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 Souza[1,2], **Marilene Loewen Wall**[1], **Andrea Cristina de Morais Chaves Thuler**[3], **Ingrid Margareth Voth Lowen**[4], **Aida Maris Peres**[1]

* Extracted from the dissertation: "Vivência do Acompanhante da Parturiente no Processo de Trabalho de Parto e Parto", Programa de Pós-Graduação em Enfermagem, Universidade Federal do Paraná, 2015.

[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.

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.

Corresponding author:
Marli Aparecida Rocha de Souza
Av. Pref. Lothário Meissner. 632,
3° andar – Jardim Botânico
CEP 80210-170 – Curitiba, PR, Brazil
marlirochasouza2@gmail.com

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 the organization and 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 Ratinaud, 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[1].

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[5].

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[5-6].

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[5].

The program uses the chi-square test (χ^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 (χ^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.
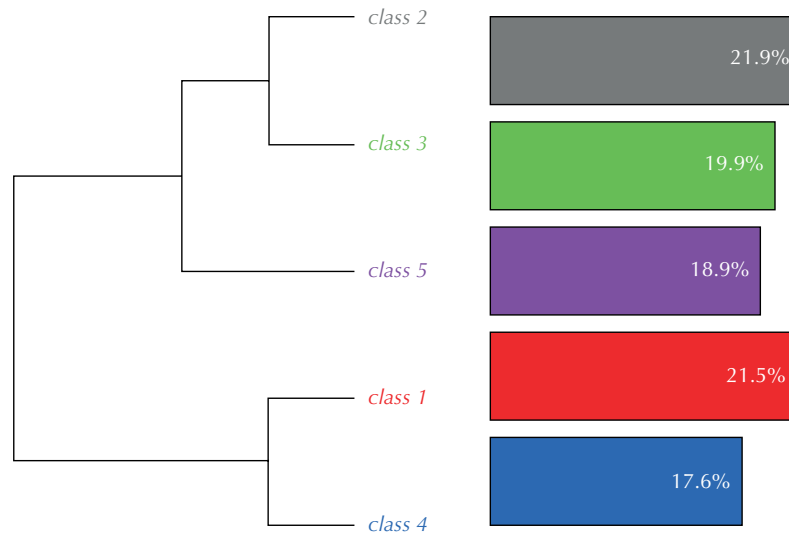

Figure 1 – Dendogram of the classes provided by IRAMUTEQ software – Curitiba, PR, Brazil, 2015.

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 (χ^2)[5,16] for each class.

The software provides more than one type of dendogram 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).
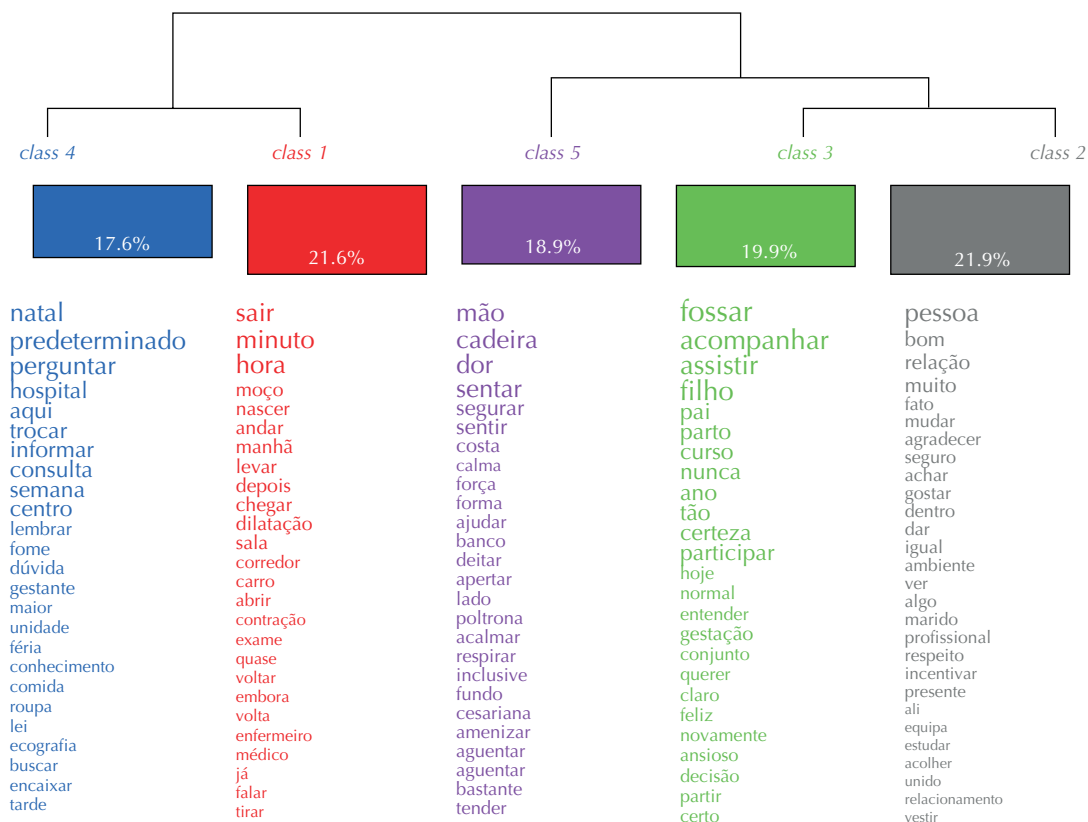
Figure 2 – Dendogram with the percentage of ECU in each class and words with greater chi-square (χ^2) provided by the IRAMUTEQ software – Curitiba, PR, Brazil, 2015.

The dendogram 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 (χ^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 ECUs 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[1]: 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 should start in the prenatal period, so that the companion's participation is carried out with greater security. Finally, Class 5 revealed that the information given in advance made it possible for the companion to take more action during the labor and delivery process.

Class 5 is related to the actions performed by the 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. 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 were associated to class 5.

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].

Figure 3 – Factorial representation provided by IRAMUTEQ software – Curitiba, PR, Brazil, 2015.

Therefore, the factorial representation provided by the IRAMUTEQ software confirmed the interconnection of classes in comparison with those performed in the research during the qualitative data analysis, and enabled recovery of the context in which the words were associated with statistical significance[17].

Several other studies in the primary care setting have used the DHC analysis provided by the software. One of them evidenced the nursing consultation as an innovation when implementing the Family Health Strategy (FHS) linked to the adequacy of services, which promoted a recovery of autonomy of primary care professionals[18].

Another study analyzed the service provided by professionals of the FHS health team to patients with Diabetes mellitus, pointing out that this is a service which promotes medication valorization and food imposition in detriment to and devaluing individual reality, which characterizes fragmented care[19].

A study on social representations used processing by the DHC method to describe rejuvenation practices used by men and women between the ages of 40 and 60, concluding that they used physical activity and mental exercises as a way to delay aging[20].

Another processing method used on a smaller scale provided by the software is the word cloud, which was evidenced in a study that analyzed the perception of professionals working in street clinics, highlighting the need to promote articulated attention between knowledge, practices and subjects[9].

CONCLUSION

Based on the approach of the IRAMUTEQ software as a tool for qualitative research data processing in the health area, we can highlight the need for exhaustive readings, discussions and also the development of skills for its use. By describing its use in a detailed manner, this study intended to highlight some relevant aspects and contribute to disseminating this tool, which requires knowledge and skill, in addition to not disregarding the importance of the researcher's analysis and the ethical aspects inherent to qualitative research.

Previous studies which refer to software with possibilities for processing and data analysis in qualitative research have significantly contributed to the learning and mastery of using IRAMUTEQ. However, the scarcity of published articles describing other forms of analysis in addition to the DHC is still a noted limitation.

We can emphasize that all research steps require seriousness and knowledge which encompass the choice of the qualitative research approach and the quality of the instrument used to perform the data collection interviews, so that the produced corpus enables better processing and presentation of the classes. In this sense, the researcher continues to be the research driver and their role is valued by IRAMUTEQ software, which enables interpretation of results which have already been processed with scientific rigor. The commitment to qualitative research and its social impact is a great responsibility, and the researcher is one of its protagonists in producing quality knowledge.

RESUMO

Objetivo: Descrever o uso do sof*twar*e IRAMUTEQ na análise de dados qualitativos. **Método:** Descrição da utilização de um *software* como ferramenta de apoio ao processamento de dados na pesquisa qualitativa, realizada em 2015. A coleta de dados foi realizada por meio de entrevistas, nas quais foi aplicado um roteiro semiestruturado. **Resultados:** Participaram da pesquisa 21 acompanhantes. As cinco classes resultantes do processamento de dados pelo *software* permitiram a análise e a interpretação da atuação da enfermagem junto à parturiente como participante ativa no processo de trabalho de parto e parto, e do papel do acompanhante nesse período. **Conclusão:** O uso do *software* IRAMUTEQ como ferramenta no processamento de dados qualitativos, por meio da Classificação Hierárquica Descendente, da qual emergiram as classes e a ligação entre elas, permitiu a análise dos dados com segurança e credibilidade. Faz-se necessário explorar com maior profundidade as demais possibilidades de uso desta ferramenta.

DESCRITORES

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

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 y la interpretación de la actuación de la enfermería junto a la parturienta con participante activa en el proceso laboral 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.

DESCRIPTORES

Programas Informáticos; Investigación en Enfermería; Investigación Cualitativa; Enfermería Obstétrica.

REFERENCES

1. Creswell JW. Projeto de pesquisa. Porto Alegre: Artmed; 2010.

2. Paula MC, Viali L, Guimarães GTD, Godoy AS. A pesquisa qualitativa e o uso de CAQDAS na análise textual: levantamento de uma década. Internet Latent Corpus J [Internet]. 2016 [citado 2016 dez. 15];6(2). Disponível em: http://revistas.ua.pt/index.php/ilcj/article/view/4534

3. Creswell JW, Clark VLP. Pesquisa de métodos mistos. São Paulo: Penso; 2013.

4. Garcia MN, Santos SMB, Pereira RS, Rossi GB. Software livre em relação ao software proprietário: aspectos favoráveis e desfavoráveis percebidos por especialistas. Gest Reg [Internet]. 2010 [citado 2010 dez. 15];26(78):106-20. Disponível em: http://seer.uscs.edu.br/index.php/revista_gestao/article/viewFile/1061/847

5. Camargo BV, Justo AM. Tutorial para uso do software de análise textual IRAMUTEQ. Universidade Federal de Santa Catarina [Internet]. 2013 [citado 2015 out 15]. Disponível em: http://www.iramuteq.org/documentation/fichiers/tutoriel-en-portugais

6. Camargo BV, Justo AM. IRAMUTEQ: um software gratuito para análise de dados textuais. Temas Psicol [Internet]. 2013 [citado 2015 out. 15];21(2):513-8. Disponível em: http://pepsic.bvsalud.org/pdf/tp/v21n2/v21n2a16.pdf

7. Lowen IMV, Peres AM, Crozeta K, Bernardino E, Beck CLC. Managerial nursing competencies in the expansion of the Family Health Strategy. Rev Esc Enferm USP [Internet]. 2015 [cited 2016 Nov 20];49(6):964-70. Available from: http://www.scielo.br/pdf/reeusp/v49n6/0080-6234-reeusp-49-06-0967.pdf

8. Mendes FRP, Zangão MOB, Gemito MLGP, Serra ICC. Social representations of nursing students about hospital assistance and primary health care. Rev Bras Enferm [Internet]. 2016 [cited 2017 Apr 10];69(2):343-50. Available from: http://www.scielo.br/pdf/reben/v69n2/en_0034-7167-reben-69-02-0343.pdf

9. Kami MTM, Larocca LM, Chaves MMN, Lowen IMV, Souza VMP, Goto DYN. Working in the street clinic: use of IRAMUTEQ software on the support of qualitative research. Esc Anna Nery [Internet]. 2016 [cited 2016 Nov 20];20(3):e20160069. Available from: http://www.scielo.br/pdf/ean/v20n3/en_1414-8145-ean-20-03-20160069.pdf

10. Reinert M. Alceste une méthodologie d'analyse des données textuel les e tune application: Aurelia de Gerard de Nerval. Bull Methodol Sociol.1990;26(1):24-54.

11. Chartier JF, Meunier JG. Text mining methods for social representation analysis in large corpora. Papers Soc Represent [Internet]. 2011 [cited 2015 Oct 15];20. Available from: http://www.psych.lse.ac.uk/Psr/PSR2011/20_38.pdf

12. Lahlou S. Text mining methods: an answer to Chartier and Meunier. Papers Soc Represent [Internet]. 2012 [cited 2015 Dec 15];20. Available from: http://www.psych.lse.ac.uk/psr/PSR2011/20_39.pdf

13. Oltramari LC, Camargo BV. Aids, relações conjugais e confiança: um estudo sobre representações sociais. Psicol Estud (Maringá) [Internet]. 2010 [citado 2015 abr. 20];15(20):275-83. Disponível em: http://www.scielo.br/pdf/pe/v15n2/a06v15n2.pdf

14. Nascimento ARA, Menandro PRM. Análise lexical e análise de conteúdo: uma proposta de utilização conjugada. Estud Pesq Psicol UERJ [Internet]. 2006 [citado 2016 jul. 15]; 6(2):72-88. Disponível em: http://www.revispsi.uerj.br/v6n2/artigos/pdf/v6n2a07.pdf

15. Rehem TCMSB, Egry EY, Ciosak SI. Internações sensíveis à atenção primária: uso de ferramenta decodificadora para estudo das percepções dos profissionais da saúde. Indagatio Didactica [Internet]. 2013 [citado 2015 out. 10];5(2). Disponível em: http://revistas.ua.pt/index.php/ID/article/view/2448

16. Ratinaud P. IRAMUTEQ: Interface de R pour les analyses multidimensionnelles de textes et de questionnaires (computer software) [Internet]. 2009 [cited 2015 Oct 15]. Available from: http://www.iramuteq.org

17. Trigueiro DRSG, Almeida AS, Monroe AA, Costa GO, Bezerra VP, Nogueira JA. AIDS and jail: social representations of women in freedom deprivation situations. Rev Esc Enferm USP [Internet]. 2016 [cited 2016 Aug 15];50(4):554-61. Available from: http://www.scielo.br/pdf/reeusp/v50n4/0080-6234-reeusp-50-04-0554.pdf

18. Lowen IMV, Peres AM, Ros C, Poli Neto P, Faoro NT. Innovation in nursing health care practice: expansion of access in primary health care. Rev Bras Enferm. 2017;70(5):898-903. DOI: http://dx.doi.org/10/1590/0034-7167-2016-0131

19. Fé MMM, Ramos CV, Moura ME, Nunes BMVT, Martins MCC. Orientação alimentar para os diabéticos na Estratégia Saúde da FamíliaJ Nurs UFPE On Line [Internet]. 2016 [citado 2016 jan. 20];10 Suppl. 1:202-10. Disponível em: https://periodicos.ufpe.br/revistas/revistaenfermagem/article/viewFile/10941/12244

20. Berri B, Castro A, Camargo BV. Representações sociais relacionadas às práticas de rejuvenescimento. Psicol Pesq UFJF [Internet]. 2016 [citado 2016 jul. 10];10(2):21-30. Disponível em: http://pepsic.bvsalud.org/pdf/psipesq/v10n2/04.pdf