

## Overview

### Three approaches to text analysis

- Qualitative Analysis
- Quantitative Content Analysis
- Text Mining

## Introduction to Provalis Research software

### QDA Miner

- Introduction and project management
- Codebook management and manual coding
- Security features and text retrieval tools
- Coding Frequency and Retrieval
- Code co-occurrence and case similarity analysis
- Assessing relationship between coding and variables
- Using the Report Manager and the Command Log
- Performing teamwork
- Miscellaneous Functions

### WordStat

- Content Analysis or Text Mining
- Analyzing words without dictionaries - a text mining approach
- Content Analysis - Principles of dictionary construction
- Importing and exporting data
- Introduction to automatic document classification

## Day 1

## QDA Miner

### Part 1 - Introduction and project management

#### Introduction to CAQDAS using QDA Miner

- The CASE x VARIABLE file structure
- The Mixed-Method approach

#### Quick overview of the work environment

- The four windows - CASE, VARIABLES, CODES, and DOCUMENT
- The menu system

#### Creating of a new project

- Creating a new project from a list of documents
- Creating a new project from an existing data file
- Creating an empty project / defining structure
- Using the document conversion wizard

#### Customizing and personalizing the project

- The PROJECT | PROPERTIES dialog
- The PROJECT | NOTES command

#### Manipulating variables

- Adding a variable - VARIABLES | ADD
- Deleting a variable - VARIABLES | DELETE
- Changing the variable data type - VARIABLES | TRANSFORM
- Recoding the values of a variable - VARIABLES | TRANSFORM | RECODE

- Reordering variables - VARIABLES | REORDER
- Changing variable properties - VARIABLES | PROPERTIES

### **Manipulating cases**

- Add a new case - CASES | ADD
- Deleting cases - CASES | DELETE
- Importing new documents in new cases - CASES | APPEND DOCUMENTS/IMAGES
- Changing the case grouping and description - CASES | GROUPING/DESCRIPTOR

## **PART 2 - Codebook management and manual coding**

### **Creating codes and managing the codebook**

- Creating codes and categories - CODES | ADD
- Modifying an existing code - CODES | EDIT
- Delete existing codes - CODES | DELETE
- Moving codes in the codebook
- Merging codes in the codebook - CODES | MERGE
- Splitting codes in the codebook - CODES | SPLIT
- Importing an existing codebook - CODES | IMPORT

### **Manual coding of documents (versus autocoding)**

- The four basic methods for assigning codes to text segments:
  1. Highlight text segment then drag a code
  2. Highlight text segment then double-click a code
  3. Highlight text segment then select code and button (toolbar)
  4. Drag and drop a code over a paragraph (or a sentence – press ALT)
- Assignment of multiple codes to the same segment (press CTRL)

### **Modifying existing coding**

- Working with code marks
- Viewing coding information
- Adding a comment to a coding -  COMMENT
- Remove a coding -  REMOVE CODING
- Change the code assigned to a text segment -  RECODE TO
- Resizing a segment -  RESIZE
- Consolidating codes - CODES | CONSOLIDATE
- Searching and replacing codes - CODES | SEARCH & REPLACE
- Hiding code marks - CODES | HIDE CODINGS
- Highlighting coded segments - DOCUMENT | CODED TEXT

## **PART 3 – Security features and text retrieval tools**

### **Using backup features**

- Creating a permanent backup - MAINTENANCE | BACKUP | CREATE
- Restoring a backup - MAINTENANCE | BACKUP | RESTORE
- Using the temporary session backup

### **Text retrieval tools (4)**

#### **1. Searching for text - ANALYSIS | TEXT RETRIEVAL**

- Performing a simple text search
- Performing a complex text search (using Boolean and wildcard)
- Performing a thesaurus search
- Using the "search hits" table
- Performing manual coding and autocoding
- Saving to disk or printing the table

#### **2. Retrieving sections in structured documents - ANALYSIS | SECTION RETRIEVAL**

#### **3. Performing a query by example - ANALYSIS | QUERY BY EXAMPLE**

- Finding text similar to a sample text segment

- Providing relevance feedback to improve search results
- Finding text similar to specific coded segments
- Performing a "fuzzy string matching"

#### **4. Performing a keyword search**

- Assigning keywords to codes
- Performing a keyword retrieval on internal codes
- Performing a keyword retrieval on WordStat dictionary files

## **PART 4 - Coding Frequency and Retrieval**

### **Coding frequency**

- Creating a frequency list of all codes - ANALYSIS | CODING FREQUENCY
- Creating a barchart or a pie chart on selected codes
- Customizing the chart

### **Coding Retrieval**

- Performing a simple coding retrieval - ANALYSIS | CODING RETRIEVAL
- Performing a complex search
- Creating a text report
- Creating a new project from
- A shortcut for simple coding retrieval -  | RETRIEVE SEGMENTS

### **Saving and Retrieving Queries**

### **Retrieving a list of comments**

## **PART 5 - Code co-occurrence and case similarity analysis**

### **Analyzing codes co-occurrences - ANALYSIS | CODING CO-OCCURRENCE**

- Hierarchical clustering of codes
- 2D and 3D multidimensional scaling plots
- Using the Proximity plots
- Assessing similarity of cases

### **Analyzing code sequences - ANALYSIS | CODING SEQUENCES**

- Choosing codes and setting minimum / maximum distances
- Using the Sequence matrix
- Searching and coding specific sequences

## **PART 6 - Assessing relationship between coding and variables**

### **Analyzing coding by variables - ANALYSIS | CODING BY VARIABLE**

- Crosstabulating coding frequency by variables
- Setting the content and format of the table
- Computing correlation or comparison statistics
- Comparing frequencies using barcharts or line charts
- Creating and interpreting 2D and 3D correspondence plots
- Creating and interpreting heatmaps

### **A quick overview of graphic coding features**

## **PART 7 - Using the Report Manager and the Command Log**

### **Using the Report Manager**

- Accessing the Report Manager - PROJECT | REPORT MANAGER
- The Report Manager interface
- Appending tables, graphics and quotes
- Moving and organizing items using the table of content
- Editing existing items / adding comments
- Adding empty documents or folders and deleting existing items
- Importing documents, images or tables

- Searching and replacing text
- Exporting results to HTML, Word or RTF files.

### **Using the Command Log**

- Introduction to the command log - PROJECT | COMMAND LOG
- Filtering log entries
- Adding comments to log entries
- Undoing previously performed operations
- Repeating previously performed operations
- Exporting the log table to disk

## **PART 8 - Performing teamwork**

Preparing projects for teamwork - PROJECT | TEAMWORK

- Creating user accounts and setting privileges
- Creating new accounts
- Defining users access rights
- Forcing users to log in
- Creating duplicate copies of a project
- Sending a project by email

### **Merging projects and assessing coding reliability**

- Merging two or more projects
- Planning teamwork for assessing coding agreement
- Adjusting colors of code marks
- Computing coding agreement - ANALYSIS | CODING AGREEMENT
- The codebook and segmentation problems
- Four levels of agreement
- 1. Presence or absence (0 or 1)
- 2. Frequency (0, 1, 2, etc.)
- 3. Coding importance (% of words)
- 4. Coding overlap
- Correcting (or not) for chance agreement.
- Identifying disagreements

## **Day 2**

# **WordStat**

## **PART 1 - Basic Word Statistics and Text Mining**

### **Content Analysis or Text Mining**

#### **Running WordStat from QDA Miner or Simstat**

#### **Analyzing words without dictionaries - a text mining approach**

#### **Data preparation - misspelling and control characters**

#### **Basic word frequency analysis**

- Application of text pre-processing methods
- Exclusion list - use with care
- Lemmatization and stemming - limits and benefits
- Setting upper and lower frequency criteria
- A few additional options
- Numeric and other non-alphabetic characters

- Braces and square brackets
- Random sampling
- Using disk or memory as the working space
- Identifying themes using word co-occurrence analysis**
- Clustering words and measuring their proximity
- Clustering documents based on the words they contains
- Correlation and comparison analysis based on word usage**
- Performing crosstabs and computing statistics
- Comparing words among the sources (document or text variables)
- Correspondence analysis and heatmaps.

## **PART 2 - Content Analysis - Principles of dictionary construction**

### **Introduction to WordStat categorization dictionary**

- Dictionary structure and functions
- Opening, saving, and creating categorization dictionaries
- Creating manually categories of words and phrases
- Principles of dictionary construction - Extracting features
- Identification of technical terms and proper names (persons, places, products)
- Identification of common misspellings
- Extracting phrases
- Creating an initial dictionary - Phrases  technical terms and proper nouns  words
- Adding words manually
- Adding words from tables
- Using the drag and drop editor
- Organizing the dictionary (drag and drop)

### **Applying the dictionary**

- Setting different levels
- Mixing dictionaries with words

### **Validating the dictionary**

- Finding words or phrases with improper meanings using the KWIC list
- WordStat evaluation order – how to use this at your advantage
- Disambiguation methods
- Manual disambiguation
- Disambiguation using phrases
- Disambiguation using rules

### **Improving categorization dictionaries**

- Creating comprehensive dictionaries using the Suggest button.
- Assessing coverage using the keyword retrieval feature

## **PART 3 – Advanced features**

### **Importing and exporting data**

- Exportation of frequency data