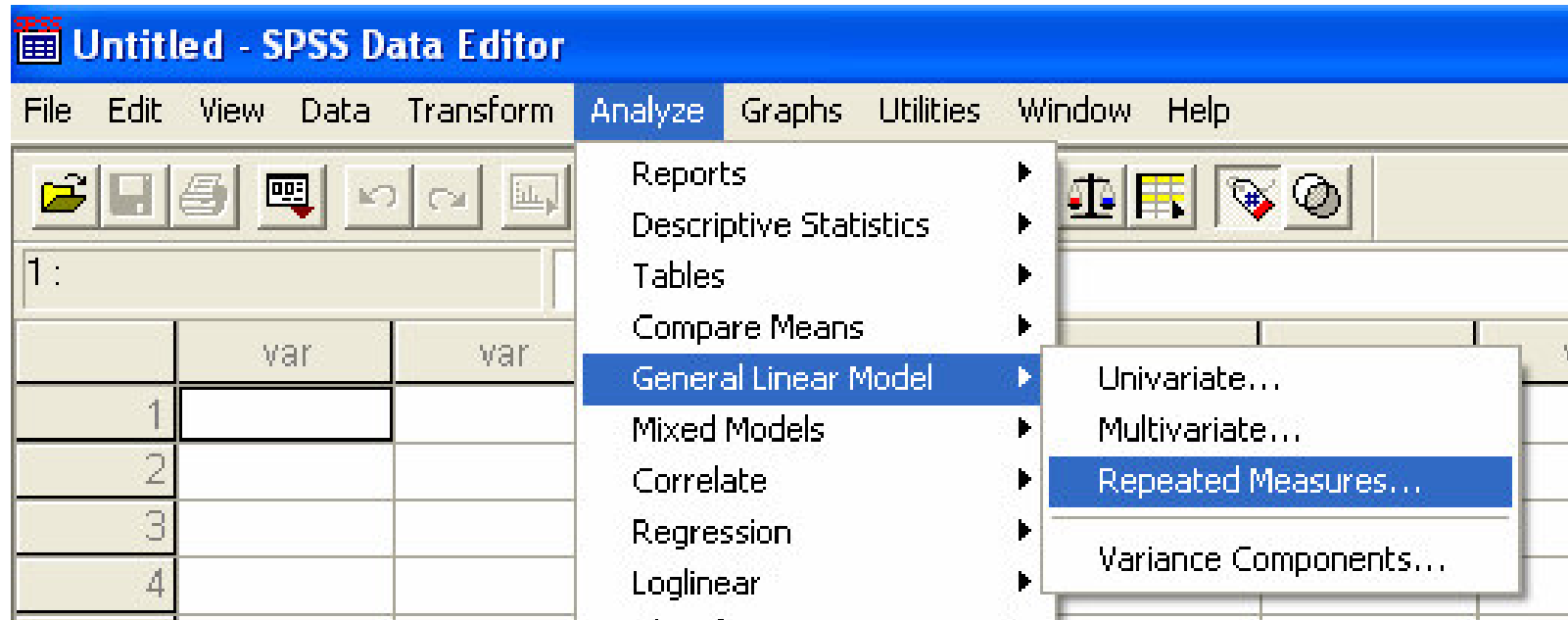
Segment of interest

- o “Point of integration”:

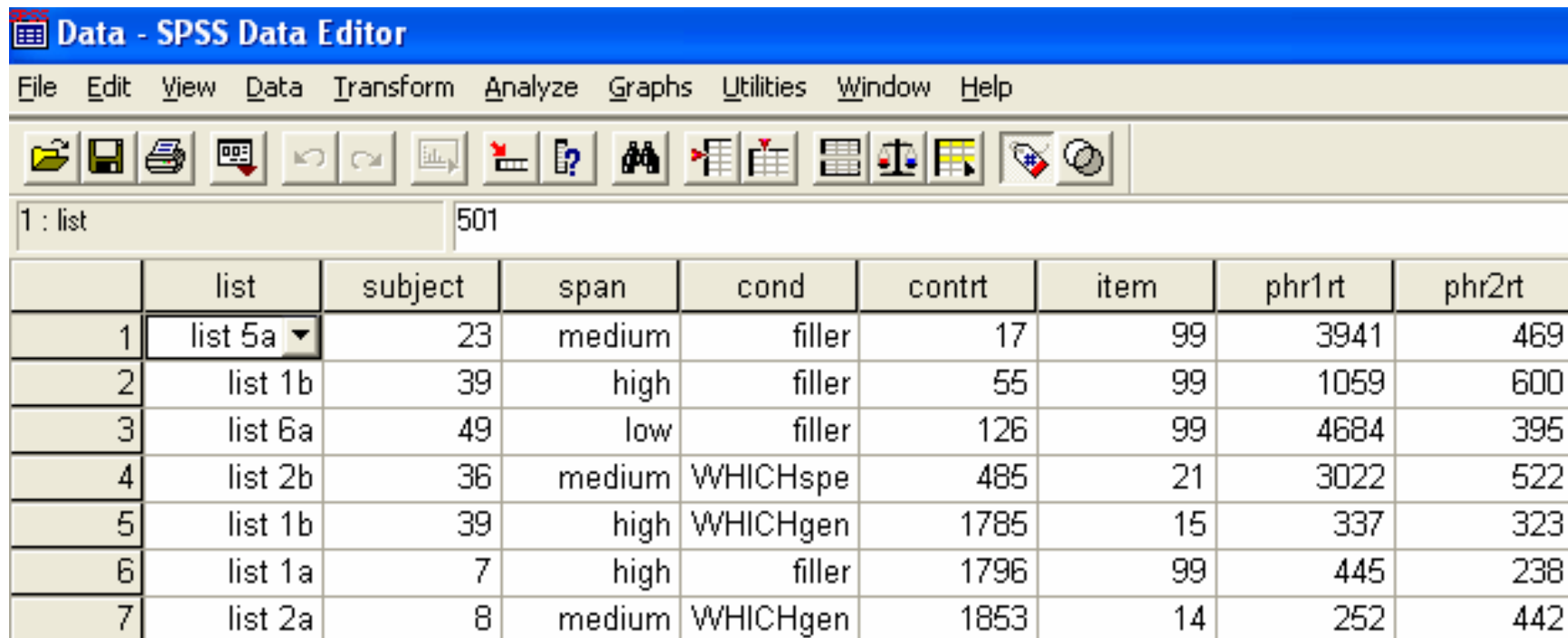
Welke bediende heeft de keizer gezocht in de kelder?

- o At this position (i.e. participle) it becomes clear which role each individual NP (Wh-phrase and “de keizer”) plays

Repeated measures/ SPSS



Data



The screenshot shows the SPSS Data Editor interface. The title bar reads "Data - SPSS Data Editor". The menu bar includes "File", "Edit", "View", "Data", "Transform", "Analyze", "Graphs", "Utilities", "Window", and "Help". The toolbar contains various icons for file operations, editing, and analysis. The data grid shows a table with 8 columns and 7 rows. The first row is highlighted. The "list" column has a dropdown menu open, showing "list 5a" selected.

	list	subject	span	cond	contrt	item	phr1rt	phr2rt
1	list 5a	23	medium	filler	17	99	3941	469
2	list 1b	39	high	filler	55	99	1059	600
3	list 6a	49	low	filler	126	99	4684	395
4	list 2b	36	medium	WHICHspe	485	21	3022	522
5	list 1b	39	high	WHICHgen	1785	15	337	323
6	list 1a	7	high	filler	1796	99	445	238
7	list 2a	8	medium	WHICHgen	1853	14	252	442

Data pre-processing (1)

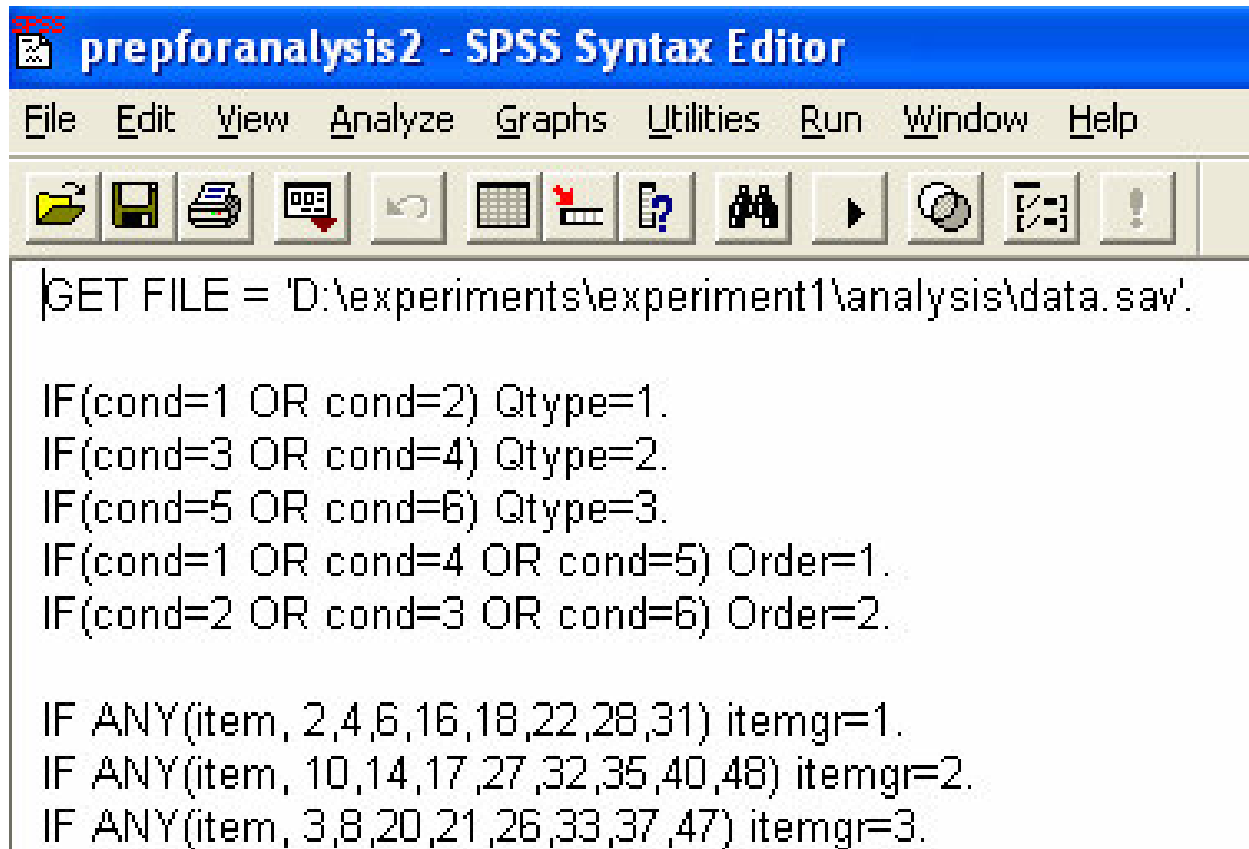
- o Define conditions (in terms of factors)
 - o Cond 1 or 2: Qtype 1 (Who)
 - o Cond 1 or 4 or 5: Order 1 (SO)
- o Define item groups
- o Define lists

Data pre-processing (2)

- o Define cut-offs (outliers)
- o Calculate segment means and sds
 - o Per subject
 - o Per item
- o Define limits (mean+ 2sd)
- o Replace outliers and >limits by mean+2sd
- o Data-transformation (suitable for SPSS)

Data pre-processing (3)

- o Write syntax script



```
GET FILE = 'D:\experiments\experiment1\analysis\data.sav'.  
  
IF(cond=1 OR cond=2) Qtype=1.  
IF(cond=3 OR cond=4) Qtype=2.  
IF(cond=5 OR cond=6) Qtype=3.  
IF(cond=1 OR cond=4 OR cond=5) Order=1.  
IF(cond=2 OR cond=3 OR cond=6) Order=2.  
  
IF ANY(item, 2,4,6,16,18,22,28,31) itemmgr=1.  
IF ANY(item, 10,14,17,27,32,35,40,48) itemmgr=2.  
IF ANY(item, 3,8,20,21,26,33,37,47) itemmgr=3.
```

Data pre-processing (4)

- o Write syntax script

```
COMPUTE phr1rt1=phr1rt.  
IF(phr1rt<150 OR phr1rt>5000) phr1rt1=0.  
MISSING VALUES phr1rt1(0).
```

```
COMPUTE phr2rt1=phr2rt.  
IF(phr2rt<150 OR phr2rt>5000) phr2rt1=0.  
MISSING VALUES phr2rt1(0).
```

```
COMPUTE phr3rt1=phr3rt.  
IF(phr3rt<140 OR phr3rt>5000) phr3rt1=0.  
MISSING VALUES phr3rt1(0).
```

```
COMPUTE phr4rt1=phr4rt.
```


Data pre-processing (5)

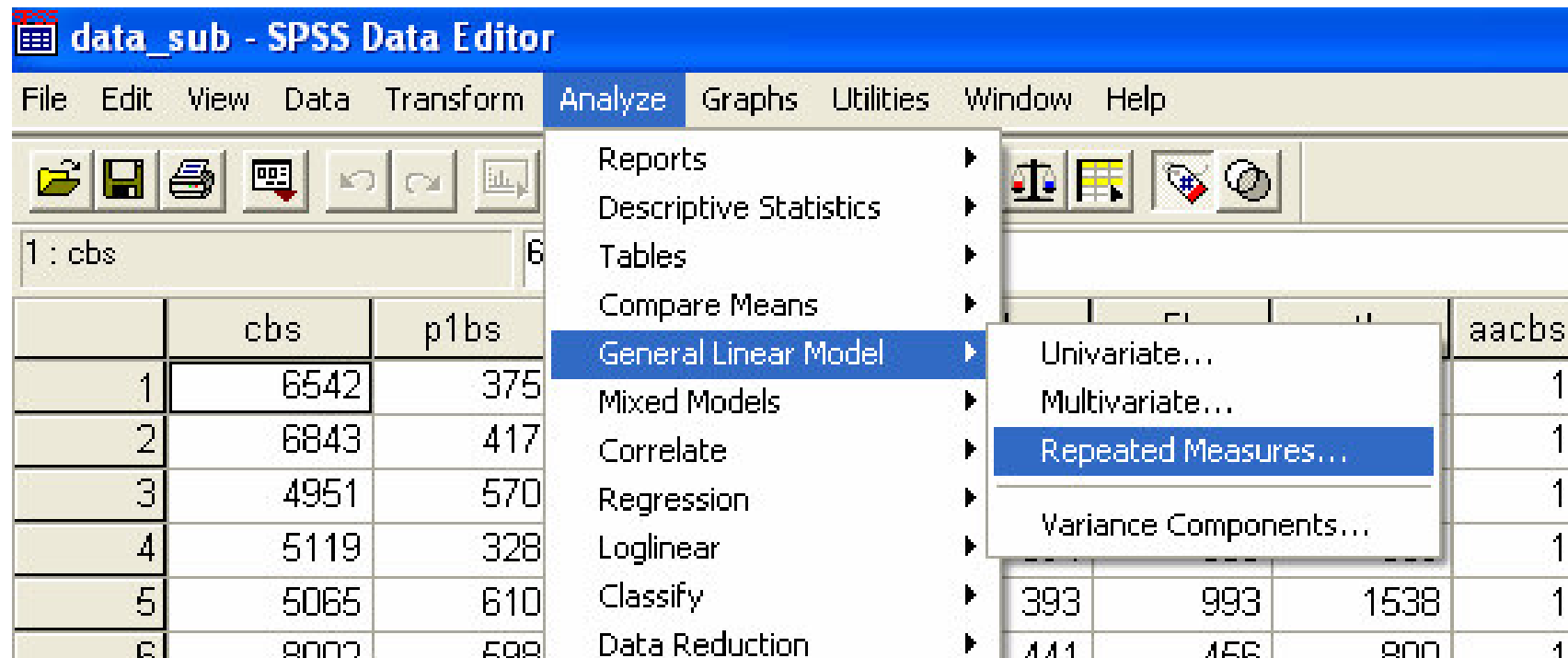
- o Write syntax script

```
COMPUTE xphr1rt = phr1rt1.  
IF (phr1rt1 < lopp1) xphr1rt = lopp1.  
IF (phr1rt1 > hipp1) xphr1rt = hipp1.  
IF (phr1rt1 < loit1) xphr1rt = loit1.  
IF (phr1rt1 > hiit1) xphr1rt = hiit1.  
COMPUTE hmean=(hipp1+hiit1)/2.  
COMPUTE lmean=(lopp1+loit1)/2.  
IF ((phr1rt1 > hipp1) AND (phr1rt1 > hiit1)) xphr1rt=hmean.  
IF ((phr1rt1 < lopp1) AND (phr1rt1 < loit1)) xphr1rt=lmean.  
  
COMPUTE xphr2rt = phr2rt1.  
IF (phr2rt1 < lopp2) xphr2rt = lopp2.  
IF (phr2rt1 > hipp2) xphr2rt = hipp2.
```

GLM – Repeated measures

- o SPSS data file with **subject means**
- o SPSS data file with **item means**

GLM – Repeated measures



The screenshot shows the SPSS Data Editor interface. The title bar reads "data_sub - SPSS Data Editor". The menu bar includes "File", "Edit", "View", "Data", "Transform", "Analyze", "Graphs", "Utilities", "Window", and "Help". The "Analyze" menu is open, showing options: Reports, Descriptive Statistics, Tables, Compare Means, General Linear Model (highlighted), Mixed Models, Correlate, Regression, Loglinear, Classify, and Data Reduction. The "General Linear Model" sub-menu is also open, showing: Univariate..., Multivariate..., Repeated Measures... (highlighted), and Variance Components... The data grid shows a table with columns "cbs" and "p1bs". The first row is highlighted, and the value "1" is visible in the first column. The variable "aacbs" is also visible in the grid.

	cbs	p1bs	aacbs
1	6542	375	1
2	6843	417	1
3	4951	570	1
4	5119	328	1
5	5065	610	1
6	5000	500	1

GLM – Repeated measures

	cbs	p1bs	p2bs	p3bs	p4bs	p5bs	artbs	aacbs	cbo	p
1	6542	375	332	346	380	404	1773	1	10691	
2	6843	417	342	334	352	400	1752	1	7913	
3	4951	570	347						6003	
4	5119	328	220						6039	
5	5065	610	374						6611	
6	8002	598	333						6701	
7	5638	287	241						5241	
8	7314	474	311						8768	
9	5018	494	319						4466	
10	5997	327	260						6985	
11	8901	416	250						6459	
12	6560	344	231	281	407	400	505	1	6142	
13	7732	480	367	379	443	543	1169	1	8569	

Repeated Measures Define Factor(s)

Within-Subject Factor Name:

Number of Levels:

Add Change Remove

qtype(3)

Define Reset Cancel Help Measure >>

GLM – Repeated measures

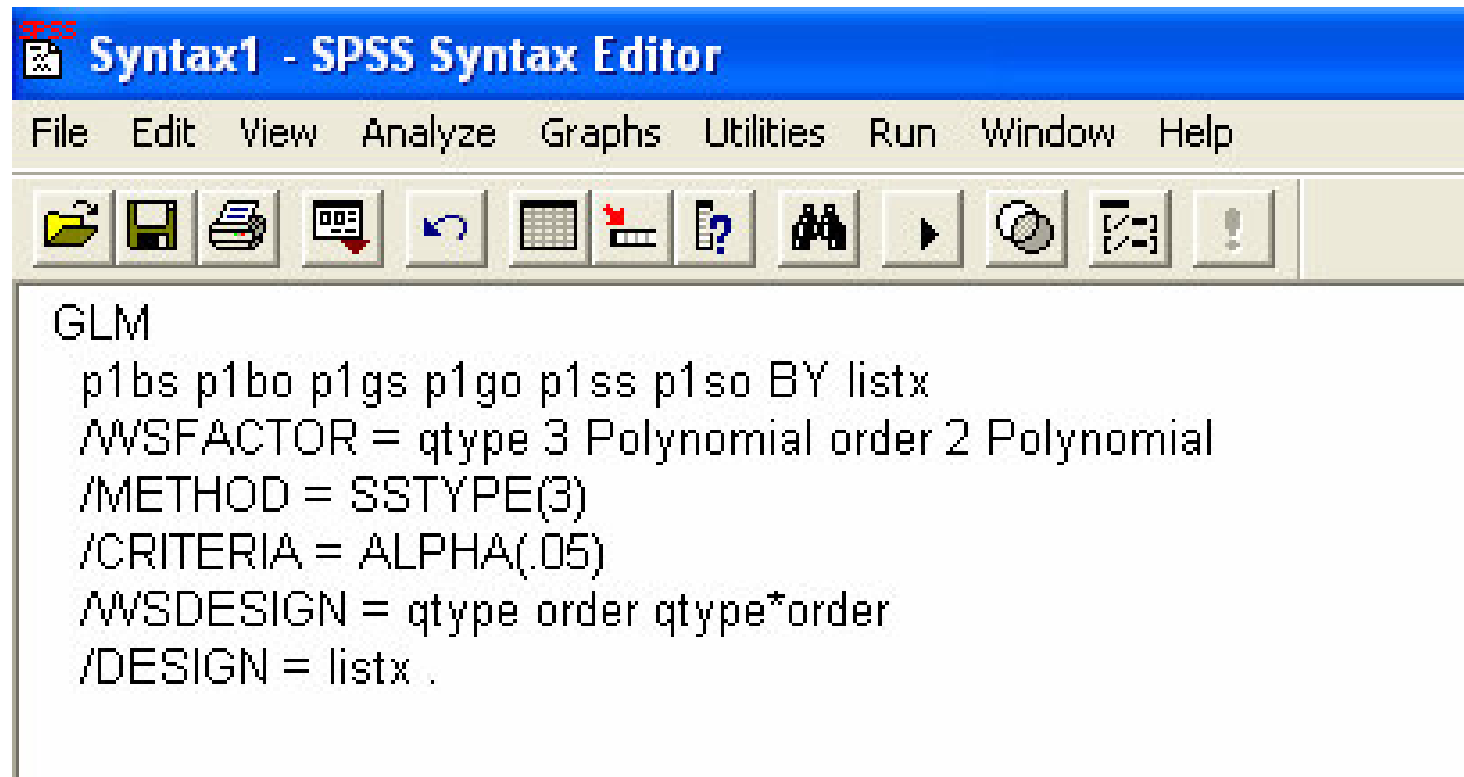
The image shows the SPSS 'Repeated Measures' dialog box overlaid on a data table. The dialog box is titled 'Repeated Measures' and has a blue header bar. It contains several sections:

- Within-Subjects Variables (qtype,order):** A list of variables including p1bs(1,1), p1bo(1,2), p1gs(2,1), p1go(2,2), p1ss(3,1), and p1so(3,2).
- Between-Subjects Factor(s):** A list containing the variable 'listx'.
- Covariates:** An empty list.
- Buttons:** 'Model...', 'Contrasts...', 'Plots...', 'Post Hoc...', 'Save...', and 'Options...' are located at the bottom of the dialog box. 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' are on the right side.

The background data table has columns 'cbs', 'bo', and 'p2bo'. The 'cbs' column contains values from 6542 to 5457. The 'bo' column contains values from 399 to 634. The 'p2bo' column contains values from 39 to 47.

	cbs	bo	p2bo
1	6542	399	39
2	6843	443	37
3	4951	593	38
4	5119	340	23
5	5065	542	40
6	8002	580	33
7	5638	304	22
8	7314	544	30
9	5018	469	34
10	5997	316	28
11	8901	372	28
12	6560	296	25
13	7732	386	36
14	5537	477	43
15	7872	450	34
16	12430	429	41
17	8199	667	33
18	6492	626	48
19	5770	494	35
20	9439	457	24
21	5457	634	47

GLM – Repeated measures



The image shows a screenshot of the SPSS Syntax Editor window. The title bar reads "Syntax1 - SPSS Syntax Editor". The menu bar includes "File", "Edit", "View", "Analyze", "Graphs", "Utilities", "Run", "Window", and "Help". The toolbar contains icons for file operations (Open, Save, Print, Save As), navigation (Undo, Redo), and execution (Run, Stop, Refresh, Help). The main text area contains the following SPSS syntax:

```
GLM  
  p1bs p1bo p1gs p1go p1ss p1so BY listx  
  /WSFACTOR = qtype 3 Polynomial order 2 Polynomial  
  /METHOD = SSTYPE(3)  
  /CRITERIA = ALPHA(.05)  
  /WSDESIGN = qtype order qtype*order  
  /DESIGN = listx .
```

GLM – Repeated measures

The screenshot displays the SPSS Output1 - SPSS Viewer window. The main content area shows the results for a General Linear Model. The dependent variable is MEASURE_1. The within-subjects factors are defined by a table with columns QTYPE, ORDER, and Dependent Variable. The between-subjects factors are defined by a table with columns LISTX and N.

General Linear Model

Measure: MEASURE_1

QTYPE	ORDER	Dependent Variable
1	1	P1BS
	2	P1BO
2	1	P1GS
	2	P1GO
3	1	P1SS
	2	P1SO

Between-Subjects Factors

LISTX	N
1	8
2	8
.	

The taskbar at the bottom shows the Windows Start button, several open applications (Novell, Microsoft Office, Indel..., 3 SPS..., Microsoft..., Analysis), and the system clock showing 17:42 on March 14, 2005.

March 14, 2005

Methodology & Statistics

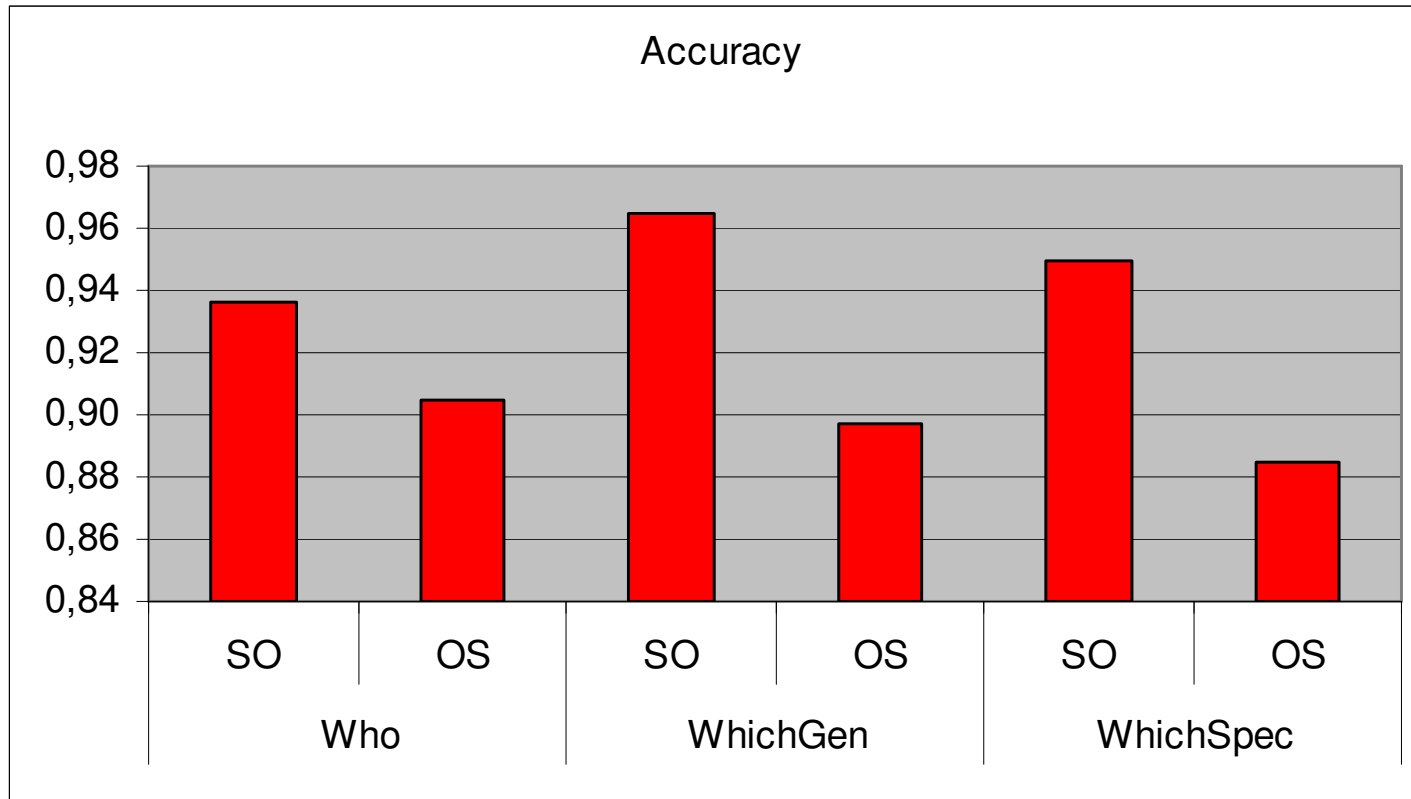
Effects (subject analysis)

- o Main effect of question type (who, whichGen, whichSpec)
- o Main effect of order (SO, OS)
- o Interaction question type by order
- o (Interaction ...by list)

Effect or interaction?

- o Plot your data!
 - o Mistake in labelling
 - o Effect can be counter-intuitive
 - o Post-hoc analyses

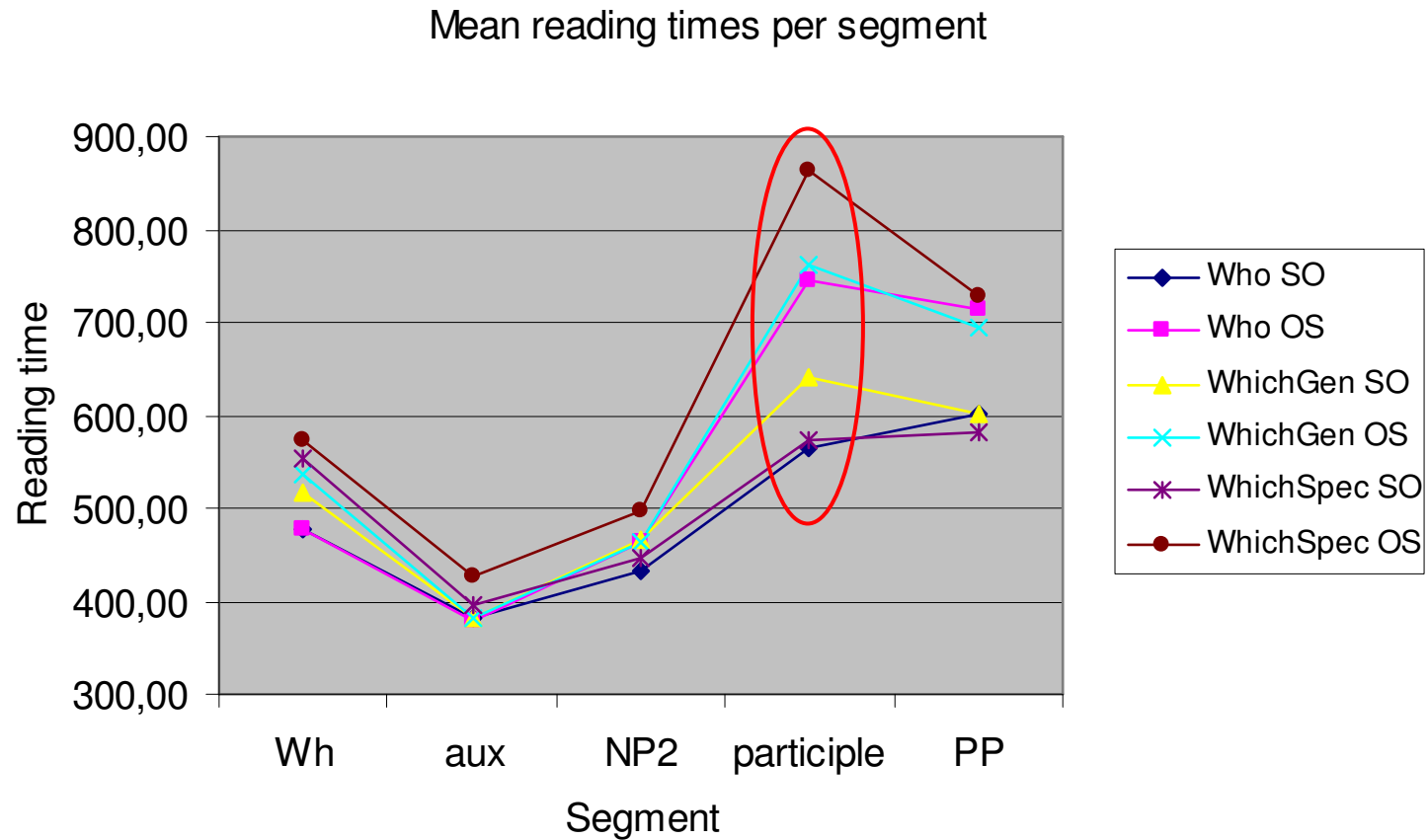
Accuracy



Accuracy

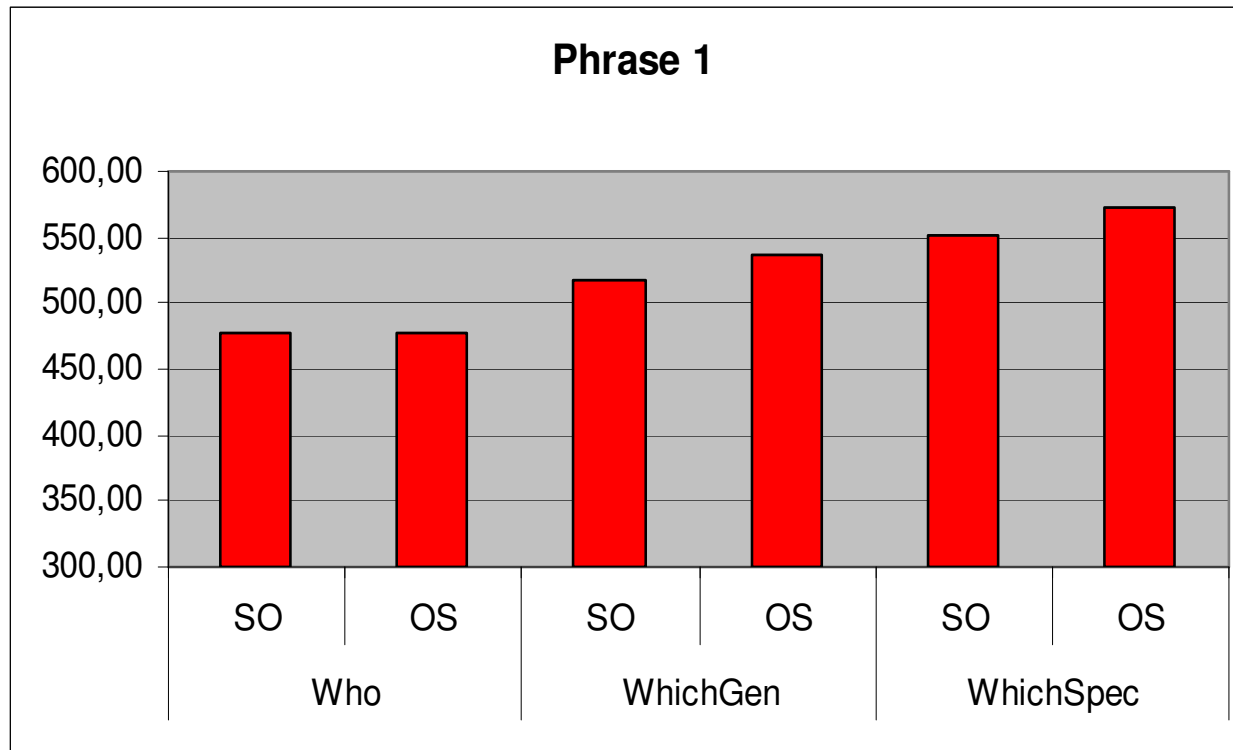
- o The data suggest that answers following OS structures are more difficult to judge than those following SO questions.

Reading times



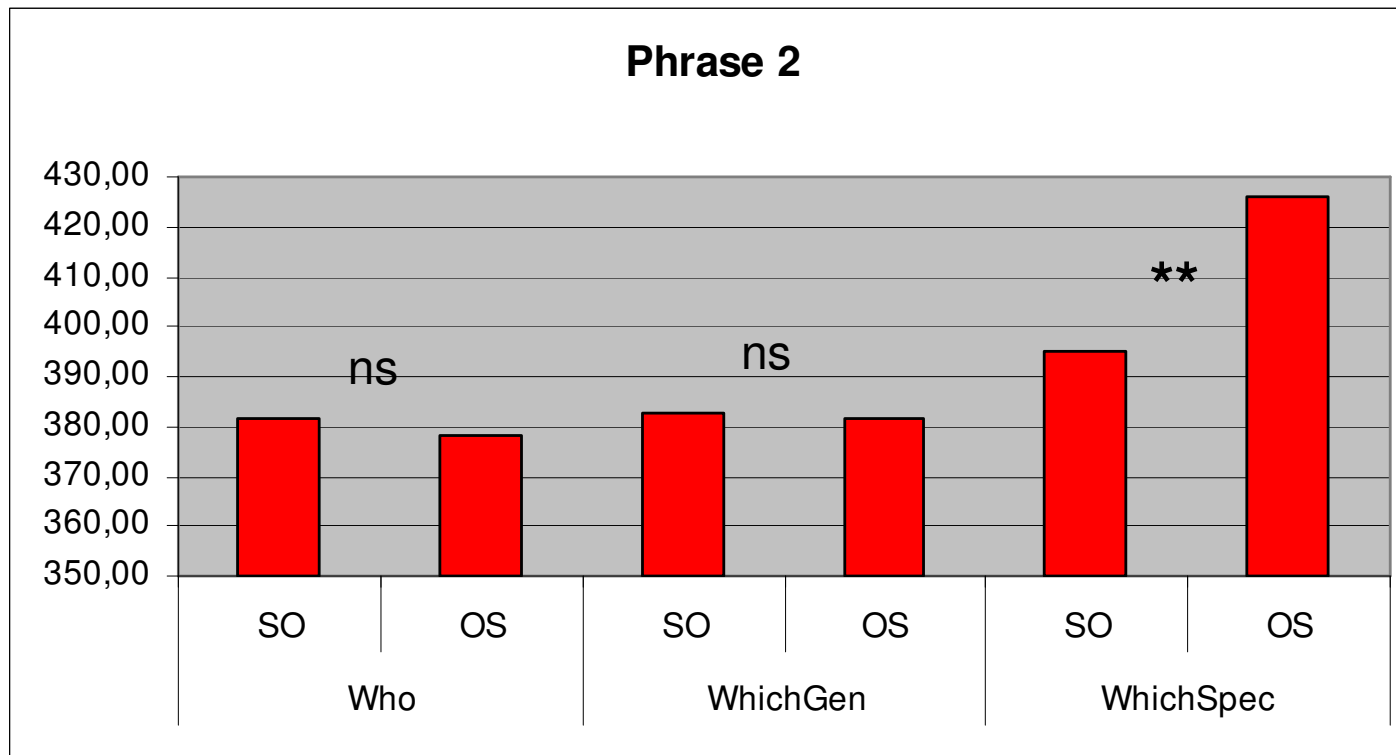
Reading times (2)

o Phrase 1: wh-element



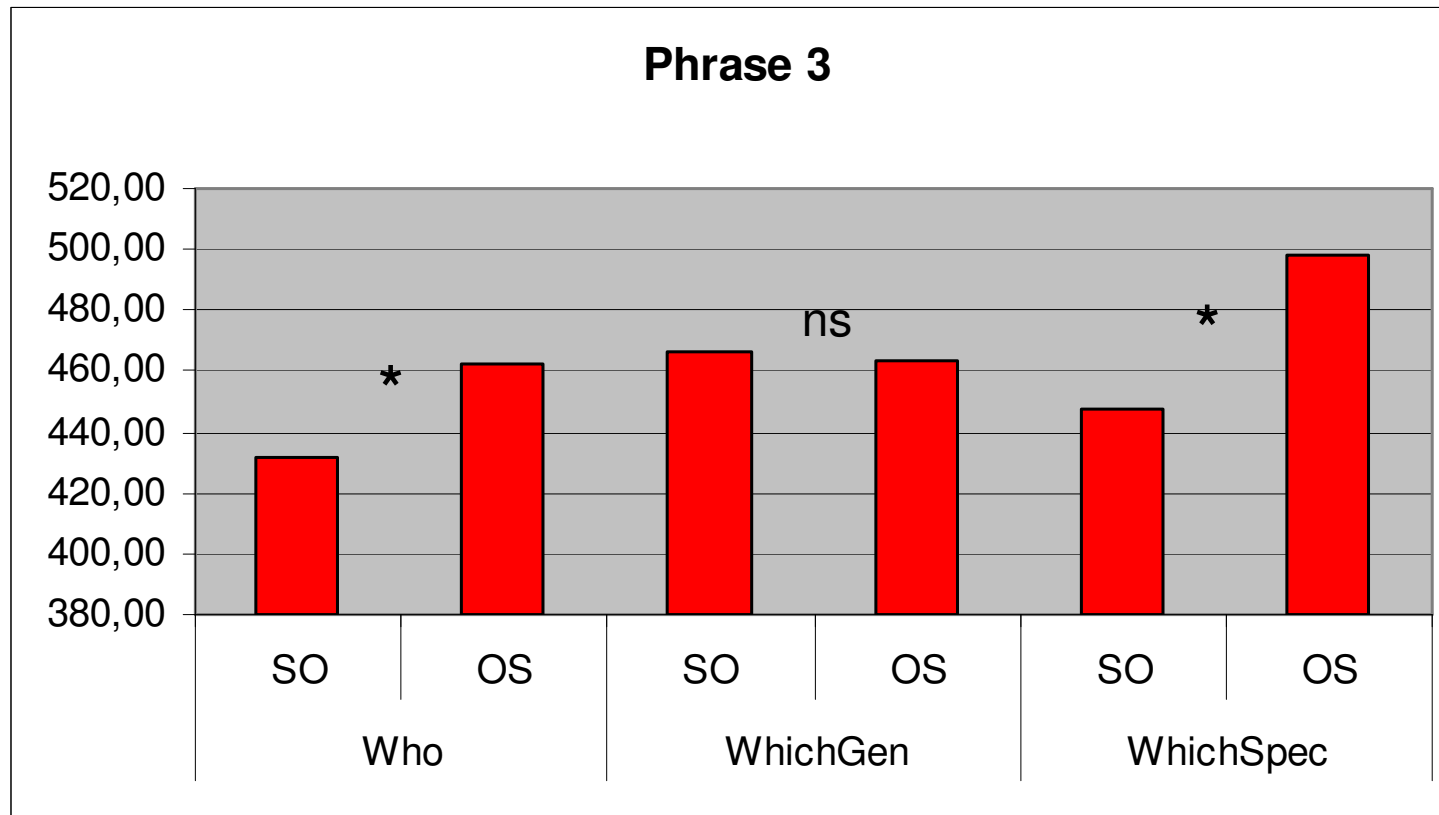
Reading times (3)

o Phrase 2: auxiliary



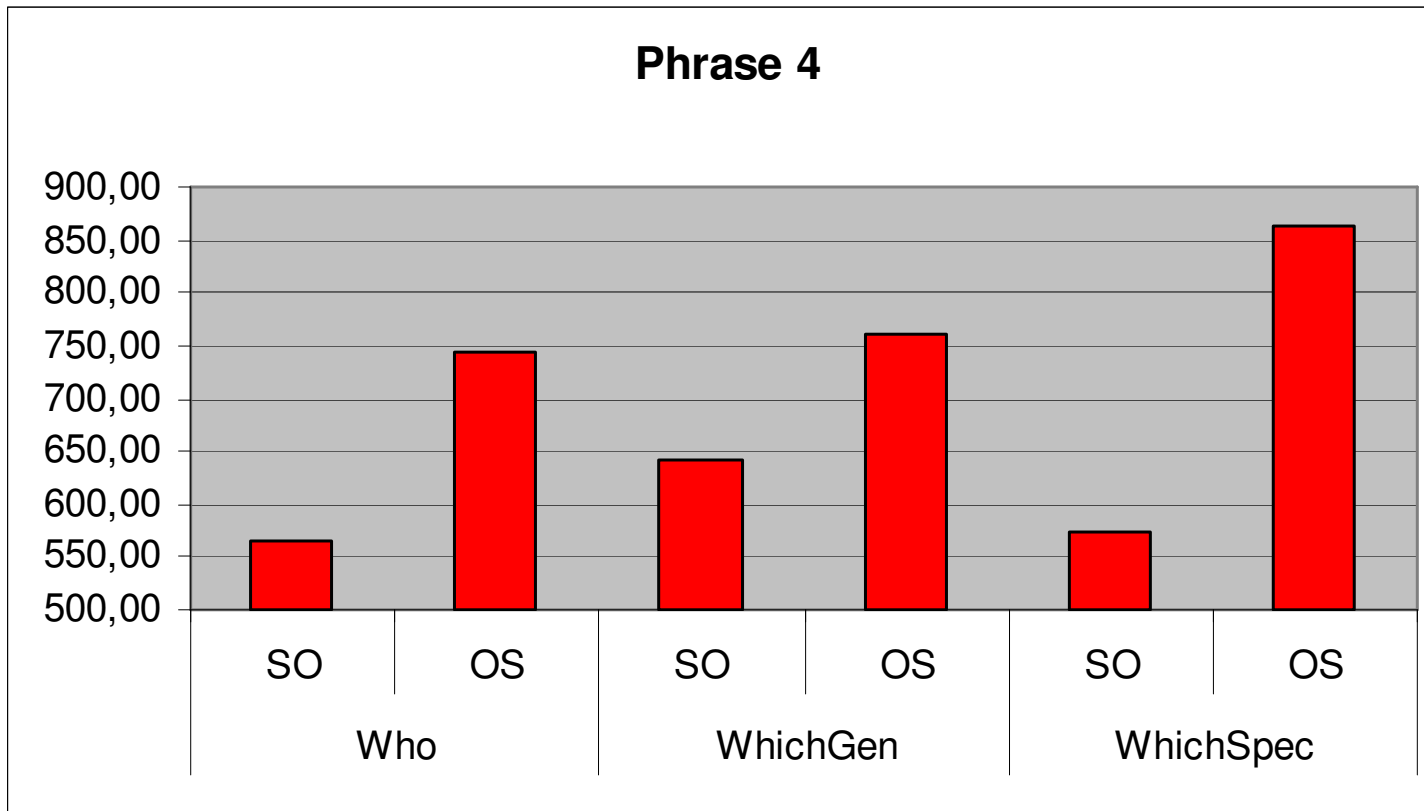
Reading times (4)

o Phrase 3: NP2



Reading times (5)

o Phrase 4: Participle

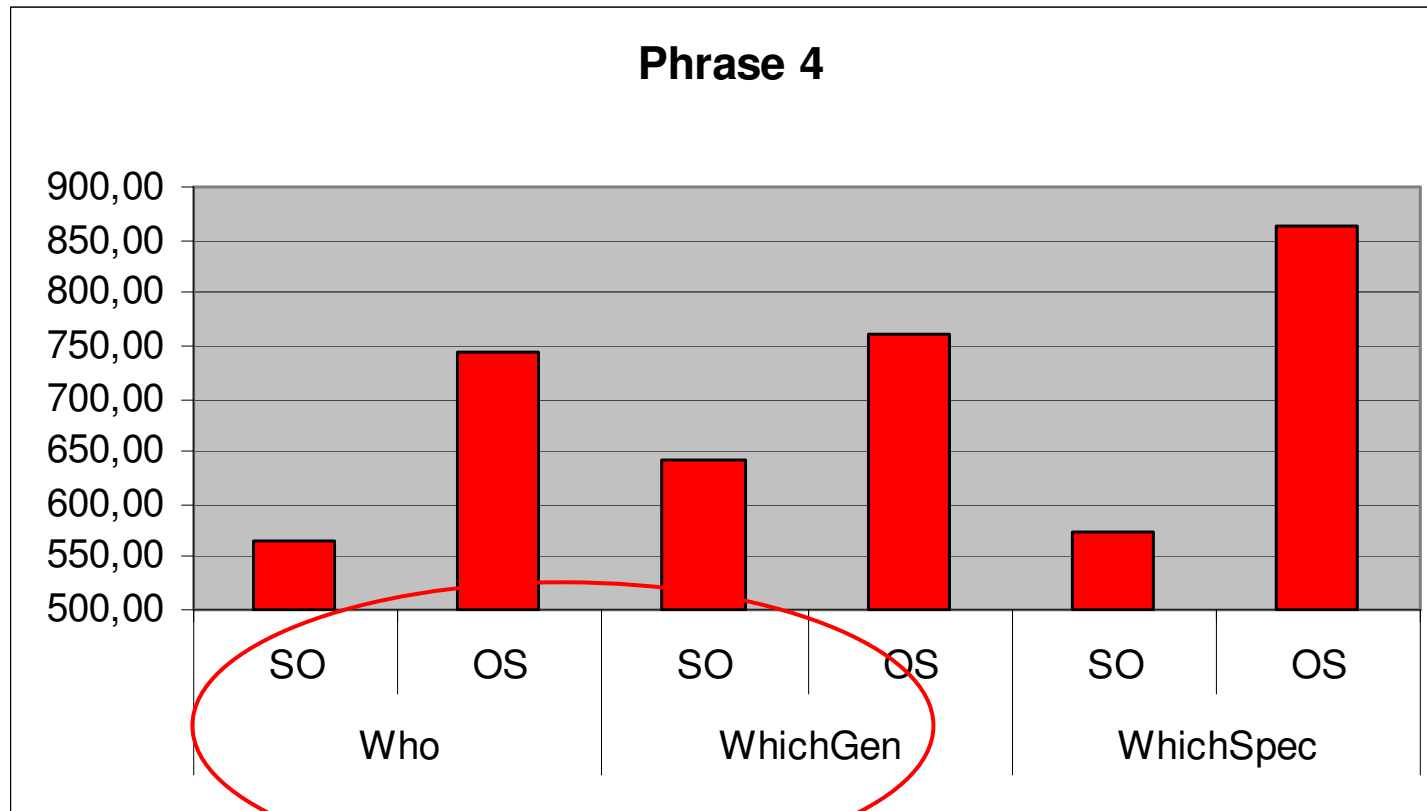


What do I want to know?

- o So the separate question types behave differently?
- o Is this connected to word order complexity?

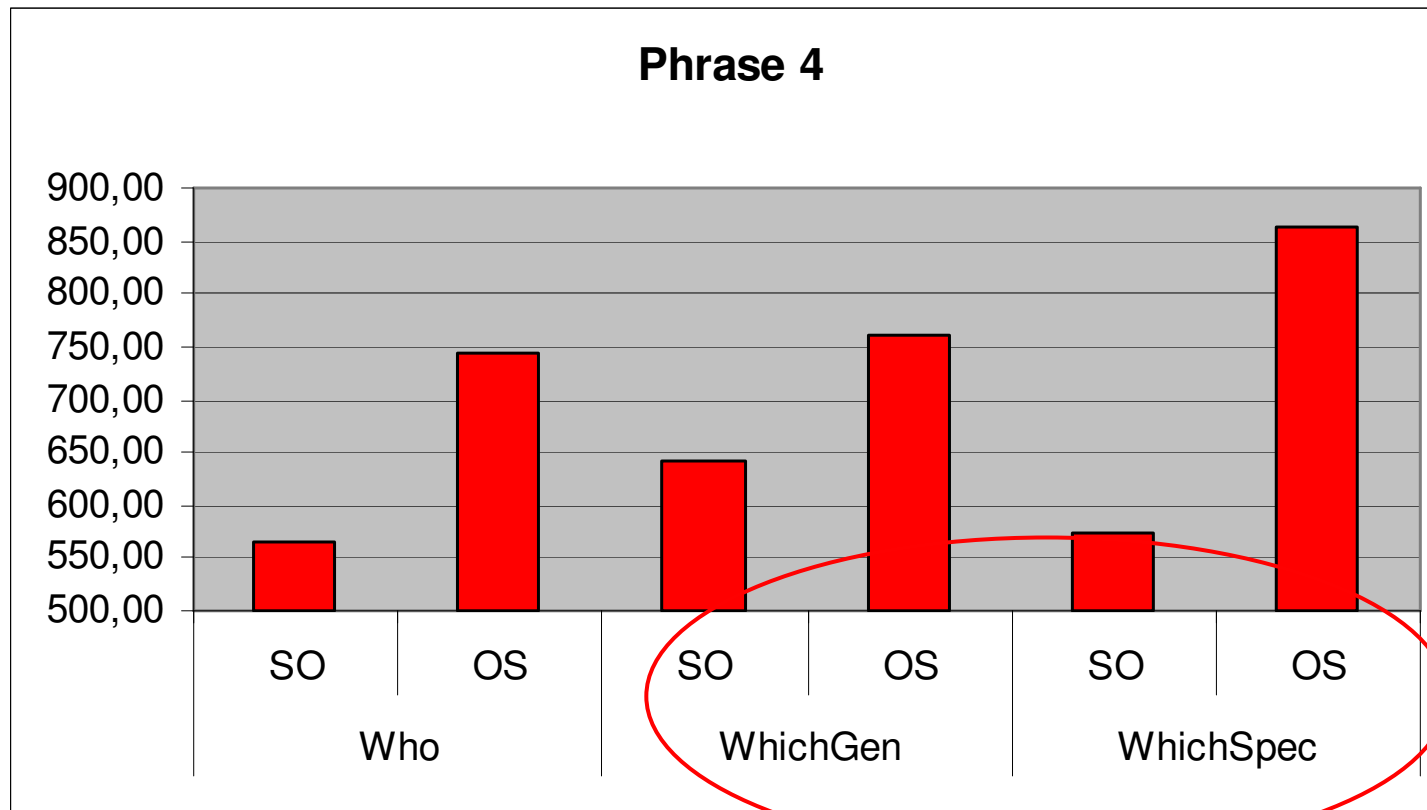
Reading times (5)

o Phrase 4: Participle



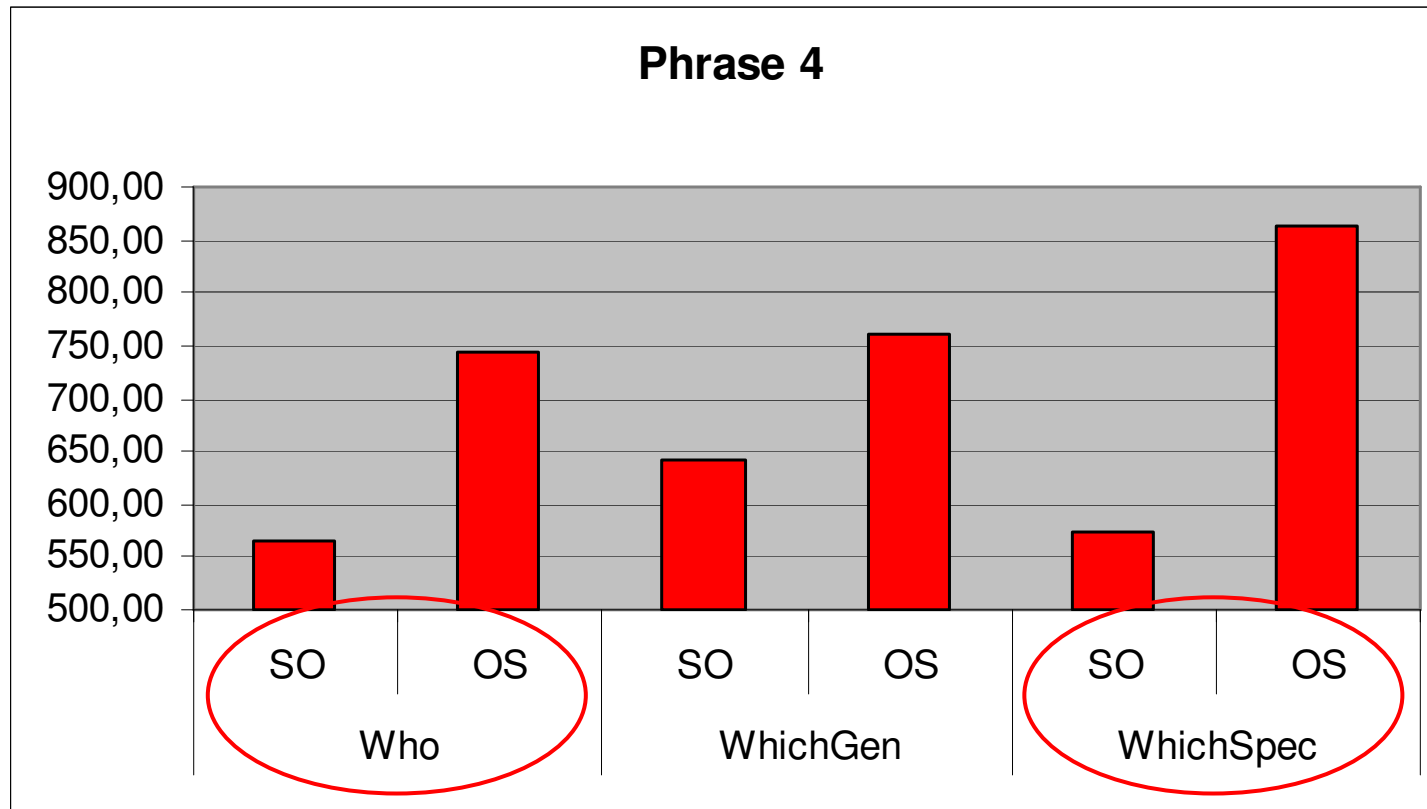
Reading times (5)

o Phrase 4: Participle



Reading times (5)

o Phrase 4: Participle

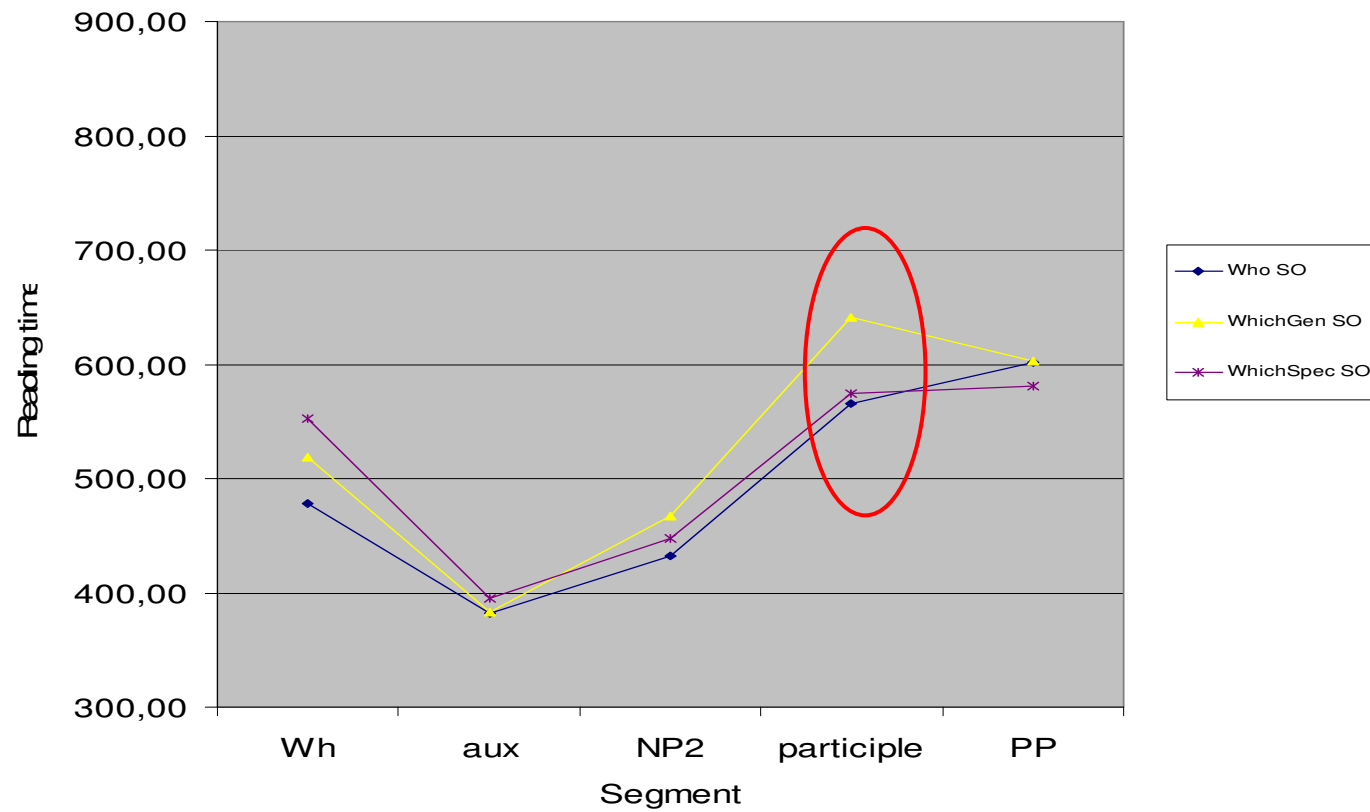


What do I want to know?

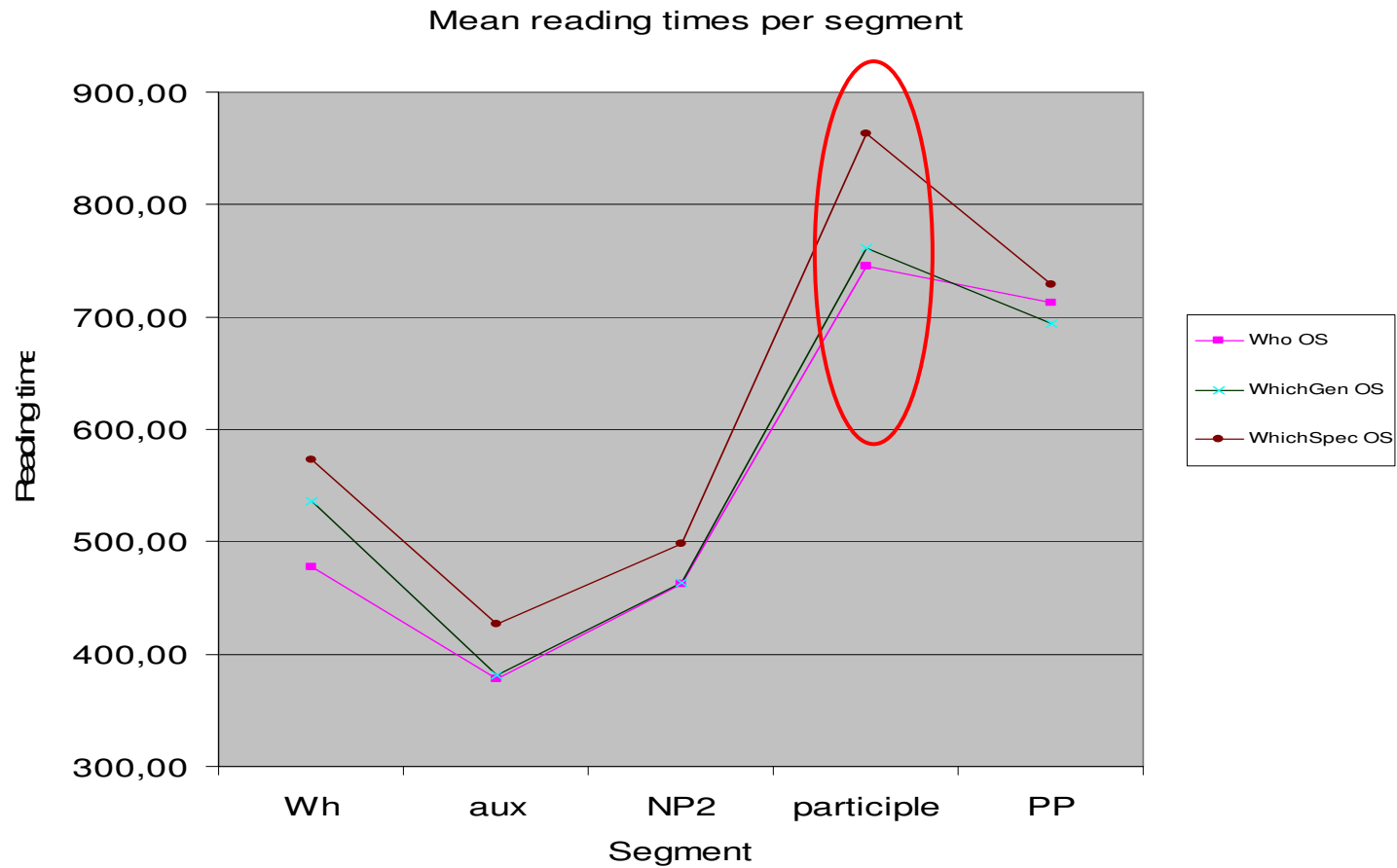
- o So the separate question types behave differently?
 - Who and WhichGen seem to pattern alike, compared to WhichSpec

Reading times SO

Mean reading times per segment



Reading times OS



March 14, 2005

Methodology & Statistics

What do I want to know?

- o Is a different pattern for the WhichSpec condition only connected to word order complexity?
- The increased reading times for WhichSpec are confined to the OS structure-conditions. In SO conditions the average reading times were comparable.

Reading times (6)

o Phrase 5: PP

