Use of the Software IRAMUTEQ in Qualitative Research: An Experience Report

Maria Marta Nolasco Chaves, Ana Paula Rodrigues dos Santos,
Neusa Pereira dos Santosa and Liliana Müller Larocca

Abstract This article describes the experience of the use of the software IRAMUTEQ to support qualitative research, and thus concepts and some characteristics of the software are addressed to aid in understanding and to illustrate its features. The observations come from two surveys carried out in the city of Curitiba, which were centered on the vulnerability of and the promotion of a healthy lifestyle to adolescent students. The tool enabled the organization of data collected from the participants' speeches acquired through semi-structured interviews. Both studies went through the same steps in the processing of the data from the interviews because, initially, they were transcribed for the preparation of the *corpus* of each of the studies so as to be inserted in the software. Categorization of the data was performed using Descending Hierarchical Classification (DHC). This tool allowed us to show and confirm the thematic categories in the same way that it added quality to the presentation and supported the results.

Keywords Software tools · IRAMUTEQ · Qualitative research · Nursing · Collective health

M.M.N. Chaves (✉) · A.P.R. dos Santos · N.P. dos Santosa · L.M. Larocca
Departamento de Enfermagem, Universidade Federal Do Paraná,
Curitiba, PR, Brazil
e-mail: mnolasco@terra.com.br

A.P.R. dos Santos
e-mail: ana25rodrigues@gmail.com

N.P. dos Santosa
e-mail: neusapesantos@gmail.com

L.M. Larocca
e-mail: larocca_m@terra.com.br

1 Introduction

This article reflects on the use of the software IRAMUTEQ in the organization of data and as support in the analysis of said data in two qualitative exploratory searches: "Vulnerabilidade na adolescência: a perspectiva de gestores e líderes do movimento social organizado em um território de Curitiba-PR" ("Vulnerability in adolescence: the perspective of managers and leaders of the social movement organized in a territory of Curitiba, PR") and "Promoção da Saúde do escolar adolescente segundo as diretrizes do programa de saúde do escolar: uma experiência em um município do Sul do Brasil." ("Health promotion to adolescent students according to the school health program's guidelines: an experience in a Southern Brazilian municipality"). The studies were grounded in the theoretical and methodological framework of the Theory of Praxis Intervention in Nursing in General Health and critical epidemiology, both of which are founded in historical and dialectical materialism and thus advise that the praxis in nursing and public health should occur via the dynamic systematization of capturing, interpreting, and intervening in health phenomena, relating this to the processes of production and reproduction of society in its historicity and dynamicity (Egry 1996; Breilh 2006).

To uncover these processes, which are not directly observable in reality, researchers have sought, through the guidance of critical epidemiology (Breilh 2006), to identify the processes of wear and protection present in the lives of the population segment defined for the studies—adolescents living in an area covered by a local health service. To further the discussion and meet the objectives proposed in each survey, data regarding the processes that are directly or indirectly encountered in the lives of those subjects was examined. These processes were related to current public policies that guided actions which are the responsibility of local services in the areas of health, education, sports and leisure, security, and social action. Information about the processes related to the social movement organized by its demands that seeks to establish the necessary confrontations in order to change the local reality was also gathered.

To handle this stage of the studies, in which the particular and structural dimension related to the determination of the local reality was examined, data available in secondary banks on public websites was used. To collect primary data, which was done through semi-structured interviews, it was sought to explore the singular dimension of the reality and objects defined for the referred studies, therefore there were 88 interviews, with 73 participants in the study of health promotion and 15 participants in the study about vulnerability.

Among the participants, managers and professionals of the local services were encountered. These were responsible for developing actions that were, in some way, directed at the adolescents of that region. These health education actions, which were developed in six workshops at the local health care facility, met one of the objectives of the study of health promotion. Adolescent students who partici-pated in these actions were also interviewed. In this phase of data collection, ethical

aspects of research with human beings were respected, according to the current resolution in Brazil (Brasil 2012).

Thus, faced with scores of information and the need to discuss in depth the data collected, a qualitative research support software that helps organize and systematize data was used, guided by discourse analysis (Bardin 2011). It is believed that this process enhances the construction of knowledge in studies with little subjective interference from the researchers involved in the presentation of the results, and consequently helps the development of data analysis, since the support of the thematic categorization of the data can be both proven and disproven through textual analysis, which is done by clustering the fragments according to the frequency of the related term or terms, seeing as the system itself performs the data crossing to the extent that the researcher directs the command provided by the software developed to this end.

With the use of the software in the aforementioned studies, the authors sought to overcome some of the challenges identified around the methods and techniques used in qualitative research. At present, the relevance of the development of such research—with emphasis on scientific rigor, transparency, and the inclusiveness of the stages of collection and processing of data so as to ensure the reliability of their analysis, the magnification of subjects/source, etc., and, consequently, the dissemination of knowledge built of the study of means of impact—is debated (Latimer 2005; Egry and Fonseca 2015; Tong et al. 2007).

Such overcoming is in the historical difficulty of acceptance of qualitative studies in contrast to quantitative studies, that, for the most part, according to the chosen methods, can be replicated and permit comparison of results (Sánchez 2015); however, even with greater acceptance, the knowledge constructed from this thinking has its limits and does not allow for certain seizures of objects and realities chosen for the study, especially those that are of interest to the social sciences. Therefore, regardless of the nature of the study, quantitative or qualitative, it is necessary to reflect on these limits and outline studies that help to overcome or complement the acquired knowledge.

2 The Process Organizing of Empirical Data Using the Software IRAMUTEQ in the Discussion and Analysis of Data

Qualitative research explores complex phenomena, which are obtained from non-quantitative data, usually from interviews and focus groups. The interviews—chosen method for the collection of primary data in the aforementioned studies—allowed for the analysis of the experiences and meanings attributed to a particular phenomenon, namely vulnerability of adolescents and health promotion to adolescent students (Tong et al. 2007). And, in order to analyze all of the verbal text material produced, the use of specific software has been increasingly present,

especially in studies in which the *corpus* to be analyzed is large (Camargo and Justo 2013a).

In this sense, in order to analyze the amount of *corpus* originating from the interviews, the software IRAMUTEQ (*Interface de R pour les Analyses Multimensionnelles de textes et de Questionnaires*) was used. IRAMUTEQ is a free software anchored in the R software, which enables different processing and statistical analysis of texts—in this case produced from interviews—documents, and other modalities, as it organizes the distribution of vocabulary in an understandable and visually clear manner, thus facilitating the process of organizing the collected material (Camargo and Justo 2013a).

IRAMUTEQ was developed by the French researcher Pierre Ratinaud in 2009, and, despite originally being in French, this software already has full dictionaries in English and Italian, and more recently in Portuguese, since it has been used in Brazil since 2013, and has since become innovated qualitative health research (Lowen et al. 2015; Camargo and Justo 2013b).

This software makes different types of analysis of textual data possible, from the simplest—such as basic lexicography, which covers, primarily, the lemmatization and the calculation of word frequency—to multivariable analysis, such as Descending Hierarchical Classification (DHC), post-factorial correspondence analysis, similitude analysis, which enables a lexical analysis of textual material, providing contexts (lexical classes) characterized by a specific vocabulary e by segments of text that share this vocabulary (Camargo and Justo 2013a). One of the possibilities of this lexical analysis is verified in the word cloud that can be made through the researcher's commands, as shown in Fig. 1. In this example, the relationship between lexical terms and the term adolescent in the study that analyzed vulnerabilities is shown.

The software also "classifies segments of text according to their respective vocabularies and their conjunction is distributed based on the frequency of reduced forms, from matrices, crossing segments of texts and words" (Camargo 2005). In this phase of organization of empirical material, the software aims to obtain Elementary Context Unit (ECU) classes from the Initial Context Units (ICU) as the program performs the scaling of segments of text or ECU, which have an average of three lines, classified according to the most frequent vocabulary and the highest chi-squared values in the class, considering the understanding that they were significant for the qualitative analysis of the data (Camargo and Justo 2013a; Ratinaud 2009).

In the development of the studies, the *corpus* was prepared, which is the set of Initial Context Units (ICU) to be analyzed, and in this sense, the development of the *corpus* resulted from the transcription of the interviews and, later, was grouped into a single text file using the software OpenOffice.org and saved as text.txt, separated by command lines according to the research variables and, from the *corpus* prepared for the study proceeded the textual analysis, being that, in both studies, the method used was the Descending Hierarchical Classification (DHC), as the *corpus* of each study worked with a textual set centered on a theme (Camargo 2005).

The ICUs or texts were build based on each guiding question of their respective studies and not on each interview with the participants, i.e., care was taken with the

Fig. 1 Word cloud from the study of vulnerability in adolescence. *Source* dos Santos (2015)

preparation of the *corpus* through questions asked to each participant to not use the complete interviews. This choice was made to facilitate the organization of the empirical material and, later, to facilitate the process of analysis of the results. Therefore, the participants' speeches were grouped according to the category to which they belonged (manager, employee, teacher, adolescent) and, thus, the issuer was evidenced by means of encoding previously established by the researcher (Camargo and Justo 2013a; Camargo 2005).

From the analysis performed by the software, which is done by commands from the researcher, IRAMUTEQ organized the classes, which are composed of several segments of texts according to a classification based on the distribution of vocabulary, originating from a dendrogram of the DHC from the *corpus* that illustrated the relationship between these. IRAMUTEQ allowed for the description of each of these classes, chiefly through their characteristic vocabulary (lexicon) and through the words with asterisks (variables).

The command lines and the variables containing asterisks were: **** *n_01, in an increasing ordinal sequence that would meet the category and the number of

respondents; *gest_1, manager; *educ_1, educator; *saud_1, health; *soc_1, social action; *adolesc_1, adolescent student; *func_1, school employee from the area studied; *prof_1, high-school teacher from the area studied. It should be noted that this procedure, identifying the dendrogram of the DHC, lasted 12 s for the study that worked with vulnerabilities in adolescents and 25 s for the study of health promotion, which was a significant advantage at this stage of the studies when comparing the procedure to other forms of qualitative data processing.

After processing the *corpus*, from which the dendrogram of the DHC originated, the software showed the results in another way, or in other words, presented the results using a factorial correspondence analysis made from the DHC, seeing that the program, through calculation, provided the most characteristic segments of text from each class (*corpus* in color, also known as *corpus* cooler), which allowed the researchers to contextualize the typical vocabulary in each class. Based on this contextualization, the classes originated in each study were interpreted by means of content analysis (Bradin 2011), which is defined by a set of communication analysis techniques, which use systematic and objective procedures to describe the content of the messages, allowing the inference of knowledge related to the condition of receiving such messages (Bradin 2011).

In the study of vulnerability in adolescence, 74 texts were analyzed and processed by the software, from which 510 text segments were obtained; of these, 455, or 89.22 %, were used. After sizing the text segments, classified according to their vocabularies, the text classes were defined, as shown in Fig. 2, in which the dendrogram of the Descending Hierarchical Classification (DHC), is illustrated.
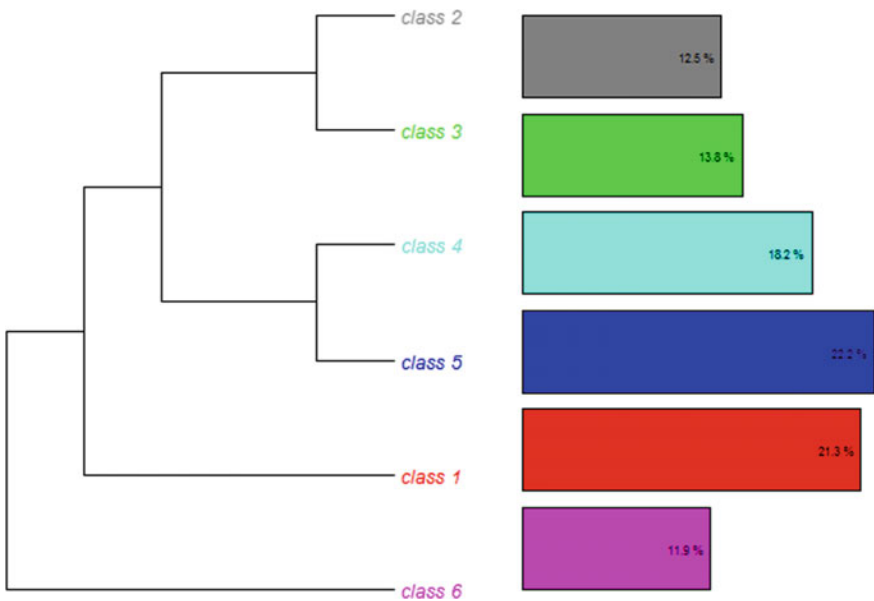

Fig. 2 Dendrogram of the classes provided by IRAMUTEQ. *Source* dos Santos (2015)

According to Fig. 2, the *corpus* was divided into four *sub-corpora*. Class 6, which consists of 54 Elementary Context Unit (ECU) and which concentrates 11.9 % of the total ECUs of the *corpus*, was obtained from one of the *sub-corpora*. Class 1 was obtained from another *sub-corpus*, with 97 ECU, which corresponds to 21.3 % of the ECUs, and, from this, three more divisions from which classes 2, 3, 4, and 5 originated, with 57, 63, 101, and 83 ECUs and corresponding to 12.53, 13.8, 22.2, and 18.24 %, respectively, of ECUs from the whole *corpus*. For each class, a list of words generated from chi-square tests (X2) was computed, or rather, this analysis aims to obtain the ECU classes, which present similar vocabulary to each other and, at the same time, different vocabulary from the ECU of the other classes, that is, as explained above, the program performs calculations and provides results that allow for the description of each of the classes, mainly through their characteristic vocabulary (lexicon). The percentage, on the other hand, refers to the occurrence of the word in the text segments in that class in relation to its occurrence in the *corpus*, while the chi-square refers to the association of the word with the class (Camargo and Justo 2013b) (Table. 1).

After reading and analyzing the *corpus*, it was noted what classes 2, 3, 4, and 5 had the same logical sequence of subjects demonstrated by the software, that is, after processing the *corpus* and reading thoroughly the material, by means of the *corpus* cooler, analysis of the results was done through content analysis (Bradin 2011). From this analysis, classes 2 and three were joined together, as well as

Table 1 Distribution of the terms by chi-square (X2) and the frequency of the term in the classes

Class	Word	X2	%
Class 1	You	63.34	65.31
	Phase	41.6	100.0
	Prepared	37.74	90.91
Class 2	Community Health Center	86.91	80.0
	Nurse	63.2	90.91
	Social Assistance Regional Council (CRAS)	37.79	62.5
Class 3	School Community	44.24	100.0
	Accomplish	41.76	44.68
	Sports Field	37.03	87.5
Class 4	Family	41.33	73.08
	Also	33.76	52.73
	People	32.11	43.0
Class 5	Neighborhood	57.6	62.5
	Grow	31.86	100.0
	Demand	30.55	81.82
Class 6	Violence	138.22	95.0
	Drug	115.9	82.61
	Pregnancy	99.38	100.0

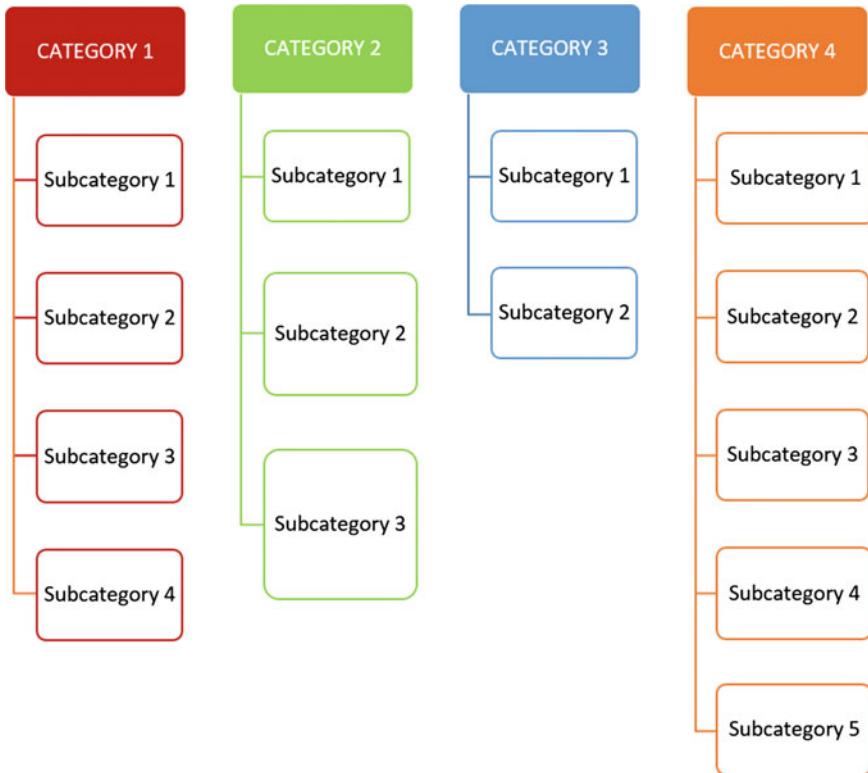*Source* dos Santos (2015)

Fig. 3 Diagram of thematic categories and subcategories demonstrated in the participants' speeches in a study. *Source* dos Santos (2015)

classes 3 and 4, thus forming four major thematic categories, which were based on the theoretical framework that guided the study, as shown in Fig. 3.

Content analysis (Bradin 2011) consists of phases that will help the researcher organize the material, and here, it is worth mentioning that the organization of said material was done using the formatting of the *corpus* to be analyzed by the software IRAMUTEQ, and, with the disclosure of the classes, we proceeded to the stage of superficial reading and defining of analysis categories, which was done through analytical categories defined for each of the studies. The objective of this phase was to go beyond common sense and subjectivity in the interpretation of the data collected and, thus, interpret critically the participants' speeches through the identification of core meanings in each thematic category shown. The core meanings were identified in each class originated in the dendrogram of the DHC, as well as the relationships between the classes and the frequency of texts in each of these.

Through the experience of the studies, the core meanings evident were: violence, drugs, pregnancy, STIs, vulnerability, gender and generation, territory, and intersectionality, whose presence and frequency had greater chi-square values and

significance $p > 0.001$, indicating a significant association for the proposed objectives. The chi-square test is used to verify the association of ECU with a particular class; therefore, the higher the value, the greater the association, whereas all of the selected words had a $p > 0.001$, indicating a significant association (Chartier and Meunier 2011).

It is believed that statistical calculations of qualitative variables, from the entered texts, do not make the software a proper research method, since this does not analyze the organized data; however, in the process of organizing data, this tool allows for more thorough exploration of data. Therefore, the use of software in qualitative studies can facilitate the processing of long and numerous texts, but it does not replace the essential role of the researcher in the collection, preparation, and analysis of data. The outline of the study, the methodological approach, the interpretation, and the analysis will always be the responsibility of the author of the study, who should respect the ethical aspects and quality parameters of qualitative research (Egry and Fonseca 2015; Camargo and Justo 2013a; Chartier and Meunier 2011; Lahlou 2012; Coutinho 2015).

3 Conclusion

The use of the software IRAMUTEQ was decisive for the processing of the research data, as it allowed for the development of a more critical look at the material—through the frequency of the terms, and the possibility to confirm the participants' speeches using the core meanings in each thematic category outlined by the DHC dendrogram. This process allowed for the qualification, the consistency, and the visualization of empirical texts in the categorization process of the speeches and, consequently, the qualification of the results and the analysis done in both studies.

Although the software's interface is relatively easy to use, it is important to note that the free access of this tool has contributed significantly to its use in research. Another relevant issue to be emphasize is the role of the researcher as of fundamental importance in the design of the study, organization of collected material, and the analysis process, as the software does not do the research itself, but aids in the organization, processing, and in the support of the findings to carry out data analysis, mainly in studies with large volumes of text and, for this to happen, it is necessary that the researcher has knowledge for the use of the selected tool.

References

Bardin, L. (2011). Análise de conteúdo. São Paulo: Editions 70.

Brasil. (2012). Resolução 466/2012. Regulatory guidelines and standards for research involving humans. Ministério da Saúde/Conselho Nacional de Saúde (Ministry of Health/National Health Council), Brasília, December 12, 2012.

Breilh, J. (2006). *Epidemiologia crítica: Ciência emancipadora e interculturalidade*. Rio de Janeiro: Hucitec.

Camargo, B. V. (2005). ALCESTE:um programa informático de análise quantitativa de dados textuais. In A. S. P. Moreira, B. V. Camargo, J. C. Jesuíno, & S. M. Nóbrega (Eds.), *Perspectivas teórico-metodológicas em representações sociais* (pp. 511–539). João Pessoa, PB: Editora da Universidade Federal da Paraíba.

Camargo, B. V., & Justo, A. M. (2013a). IRAMUTEQ: um *software e* gratuito para análise de dados textuais. Rev. Temas em psicologia, Ribeirão Preto, Vol. 21, no. 2, pp. 513–518. Available at: http://pepsic.bvsalud.org/pdf/tp/v21n2/v21n2a16.pdf. Accessed on: 10 May 2015.

Camargo, B. V., & Justo, A. M. (2013b). Tutorial para uso do software de análise textual IRAMUTEQ. Universidade Federal de Santa Catarina. Available at: http://www.iramuteq.org/documentation/fichiers/tutoriel-en-portugais. Accessed on 10 May 2015.

Chartier, J. F., & Meunier, J. G. (2011) *Text mining methods for social representation analyses in large corpora*. Papers on Social Representations, Paris (FR), Vol. 20, no. 37, pp. 1–47. Available at: http://www.psych.Ise.ac.uk/psr/

Coutinho, C. (2015). Avaliação da qualidade da investigação qualitativa: algumas considerações teóricas e recomendações práticas. In de Souza, F. N., de Souza, D. N., & Costa, A. P. (Eds.), Investigação qualitativa: inovação. Dilemas e desafios. Aracaju: EDUNIT.

Egry, E. Y. (1996). *Saúde coletiva: construindo um novo método em enfermagem*. São Paulo: Editora Ícone.

Egry, E. Y., & Fonseca, R. M. G. S. (2015). Acerca da qualidade nas investigações qualitativas em enfermagem. In de Souza, F. N., de Souza, D. N., & Costa, A. P. (Eds.), Investigação qualitativa: inovação. Dilemas e desafios. Aracaju: EDUNIT.

Lahlou, S. (2012). *Text mining methods: An answer to Chartier and Munier*. Papers in Social Representations, London (UK), Vol. 20, pp. 38.1–39.7. Available at: http://psych.Ise.ac.uk/psr/PSR2011_39.pdf

Latimer, J. (2005). *Investigação qualitativa avançada para a enfermagem*. Lisbon: Sersílito.

Lowen, I. M. V., Peres, A. M., Crozeta, K., Bernardino, E., & Beck, C. L. C. (2015). Competências gerenciais dos enfermeiros na ampliação de estratégia saúde da família. RevEscEnferm USP, São Paulo, Vol. 49, no. 6, pp. 967–973. Available at: http://www.scielo.br/pdf/reeusp/v49n6/pt_0080-6234-reeusp-49-06-0967.pdf

Ratinaud, P. (2009). IRAMUTEQ: Interface de R pourLesanlysesmultidimensionnlles de textes et de questionnaires [computer software]. Available at: http://www.iramuteq.org

Sánchez, C. (2015). Orígenes y evolución de la investigación cualitativa en educación. In de Souza, F. N., de Souza, D. N., & Costa, A. P. (Eds.), Investigação qualitativa: inovação. Dilemas e desafios. Aracaju: EDUNIT.

dos Santos, A. P. R. (2015). *Vulnerabilidade na adolescência: a perspectiva de Gestores e Líderes do movimento social organizado em um território de Curitiba-PR Dissertação (mestrado) – Programa de Pós-Graduação em Enfermagem*. Setor de Ciências da Saúde: Universidade Federal do Paraná.

Tong et al. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups, Vol. 19, no. 6, pp. 349–357. Available at: https://intqhc.oxfordjournals.org/content/intqhc/19/6/349.full.pdf