The use of IRAMUTEQ software for data analysis in qualitative research*

O uso do software IRAMUTEQ na análise de dados em pesquisas qualitativas
El empleo del software IRAMUTEQ en el análisis de datos en investigaciones cualitativas

Marli Aparecida Rocha de Souza1,2, Marilene Loewen Wall1, Andrea Cristina de Morais Chaves Thuler3, Ingrid Margareth Voth Lowen4, Aida Maris Peres1

How to cite this article:

ABSTRACT
Objective: To describe the use of IRAMUTEQ software in qualitative data analysis.
Method: A description for using a software program as a tool to support data processing in qualitative research, carried out in 2015. Data collection was done through interviews using a semi-structured script. Results: Twenty-one (21) companions participated in the study. The five classes resulting from data processing by the software allowed for analysis and interpretation of the nursing performance with the parturient as an active participant in the process of labor and delivery, and in the role of companion in this period. Conclusion: The use of IRAMUTEQ software as a tool for processing qualitative data through the Descending Hierarchical Classification emerged classes and a connection between them, which further allowed for secure and credible data analysis. It is necessary to explore the other possibilities of using this tool in greater depth.

DESCRIPTORS
Software; Nursing Research; Qualitative Research; Obstetric Nursing.

1 Universidade Federal do Paraná, Programa de Pós-Graduação em Enfermagem, Curitiba, PR, Brazil.
2 Faculdade Dom Bosco, Departamento de Enfermagem, Curitiba, PR, Brazil.
3 Secretaria Municipal de São José dos Pinhais, São José dos Pinhais, PR, Brazil.
4 Secretaria Municipal de Curitiba, Curitiba, PR, Brazil.
INTRODUCTION

Qualitative research in the health area has been consolidated in recent decades, and its diffusion has been intensified in the Nursing area. Researchers who choose this approach seek to understand the participants’ perspective, interpreted according to researchers’ experiences(1).

Thus, since the 1980s the creation of computational programs that offer assistance in qualitative research data analysis have increased in quantity and possibilities for researchers. These programs are known as CAQDAS (Computer Aided Qualitative Data Analysis Software), and the debate over its use has grown, although its efficiency in the management and recovery of qualitative data has consensus(2).

Among the advantages in the data analysis process using software are help in organizing the data, separation of information, an increase in the efficiency of the process and the ease in locating text segments, in addition to more agility in the codification process compared to that performed manually(3).

Among several software programs available are some free programs which have resulted from a movement for sharing technological knowledge based on principles such as freedom of use, copying, modification and redistribution. These fundamentals are made possible by the distribution of the programs’ source code, transforming them into public goods(4).

One of these free software programs is IRAMUTEQ (Interface de R pour les Analyses Multidimensionnelles de TEXTes et de Questionnaires) created by Pierre Ratinou and, until 2009 it was only available in the French language; however, it currently has complete dictionaries in several languages. IRAMUTEQ is developed in the Python language and uses functionalities provided by the statistical software R. It began to be used in Brazil in 2013 in social representation studies, although other areas have also been using it and contribute to the dissemination of the various possibilities of qualitative data processing, since it allows different forms of statistical analysis of texts produced from interviews and documents, among others(5–10).

Descending Hierarchical Classification (DHC) is one of the analyses performed by IRAMUTEQ software and has already been proposed by the software ALCESTE (Analyse Lexicale para Context d’un Ensemble de Segments de Texte). In addition to DHC, IRAMUTEQ includes other forms of analysis such as classic textual analyses, analyses of specificities, similarity analysis and a word cloud. Both software programs are considered data processing tools rather than a research method, which makes their results instruments for exploration, search and association in research material(1,5,10–12).

Faced with the challenge of understanding and describing the use of this tool and the limitation of published materials on the use of IRAMUTEQ software in health and nursing research, this article had the objective to describe the use of IRAMUTEQ and one of its forms of qualitative data processing. The research project in which data analysis used IRAMUTEQ was entitled: Experience of the parturient’s assistant in the delivery process (Vivência do acompanhante da parturiente no processo de trabalho de parto e parto) which sought to discuss women’s health in the context propagated in institutions in relation to the presence of a companion during labor and delivery.

METHOD

This study describes the use of a software program as a tool to support data processing in qualitative research. The research was carried out in a teaching hospital in the southern region of Brazil, which is part of the Stork Network (Rede Cegonha) and has the title of a Child Friendly Hospital since 1995. Data collection was performed through interviews for which a semi-structured script was applied with 21 companions of puerperal women, above 18 years of age who had witnessed labor and delivery regardless of their gender or degree of kinship. All the interviews were recorded and transcribed, having an average duration of 40 minutes. They focused on the meaning given to the situation experienced, without the influence of the researcher’s conceptions(11).

Data collection was performed in January 2015, approved on 10/16/2012 by the Ethics Committee of a Brazilian University, under registration number 120.892/2012, CAAE 08200912.1.0000.0096, in accordance with Resolution 466/2012 of the National Health Council. This study is a subproject linked to a primary research entitled “Care in the puerperium: application of a nursing care model” (O cuidado no puerpério: aplicação de um modelo de cuidado de enfermagem), and it started after each participant signed the clear and Informed Consent Form (ICF).

DHC was used for the textual analysis of the research, in which segments of text are classified according to their respective vocabularies and summarized, mostly, into about three lines, and variation occurs according to the transcription by the researcher and the size of the corpus, which is characterized by the set of text that needs to be analyzed. The set of these segments is then divided according to the frequency of the reduced forms(9).

This interface based on the original corpus allows for retrieval of the text segments and their association, which then enables grouping statistically significant words and the qualitative data analysis, in which each interview is known as the Initial Context Unit (ICU). The Elementary Context Units (ECU) or text segments that make up each class are obtained from the ICU and have a similar vocabulary between them, differing from other ECU classes(5–9).

RESULTS

The use of a computer program provides the advantage of coding, organizing and separating information, which allowed for rapidly locating the entire text segment used in the qualitative writing.
Three steps are taken to perform the DHC: preparation and coding of the initial text, the descending hierarchical classification performed by the data processing, and interpretation of the classes. Preparing the initial text in the qualitative research means transcribing the interviews, which is a set of texts that constitutes the corpus of analysis; it is recommended that this corpus has at least 20 texts\(^{[5-11]}\). Therefore, the 21 interviews originated 21 texts organized into a single file, which originated 21 ICUs. Each are separated by a command line, comprising only one variable (n) which is chosen according to the number given to each participant (n\(_1\), n\(_2\) up to n\(_{21}\)). After the transcription was carried out using the LibreOffice Writer of the LibreOffice.org package, the file was saved as a text document that uses UTF-8 standard encoding (Unicode Transformation Format 8 bit codeunits). The questions were not included, only the answers were kept in full and referencing the question.

Next, a revision of the entire file was conducted by correcting typing and punctuation errors, standardizing acronyms and joining compound words together; for example, the term “nursing assistant”, which if included without separation by an underline instead of spaces is processed by the system as two different words. All observations must be made carefully by the researcher so that the processing is done using a greater number of compound words in the corpus.

Text segments presented in each class were obtained from statistically significant words based on the corpus, thus enabling qualitative data analysis to be performed. Corpus processing was performed in 23 seconds and 964 ECU were classified, of which 873 were used corresponding to 90.56% of the total corpus. An index of 75% or more is considered as efficient use of ECUs\(^{[9]}\).

The program uses the chi-square test (\(\chi^2\)) for creating a dictionary of words, which reveals the associative strength between words and their respective class. This associative strength is analyzed when the test is greater than 3.84, representing \(p<0.0001\). A lower chi-square value represents a lower relationship between the variables\(^{[12-13]}\).

The classes are formed according to the relationship of the several processed ICUs and which present homogeneous words. The ICUs are grouped based on occurrences of the words according to their roots for classification and the relationship of the classes, giving rise to the ECUs which result in the creation of a dictionary with reduced forms by using the chi-square test (\(\chi^2\))\(^{[5,12-15]}\).

After processing and grouping according to word occurrences, DHC creates a class dendogram. In addition to presenting the classes, this image demonstrates the connection between them since they are associated with one another. Each class has a different color and the ECUs of each of them have the same color of the class, as shown in Figure 1.

![Figure 1 – Dendogram of the classes provided by IRAMUTEQ software – Curitiba, PR, Brazil, 2015.](image-url)

Reading the relationship between classes performed in this step is carried out from left to right. The corpus in the dendogram was divided into two subcorpus. In the first, class 5 was comprised by 165 ECUs, which corresponded to 18.95% of the total. The same subcorpus had a second subdivision, which comprised Class 3 with 174 ECUs and Class 2 with 191 ECUs, in which 19.93% corresponded to class 3 and 21.88% to class 2 from the total ECU. From the other subcorpus, class 1 was obtained with 189 ECUs, corresponding to 21.65% of the ECUs, and Class 4 consisting of 154 ECU, which accounts for 17.6% of the ECUs of the total corpus. A list of words generated from the chi-square test was created (\(\chi^2\))\(^{[5,10]}\) for each class.

The software provides more than one type of dendrogram for the chosen analysis type, and it keeps the ECUs available so that the researcher can go back to them to read and understand the results, and to choose a title for each class at any time so that it represents the central interpreted theme based on the ECUs which each of them compose (Figure 2).
The use of IRAMUTEQ software for data analysis in qualitative research

The dendrogram allows visualization of the words that obtained the highest percentage of the average frequency between them and different between them. Through use of the chi-square ($\chi^2$), this dictionary of words enables an analysis of the words that presented a value greater than 3.84 and $p<0.0001$.

After the data processing, the five classes provided by the IRAMUTEQ software in which the ECU's grouped in each class are comprehensively read in order to understand and name each class.

This data processing stage was followed by the data analysis phase, which is based on the precepts of a qualitative research approach guided by six steps(3): Step 1 – organizing and preparing the data for the analysis, a step carried out according to the guidelines for creation of the corpus; Step 2 – reading all the data, with re-readings for evaluating the transcribed content; Step 3 – a detailed analysis with the coding process carried out by the IRAMUTEQ software, which created the dictionary of words; Step 4 – using the coding process to describe the scenario or people and the categories or themes for analysis by evaluating all classes presented in the dendogram and a new round of listening to the interviews; Step 5 – information on how the description and themes are represented in the qualitative narrative, supported by the literature consulted after an analysis of the categories; and Step 6 – extraction of the meaning of the data, and presentation of the results by the researcher after the analysis through their personal interpretation, sustained in the literature.

By reading the highlighted words and their insertion in the text segments, it was possible to achieve the primary research objectives aimed at verifying the connection established between the nursing team and the companion, and valuing them as a protagonist member of the labor and delivery process, as well as how this welcoming influences more active participation. Class 1 explained the relevance of the health professional’s awareness of the importance of the companion’s presence to the parturient. Class 2 corroborated the need for not only welcoming the parturient, but also their companion. Class 3 demonstrated the interest of the companion in choosing to be present in the labor and delivery process. Class 4 emphasized that the information provided by the health team and the companion, and their companion during labor and delivery, and was subdivided into classes 2 and 3 which evidenced the relationship between birth and the previously created bond by the companion’s willingness to be present in the labor and delivery process. Classes 1 and 4 were respectively characterized by valuing the companion through the information provided by the health team and their potential during the prenatal period and the hospital stay, and they were all associated to class 5.

Figure 2 – Dendogram with the percentage of ECU in each class and words with greater chi-square ($\chi^2$) provided by the IRAMUTEQ software – Curitiba, PR, Brazil, 2015.
DISCUSSION

From observation of the factorial representation (Figure 3) it was possible to observe that class 3, “The participation of the companion by choice”, and class 2, “The experience perceived by the companion and welcoming as an influencing factor”, seem to be interconnected, which evidenced the relationship between birth and family bond as characterized by companions’ willingness to be present, and how this was highlighted when they were welcomed by the health team upon arriving at the institution.

Class 5, “Knowledge and actions performed”, stands out as the class which was further and the least interconnected with the others, again corroborating the research results; this shows that when the companions performed more actions in addition to providing support and holding their hand, these actions were mostly the result of individual wills.

Class 4, “Prenatal care as an opportunity to prepare the companion”, refers to the participation and the guidance received by the companion during the prenatal period, and which is demonstrated as a class that has remained distant or minimally related to welcoming according to the figure presented by the software. Class 2, “The experience perceived by the companion and the welcoming as an influencing factor”, and Class 1, “Information as a strategy for valuing the companion’s participation”, demonstrate (due to their distancing) the need of interaction between the companion and the health team at the institution that welcomes them.

This factorial representation was used in a study on the social representations focused on the theme of AIDS and health care policies aimed at women’s health in situations of liberty deprivation in prison, which also analysed the themes through the words featured in the four axes. The most evident words were related to misunderstandings, the transmission forms and the disease itself, which generates discrimination and the need for health education in the prison system(17).

![Factorial representation](image-url)

Figure 3 – Factorial representation provided by IRAMUTEQ software – Curitiba, PR, Brazil, 2015.
El uso de IRAMUTEQ software para el análisis de datos cualitativos

**RESUMEN**

**Objetivo:** Describir el empleo del software IRAMUTEQ en el análisis de datos cualitativos. **Método:** Descripción de la utilización de un software como herramienta de apoyo al procesamiento de datos en la investigación cualitativa, realizada en 2015. La recolección de datos se llevó a cabo mediante entrevistas, en las que se aplicó un guión semiestructurado. **Resultados:** Participaron en la investigación 21 acompañantes. Las cinco clases resultantes del procesamiento de datos por el software permitieron el análisis e interpretación de la actuación de la enfermería junto a la parturienta como participante activa en el proceso de trabajo de parto y parto, y del papel del acompañante en dicho período. **Conclusión:** El uso del software IRAMUTEQ como herramienta en el procesamiento de datos cualitativos, mediante la Clasificación Jerárquica Descendente, de la que emergieron las clases y la conexión entre ellas, permitió el análisis de los datos con seguridad y credibilidad. Se hace necesario explorar con mayor profundidad las demás posibilidades de uso de esa herramienta.

**DESCRITORES**

Software; Pesquisa em Enfermagem; Pesquisa Qualitativa; Enfermagem Obstétrica.

**REFERENCES**


