The Analysis and Archiving of Heterogeneous Text Documents: Using Support of the Computer Program NUD*IST 4

Christine Plaß & Michael Schetsche

Abstract: On the basis of a current research project we discuss the use of the computer program NUD*IST for the analysis and archiving of qualitative documents. Our project examines the social evaluation of spectacular criminal offenses and we identify, digitize and analyze documents from the entire 20th century. Since public and scientific discourses are examined, the data of the project are extraordinarily heterogeneous: scientific publications, court records, newspaper reports, and administrative documents. We want to show how to transfer general questions into a systematic categorization with the assistance of NUD*IST. Apart from the functions, possibilities and limitations of the application of NUD*IST, concrete work procedures and difficulties encountered are described.

Table of Contents

1. Problem Definition

In the future, one of the greatest problems in qualitative document analysis will be a sharp increase in the number of documents we have to analyze in our projects. A cause for this is the opinion that due to the possibilities of digitization and computer-aided handling, substantially larger text quantities can be analyzed at the same time. The basic idea is to improve the present representativity of the investigation material by a larger number of documents. [1]

The possibilities of digitization and computer-aided analysis prove thereby at the same time as "benedictions" and as "curse" for qualitative document analysis. On one hand it is true that a greater number of documents can be analyzed more completely than it was possible before. On the other hand, the current scientific community expects the handling of very large text quantities—often without an appropriate knowledge how to manage them. Three questions arise at this point:

1. How can text documents be digitized?
2. Which method of text analysis do we prefer?
3. Which computer program shall be used? [2]

In answering the third question, it needs to be decided whether the selected
program is to be used only for the administration of the documents (including
simple look-up and linkage options)—or whether the program must directly
support the document analysis. The largest difficulty here is that we must decide
at the beginning of a research project (or even already during the drawing up of
the appropriate research request) which computer program is most suitable for
the planned project. In order to supply reference points for this decision-making
process, we would like to present some concrete incidents which we experienced
in our research project "Wille und Trieb" ("Will and Drive") with one of these
programs (QSR NUD*IST). [3]

In our project we examined the interpretative patterns of serial killing in Germany
in the 20th Century. The development of these patterns was pursued:

- during a long period (from the German Empire to the Federal Republic),
- by different types of fields (medical-psychiatric, legal, public, political-
  administrative and artistic),
- concerning its effects in five fields of practice. [4]

We leave out the details of the research subject as it is not necessary for the
purpose of this text. What is important to recognize is that a large number of
documents and different types of them needed to be analyzed in the project. The
main difficulty was that the texts came from the entire 20th century and belonged
to most diverse document types. We had to analyze court documents, press
reports, scientific publications and administrative documents. Already a legal
document from Nazi Germany differs in form, style and reasoning substantially
from one from the Federal Republic area. The contrast is even stronger, if one
compares a court report of a west German daily newspaper with the documents
of a federal state parliament debate from the Weimar Republic. Therefore, we
needed a program, which is able to categorize individual text segments based on
super-ordinate questions. The documents had to remain however restorable
thereby in their time- and document-typical specificity. [5]

2. Description of the Program

The computer program QSR NUD*IST (Non Numerical Unstructured Data
Indexing Searching & Theorizing) is used so far only rarely in the German-
speaking countries. In our project, we used the fourth version of NUD*IST—and
that is the version we refer to in this text. In the meantime, the fifth version has
become available (October 2000) with many different functions
(http://qsr.latrobe.edu.au/products/n5.html). [6]

We decided to use NUD*IST for three essential reasons:

- The hierarchical structure of the category pattern (*Index Tree*) enables the
  user to maintain an outline with a large number of categories (in our case
  more than 100).

- With the help of a function to merge different versions of the processed texts (*Merge*), several coworkers can code documents at the same time (and also are able to change the coding pattern).
- External documents can be integrated into the document sample. [7]

In order to be able to understand the advantages and disadvantages of the program, it is necessary to glance at its fundamental functions. One operates NUD*IST in four—usually parallel opened—windows: *Document Explorer*, *Node Explorer*, *Node* or current *Document* and *Palette*.
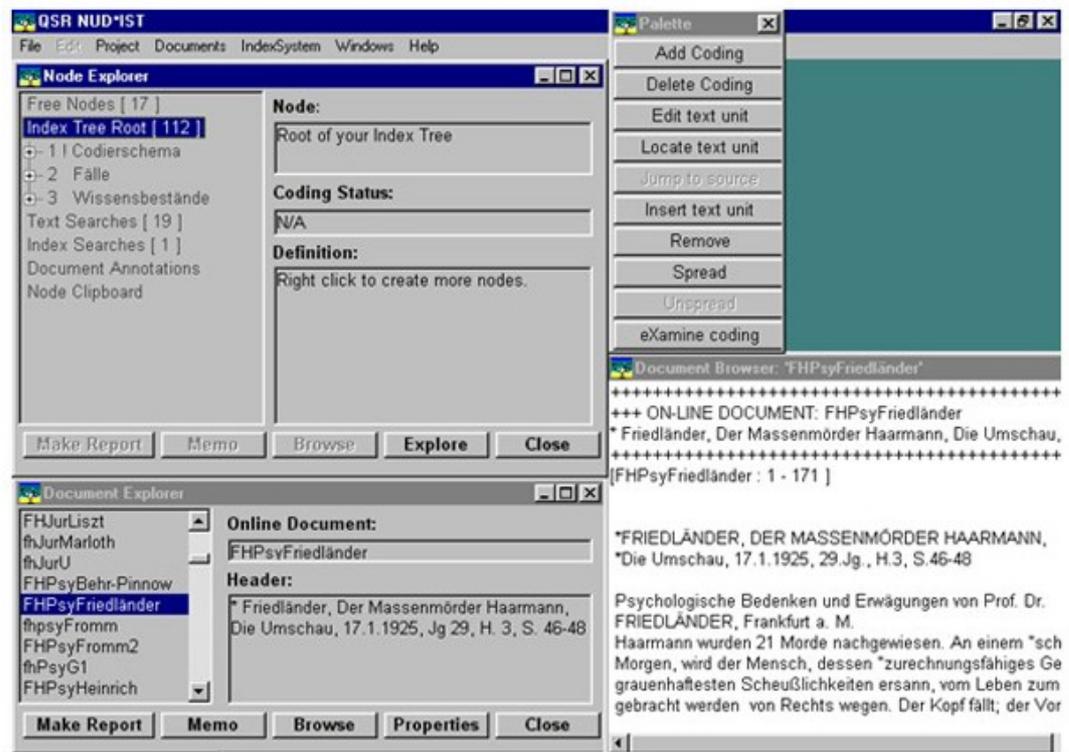
Fig. 1: Screenshot of NUD*IST (document from the author's project) [8]

The *Document Explorer* contains a list of imported documents (whether already coded or not). For each document, the *Header* is displayed. It covers freely definable information of the document (in our case among other things the specification to the source of text). In *Document Explorer,* we can create additional *Memos* (arbitrarily long notes to documents) and *Reports*. Both can be created, changed and printed out either in the text editor by NUD*IST or in a compatible text processing program (for example MS-Word). [9]

The *Node Explorer* displays the category pattern developed by the users (*Index Tree*), free categories (*Free Nodes)* and search results. Such results can be stored either separately or be inserted as (theoretical) categories (so-called *Nodes*) into the *Index Tree*. *Free Nodes* are categories, which cannot be inserted into the *Index Tree* yet. *Memos* and *Definitions* can be written for each node. Additionally, the text segments listed in a *Node* can be examined for additional

codings. The text can be coded and re-coded in the *Node* as in the *Document* at each point in time. Such codings can be checked and changed very comfortably. This proved even at the start of our project as extraordinarily helpful, when we had to further develop the coding pattern. [10]

The functions of *Make Report* provide an overview of the coded text segments and overlap with other *Nodes*. The identified texts can be processed to a larger extent and copied into text processing programs. (The size of the categories however, often exceeds the extent maximum for the internal text editor. Results of the work can only be opened then in the "READ only" mode.) [11]

There are several possibilities of using NUD*IST to support the development of a category pattern: Categories can be generated in process of the analysis of documents or be inserted from the beginning as category pattern in NUD*IST. We decided for a middle course and inserted the first design of the category pattern into the *Index Tree* before beginning the analysis. This pattern was continuously developed further in course of the coding process. Codings were added, deleted, merged and reformulated. Because of the specific programming of the code allocation and code control segments of the program, such operations can to be completed without a large expenditure of time. [12]

We followed the suggestion of the user manual (QUALITATIVE SOLUTIONS AND RESEARCH 1997) and stored—apart from the contentwise categories— also the examined "cases" completely as *Node*. Since each of these "cases" consists of a multiplicity of documents, it proved favorable to introduce additionally a third coding level: the kind of knowledge, to which the document belongs. For example, the text of a court decision was sequentially coded first of all under the category pattern, entered secondly completely under the appropriate case and thirdly under the node "legal knowledge". This three-fold coding serves not only the clear archiving of the documents, but represents also the basis for complex search strategies. [13]

Finally, the *Palette* covers the most diverse functions to work on opened *Nodes* and *Documents*:

- correction of transfer errors,
- coding and decoding of text passages,
- orientation in the text body,
- inserting text and blank lines,
- display of different codings,
- display of the total document or the surrounding text of a coding. [14]

3. Advantages and Disadvantages in Practice

The terminology of the program NUD*IST is orientated clearly at the method of Grounded Theory (GLASER & STRAUSS 1967; STRAUSS & CORBIN 1996). KELLE (1997) suggests that this should not be misunderstood as a direct connection between the program and the method. (For information on this, see MACMILLAN & MCLACHLAN (1999), who tested NUD*IST with the methods of the qualitative content analysis and the discourse analysis). The advantages and disadvantages of NUD*IST regarding to different methodical applications can be checked only empirically. For our analysis, which was orientated at the principles of the Grounded Theory, the program proved to be well suitable. However, it is not possible to write Memos to individual text segments. To support a detailed analysis of large documents is so far unsatisfying. Another problem—it was rather created in the method we used—exists in the large expenditure of a continual re-coding of the documents, if the category pattern is only developed in the run of the analysis. Therefore, we decided to develop a provisional category pattern from secondary literature and hypotheses and to insert this in NUD*IST. [15]

The basic functions of NUD*IST can be learned with the help of the manual and/or the tutorials relatively fast, although unexpected problems occurred again and again with simple processes. (For the problems that occur often, you can read the NUD*IST mailing list, the "QSR forum": http://qsr.latrobe.edu.au/resources/forum.html.) Basic functions available in the program are sufficient to archive, sort and code large document quantities as well. The more complicated search functions become relevant, only if a larger quantity of the data is coded (and the user is a little bit "familiar" with the program). An exception is the function Merge: The merging of different versions of the coding was always error free, except in (our) case of a fast emerging Index Tree, merging was complicated and time-consuming. The working procedures must be carefully organized in order to avoid the loss of relevant data. In the long run it seems more practicable, not to code parallel on different computers. [16]

Another disadvantage of the program is the necessity to pre-format the documents in a complex way before insertion: They have to be provided with specific paragraph labels in order to mark the text units. Although the smallest text unit that can be entered is theoretically a word, in practice such an allocation raises insurmountable obstacles. So finally we had to use individual lines or records as analysis units. BUSTON (1997) used NUD*IST in a study based on interviews with people who suffer from chronic diseases. She reported setting paragraph labels after each question and each response; this is really practicable only with shorter documents or interviews. [17]

NUD*IST offers almost unrestricted possibilities of representing documents or codings on screen: in summary or in details, in comparison or as a list, with given text or with given lines. In contrast to this, the possibilities for printouts are circumstantial and reduced. It is particularly problematic that the printout of a document with listed categories on the edge is only possible with restrictions. This is because of the fact that for each category of a document a letter is assigned

and so the range of these categories is limited to 26. With larger documents this causes difficulties. But even if this was not the case, the creation of a Report with many Nodes was time-consuming (because each Node had to be indicated before). [18]

A further disadvantage of NUD*IST, often mentioned in the literature, is the hierarchical coding pattern, which limits the possibilities to interconnect text segments and codes. We can agree with this criticism, however, only partly. Indeed text segments cannot be interconnected directly together, but only in codes. But this seems not to be a disadvantage in every case because there is always a reason to interconnect two text segments. The result can be recorded in a Node or Free Node. The connections of the codes themselves are possible in the Index Tree, however hierarchically only, but you can produce each conceivable relation or cross-setting of categories by using the elaborated search functions. A disadvantage is that these linkages cannot be plotted. For this application additional programs ("Decision" or "Inspiration Explorer") must be purchased. [19]

4. Result

BARRY (1998) compared NUD*IST 4 and ATLAS/ti 4 and in her opinion NUD*IST is suitable for complex and long-term based text analyses. For BARRY the most important advantages of the program are the "sophisticated" search functions which can be used in many different ways to categorize the data: *"any information about the project can used as a way to sort and view the data"* (BARRY 1998, paragraph 9.2). [20]

We agree with BARRY's judgement. In summary, one can say that NUD*IST substantially facilitates the archiving and coding of the data. On the other hand, it "devours" a lot of time resources. This is due to the fact that in order for the program to accept the documents, they must be edited and this process of editing is time-consuming. In addition: The process of the digitization and text recognition requires a great deal of time or is even impossible for many older documents. NUD*IST can also process non-digitized (external) documents, however only as references. The majority of the sophisticated coding and search functions is available only with documents, which were imported completely in NUD*IST. Therefore, an acquisition of NUD*IST is only worthwhile, if the predominant part of the documents can be digitized. [21]

References

Barry, Christine (1998). Choosing Qualitative Data Analysis Software: Atlas/ti and Nudist Compared. *Sociological Research Online*, *3*(3), http://www.socresonline.org.uk/socresonline/3/3/4.html.

Buston, K. (1997). NUD*IST in Action: Its use and its Usefulness in a Study of Chronic Illness in Young People. *Sociological Research Online*, *2*(3), http://www.socresonline.org.uk/socresonline/2/3/6.html.

Glaser, Barney & Strauss, Anselm (1967). *The discovery of grounded theory.* Chicago: Aldine.

Kelle, Udo (1997). Theory Building in Qualitative Research and Computer Programs for the Management of Textual Data. *Sociological Research Online*, *2*(2), http://www.socresonline.org.uk/socresonline/2/2/1.html.

MacMillan, Katie & McLachlan, Shelley (1999). Theory-building with Nud.Ist: Using Computer Assisted Qualitative Analysis in a Media Case Study. *Sociological Research Online*, *2*(2), http://www.socresonline.org.uk/4/2/macmillan_mclachlan.html.

Qualitative Solutions and Research (1997). *QSR NUD*IST 4: User Guide.* Victoria: Scolari.

Strauss, Anselm & Corbin, Juliet (1996). *Grounded Theory: Grundlagen Qualitativer Sozialforschung.* Weinheim: Beltz.

Authors

Dipl. Sozialwiss. *Christine PLASS* is Sociologist at the EMPAS (Institute for empirical and applied sociology), University of Bremen, Germany. Her current fields of interest are methods of qualitative research, sociology of knowledge, gender and criminality.

Contact:

Christine Plass

E-mail: plass@uni-bremen.de

PD Dr. *Michael SCHETSCHE* is an Assistant Professor at the EMPAS (Institute for empirical and applied sociology), University of Bremen, Germany. Currently he has research interests in social problems, the sociology of knowledge, and especially the coming "sociology of cyberspace". He is the author of "Die Karriere sozialer Probleme" (The rise of social problems, 1996) and "Wissenssoziologie sozialer Probleme" (Knowledge sociology of social problems, 2000).

Contact:

Michael Schetsche

E-mail: mschet@uni-bremen.de
URL: http://www1.uni-bremen.de/~mschet/

Citation