Cenas Educacionais

Thematic Dossier
RESEARCH METHODS AND EXPERIENCES

e-ISSN: 2595-4881

THE USE OF THE IRAMUTEQ SOFTWARE IN DATA ANALYSIS IN QUALITATIVE OR QUALI-QUANTI RESEARCH

EL USO DEL SOFTWARE IRAMUTEQ EN EL ANÁLISIS DE DATOS EN INVESTIGACIÓN CUALITATIVA O CUALICUANTI

O USO DO SOFTWARE IRAMUTEQ NA ANÁLISE DE DADOS EM PESQUISA QUALITATIVA OU QUALI-QUANTI

Abília Ana de Castro Neta [1]
Berta Leni Costa Cardoso [2]

Manuscript received on: May 1, 2021.
Approved on: May 29, 2021.
Published: June 23, 2021.

Abstract
This study aims to describe the use of the IRAMUTEQ software as a tool to support data processing in qualitative research (and/or quali-quanti). Although IRAMUTEQ guarantees accurate results, the use of the software does not replace the researcher's protagonism. It is highlighted, in this perspective, that all stages of the research require seriousness and knowledge of the object under investigation, which permeates the choice of approach, the coherent and consistent discussion with the literature and the quality of the instrument to conduct the interview in data collection, so that the preparation of the corpus allows a better processing and presentation of the results that are generated by the software. In this sense, the researcher continues to be the conductor of the research, and his role is valued by IRAMUTEQ, which allows the interpretation of results already processed with scientific rigor.

Keywords: IRAMUTEQ; Qualitative research; Quantitative Research.

Resumen
Este estudio tiene como objetivo describir el uso del software IRAMUTEQ como herramienta de apoyo al procesamiento de datos en la investigación cualitativa (y/o cuali-cuanti). Si bien IRAMUTEQ garantiza resultados precisos, el uso del software no reemplaza el protagonismo del investigador. En esta perspectiva, todas las etapas de la investigación requieren seriedad y conocimiento del objeto investigado, lo que impregna la elección del enfoque, la discusión coherente y consistente con la

[1] Doctoral student in Education at the Federal University of Bahia. Master in Education from the State University of Southwest Bahia. Member of the Study Group and Didactic Research, Training and Teaching Work and of the Research Group on Physical Education, Sport and Leisure.
ORCID: http://orcid.org/0000-0002-7222-5116
Contact: bia_gbi@hotmail.com
[2] PhD in Physical Education from the Catholic University of Brasília. Post-doctorate in Education from the State University of Southwest Bahia. Professor at the Graduate Program in Education at the State University of Southwest Bahia and at the State University of Bahia. Member of the Study Group and Didactic Research, Training and Teaching Work and of the Study Line Research and Extension in Physical Activity.
ORCID: http://orcid.org/0000-0001-7697-0423
Contact: bcardoso@uneb.br

e-ISSN: 2595-4881

literatura y la calidad del instrumento para realizar la entrevista en la recolección de datos, de manera que la elaboración del corpus permite un mejor procesamiento y presentación de los resultados generados por el software, en este sentido el investigador sigue siendo el conductor de la investigación, y su rol es valorado por IRAMUTEQ, lo que permite la interpretación de los resultados ya procesados con rigor científico.

Palabras clave: IRAMUTEQ; Investigación cualitativa; Investigación cuantitativa.

Resumo
Este estudo tem como objetivo descrever o uso do software IRAMUTEQ como ferramenta de apoio ao processamento de dados na pesquisa qualitativa (e/ou quali-quanti). Embora o IRAMUTEQ garanta resultados precisos, o uso do software não substitui o protagonismo do pesquisador. Ressalta-se, nesta perspectiva, que todas as etapas da pesquisa requerem seriedade e conhecimento do objeto em investigação, o que perpassa pela escolha da abordagem, pela discussão coerente e consistente com a literatura e pela qualidade do instrumento para efetuar a entrevista na coleta dos dados, para que a confecção do corpus possibilite um melhor processamento e apresentação dos resultados gerados pelo software. Neste sentido, o pesquisador continua sendo o condutor da pesquisa, e seu papel é valorizado pelo IRAMUTEQ, o qual possibilita a interpretação dos resultados já processados com rigor científico.

Palavras-chave: IRAMUTEQ; Pesquisa Qualitativa; Pesquisa Quantitativa.

Introduction

Since the 1980s, the creation of computer programs that offer assistance in data analysis in qualitative research (and/or quali-quanti) has increased in quantity and possibilities for researchers. These programs are called CAQDAS (*Computer Aided Qualitative Data Analysis Software*) and the debate regarding their use has grown, with consensus on their efficiency in managing and retrieving qualitative data. In this direction is the IRAMUTEQ (SOUZA ET AL, 2018).

IRAMUTEQ is a computer program that is anchored in the R software and allows different forms of statistical analysis on textual corpus and on tables of individuals by words (CAMARGO; JUSTO, 2013). In Brazil, it began to be used in 2013 in researches on social representations, however, other areas also appropriated its use and contribute to the dissemination of the possibilities of processing qualitative data (SOUZA *et al.*, 2018).

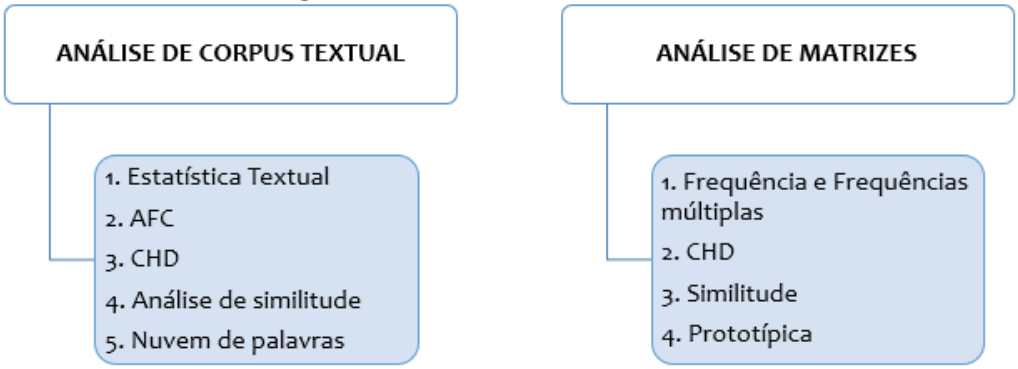
e-ISSN: 2595-4881

The IRAMUTEQ[3] (*R Interface pour les Analyzes Multidimensionnelles de Textes et de Questionnaires*) was developed by Pierre Ratinaud (2009). The software allows you to automatically carry out the lexical analysis of content. For this, the program organizes the characteristic words of the speech by classes, making it possible for the researcher to work with a large corpus of data (CAMARGO; JUSTO, 2013).

The main reason for choosing this software is due to the fact that it offers a wide range of tools for the analysis of qualitative data based on textual statistics, or lexicometry. In addition to its functional potential, IRAMUTEQ is a free computer program, but not widespread among Brazilian researchers (CAMARGO; JUSTO, 2013).

The use of IRAMUTEQ in the analysis of interview data can occur in different ways. This is because the software brings together a varied set of lexicometric procedures, such as: classic textual statistics, Specificity Analysis, Descending Hierarchical Classification (CHD), Correspondence Factor Analysis (CFA), similarity analysis, prototypical evocation analysis and word cloud (CAMARGO; JUSTO, 2013). In the software, therefore, it is possible to perform text analysis (textual corpus) and data table (textual matrix), as established in Figure 01.

Figure 01: Analysis with IRAMUTEQ



Source: Camargo and Justo (2013)

[3] In Brazil, the Laboratory of Social Psychology of Communication and Cognition at the Federal University of Santa Catarina (UFSC) is a research center that has been using the software in several researches, as well as promoting its dissemination in related research areas, including participating directly the creation and expansion of the Portuguese language vocabulary used by that program (AQUINO, 2015, p. 28).

e-ISSN: 2595-4881

Regarding the **Análise de Corpus Textual (**Analysis of Textual Corpus[4]**)** (which we will focus on in this study), Camargo and Justo (2013) recommend that **Estatística Textual** (Textual Statistics**)** identifies and reformats text units, identifies the number of words, average frequency and hapax (words with frequency one), searches the vocabulary and reduces words based on their roots (reduced forms), creates from the reduced forms dictionary, identifies active and supplementary forms. The **Especificidades e Análise Fatorial Confirmatória-AFC** (Specificities and Confirmatory Factor Analysis) associates texts with variables, that is, enables the analysis of textual production based on descriptive variables. It is possible to model the contrast analysis of the modalities of the variables and also the presentation in factorial plan.

With regard to **Classificação Hierárquica Descendente-CHD** (Descending Hierarchical Classification), text segments are classified according to their respective vocabularies and all of them are distributed according to the frequency of reduced forms. From matrices crossing text and word segments, the CHD method is applied and a stable and definitive classification is obtained (REINERT, 1990). This analysis aims to obtain classes of text segments that, at the same time, have similar vocabulary, and vocabulary different from the text segments of other classes (CAMARGO, 2005). From these matrix analyses, the software organizes the data analysis in a CHD dendrogram, which illustrates the relationships between classes. The program performs calculations and provides results that allow us to describe each of the classes, mainly, by its characteristic vocabulary (lexicon) and by its asterisked words (variables). In addition, the program provides another way of presenting the results, through a factorial correspondence analysis made from the CHD (CAMARGO; JUSTO, 2013).

The **Análise de Similitude** (Similitude Analysis) is based on the theory of graphs (MARCHAND; RATINAUD, 2012). It makes it possible to identify the co-occurrences between words and its result provides indications of the connection between the words. Finally, the Nuvem de Palavras (Word Cloud) groups the words and graphically organizes them according to their frequency. The figure generated does not follow descriptive indicators, but it can

[4] Since we are dealing with research results that were given and organized in pictures by a software, we will show these words in portuguese, so the readers can locate these words in the images provided. However, the respective translations will be shown in parenthesis.

e-ISSN: 2595-4881

facilitate a quick identification of the terms that are predominant in the set of texts analyzed (CAMARGO; JUSTO, 2013).

It should be emphasized that the choice to use one or another analysis technique depends on the characteristics of the problem and the research objectives. In this sense, the theoretical-methodological framework of the researcher, added to the support of lexicometric analysis software, can provide greater reliability to inferences made in qualitative and/or quali-quanti research.

Method

This is the description of the use of the IRAMUTEQ software as a tool to support data processing in qualitative research (and/or quali-quanti). The research was carried out in a municipality in the interior of Bahia. The sample consisted of teachers who work for the state autarchy in that municipality and who are willing to collaborate with the research. These subjects make up the faculty of the institution who are in class regency, notably 25 participants.

Data collection was carried out through interviews, in which a semi-structured script was applied. The conditions for the production of these interviews were duly agreed with the subjects, from the choice of dates to times and places. The information was collected free of charge for the interviewees and the identification data kept confidential, as a way of preserving the moral integrity of the subjects, assigning for this, fictitious names, when necessary to register these subjects, whose interests are turned, exceptionally, to the carrying out the research. The script for structuring the interview was prepared based on the discussions established by Piovezan (2017) about the phenomenon of **precariousness of teaching work** – an object investigated in the aforementioned master's research. For the textual analysis of the research, the IRAMUTEQ software was used.

Participation occurred voluntarily and was confirmed by signing the Informed Consent Form (FICF), which contains information about the study and its purposes. The empirical research started with the favorable opinion of the Research Ethics Committee (CEP). This investigation was approved by the CEP of the State University of Bahia with CAAE 43789015.4.0000.0057 and opinion number: 2,555,025.

e-ISSN: 2595-4881

Results and Discussion

Having the axes of the **precariousness of the teaching work** and the dialectical historical materialism as the guiding thread for the interviews, they were transcribed, duly adapted to the IRAMUTEQ norms and later processed by the software. The classes were named by the researcher based on the study's problematic. The possibilities for analysis will be presented below.

- Textual statistics

In this perspective, when processing the interviews in the software, it was observed that the general corpus consisted of 09 texts, separated into 162 text segments (ST). There were 5,758 occurrences (words, forms or words), with 1,572 distinct words, and 941 with a single occurrence. The analyzed content was categorized into three classes: Class 1, with 37/134 ST (27.61%); Class 2, with 49/134 ST (36.57%); and Class 3, with 48/134 ST (35.82%), which we will describe below.

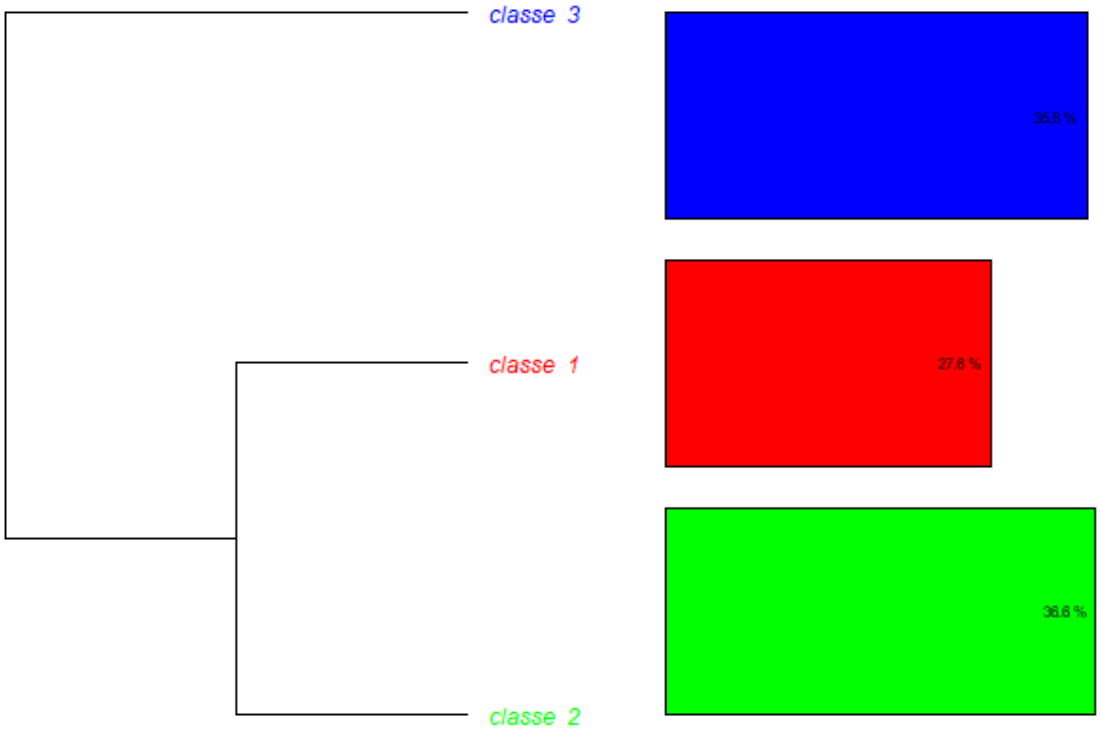
- Descending Hierarchical Classification (CHD)

In this perspective, the three classes that emerged from the CHD are divided into two branches (A and B) of the total corpus under analysis. **Subcorpus A**, "Desregulação e flexibilização da legislação trabalhista (Deregulation and relaxation of labor legislation)", composed of Class 1 ("Enfraquecimento de direitos (Weakening of rights)") and Class 2 ("Flexibilidade (Flexibility)"), was expressed, above all, by the terms "dever (duty)", "salário(salary)", "direito (right)", "lei (law)", "função (function)", "professor (teacher)" and "avaliação (evaluation)".

Subcorpus B, called "Intensificação do Trabalho (Work Intensification)", contains the speeches corresponding to Class 3 ("Intensificação (Intensification)") which contemplates another aspect of precarious work, expressed mainly by the terms "classe (class)", "sala (classroom)", "estudante (student)", "número (number)", "turma (class number)", "escola (school)" and "carga (load)". Figures 02, 03 and 04 were generated by the software, from statistical calculations that take into account the Descending Hierarchical Classification (CHD) of the lexical items present in the text segments of the corpus.

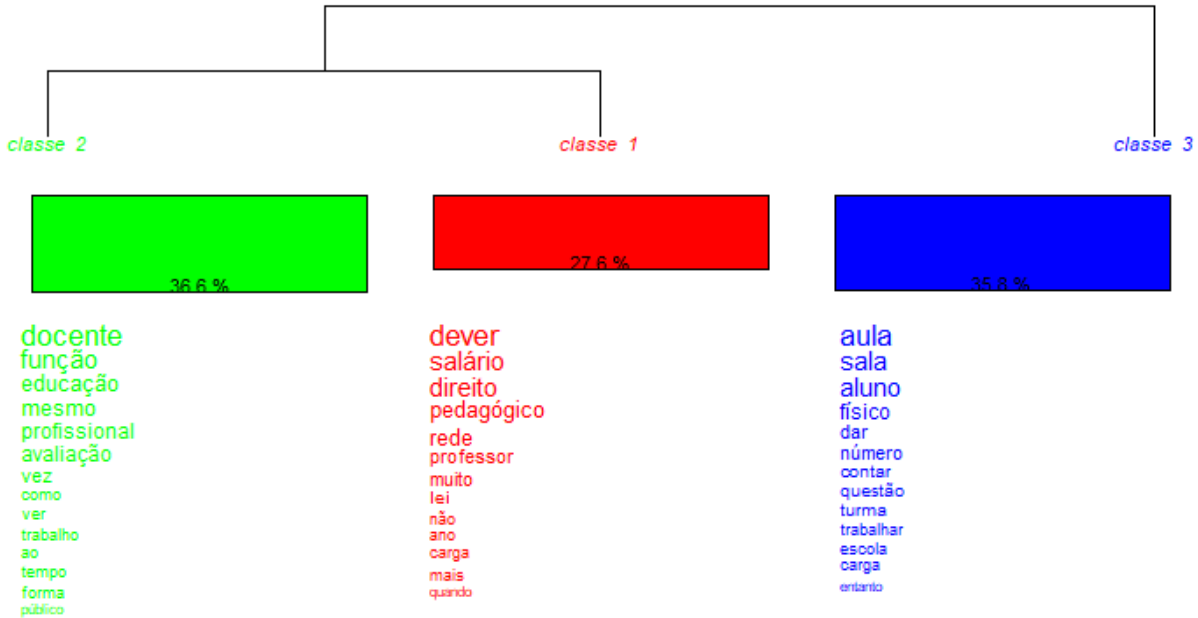
e-ISSN: 2595-4881

Figure 02: Class Classification Dendrogram (1)



Source: Research Data (IRAMUTEQ) (2020)

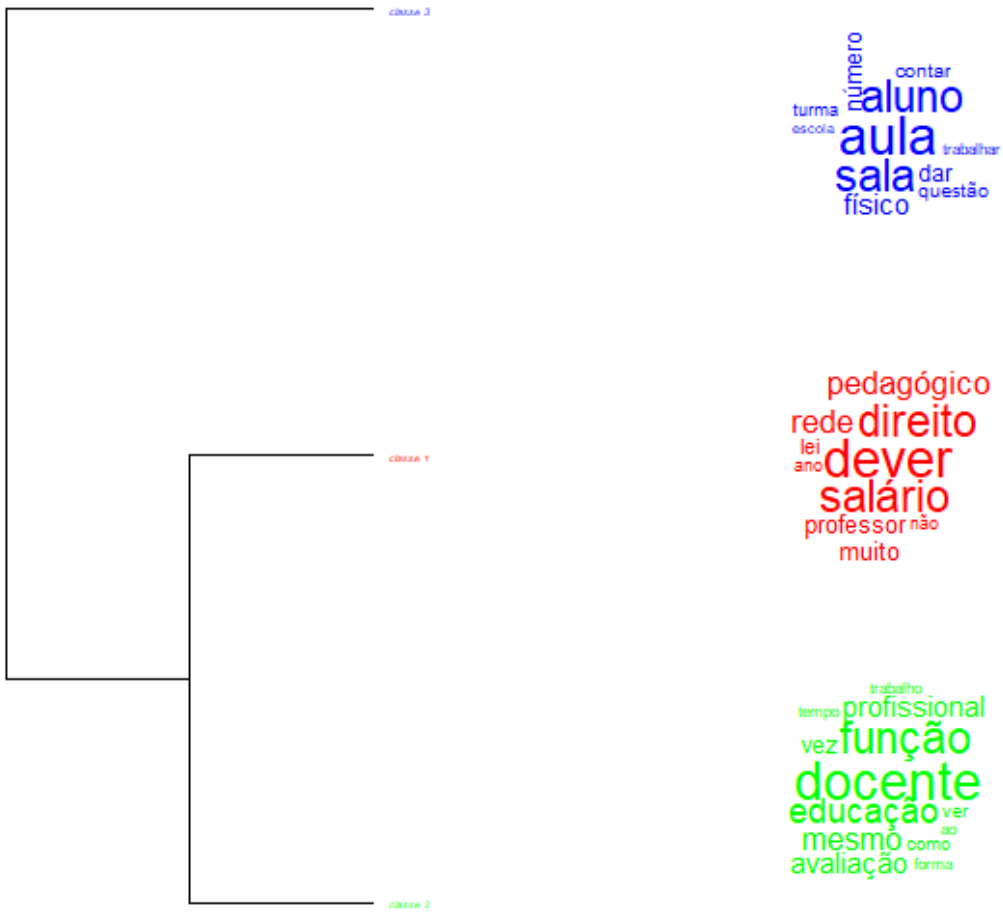
Figure 03: Class Classification Dendrogram (2)



Source: Research Data (IRAMUTEQ) (2020)

e-ISSN: 2595-4881

Figure 04: Class Classification Dendrogram (3)



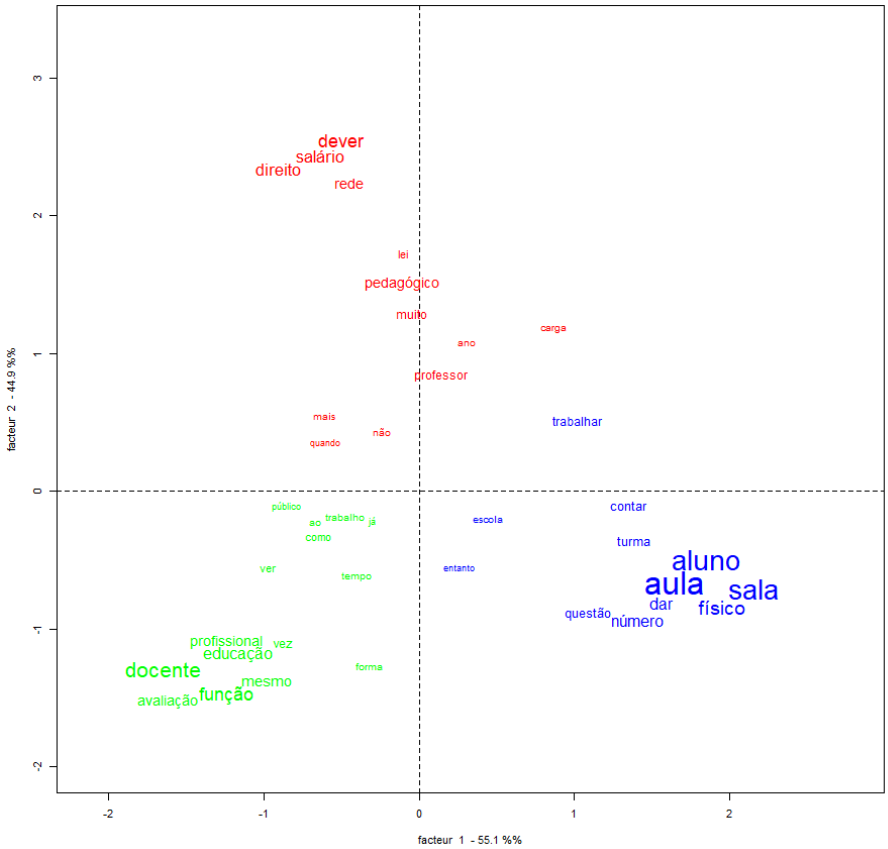
Source: Research data (IRAMUTEQ) (2020)

In **subcorpus B (Class 3)**, called "Intensificação de Trabalho (Work Intensification)", are expressed, overall, the terms "aula (class)", "sala (classroom)", "aluno (student)", "número (number)", "turma (class number)", "escola (school)" and "carga (load)".

- Factorial Correspondence Analysis (CFA)

The calculation process consists of crossing the occurrences of each lexical form of the corpus vocabulary with the classes resulting from the CHD, in order to present the relationships between these classes in a Cartesian factorial plan, divided into four quadrants of X and Y coordinates: upper right and left quadrants and lower left and right quadrants, as shown in Figure 05, below.

Figure 05: Visualization of classes distributed in the factorial plane

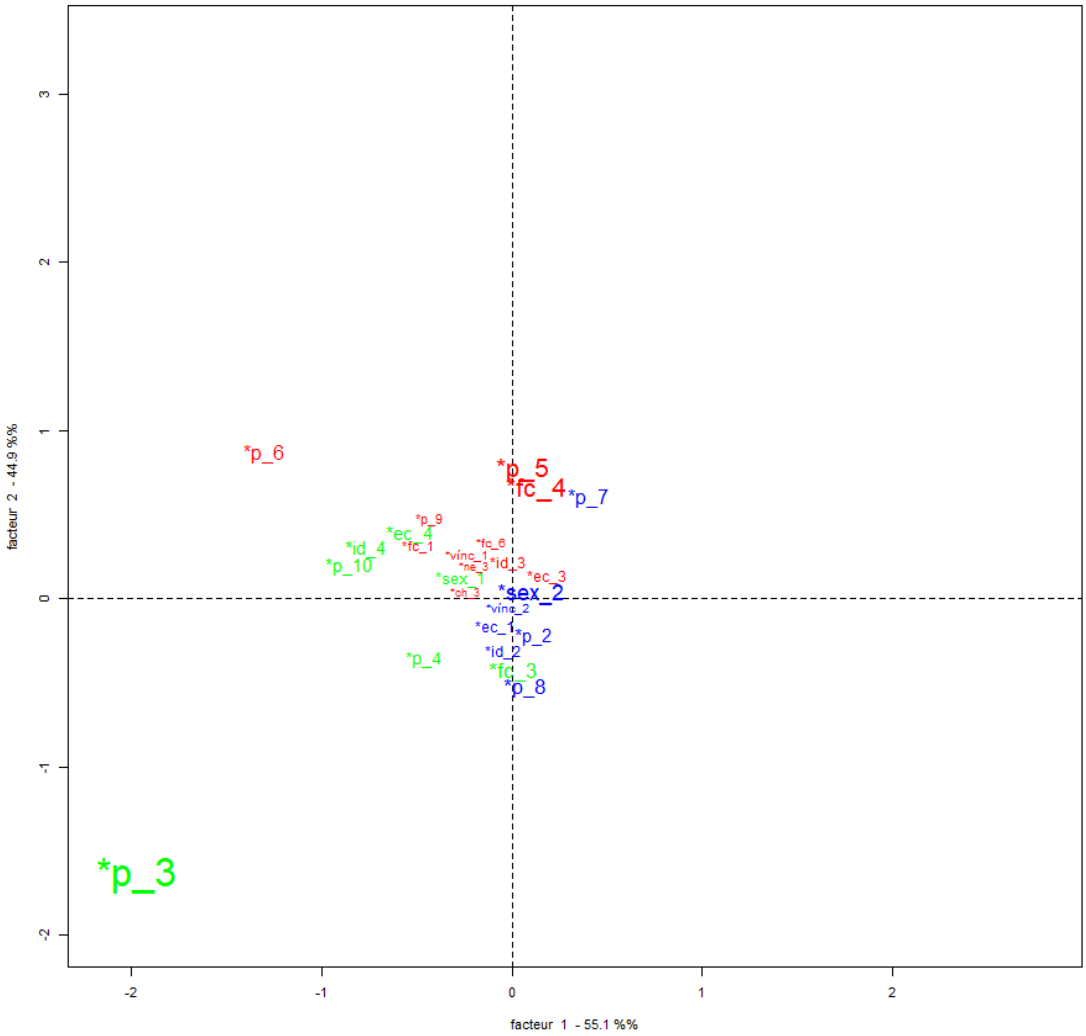


Source: Research data (IRAMUTEQ) (2020)

From the AFC it was possible to carry out text associations between words, considering the frequency of these and their respective classes, representing them in the factorial plane. In Figure 05, it can be seen that the three classes are in distinct quadrants, that is, each class covers specific semantic contexts, which refer to the semantic root of the word that most interfered in the class and allows us to perceive the action of the variables and the three observed classes. It is observed that the expressions "dever (duty)", "salário (salary)" and "direito (right)" make up Class 1, called "Fragilização de direitos (Weakening of rights)", and the terms "função (function)", "professor (teacher)" and "avaliação (evaluation)" integrate Class 2, called "Flexibilization (Flexibility)". The expressions "aula (class)", "sala (classroom)" and "estudante (student)" include Class 3, called "Intensificação (Intensification)", which were distributed by the respective quadrants in the factorial plan. Figure 06 shows the factorial plan from the sociodemographic variables listed in the organization of the textual corpus.

e-ISSN: 2595-4881

Figure 06: Visualization of sociodemographic variables distributed in the factorial plan



Source: Research data (IRAMUTEQ) (2020)

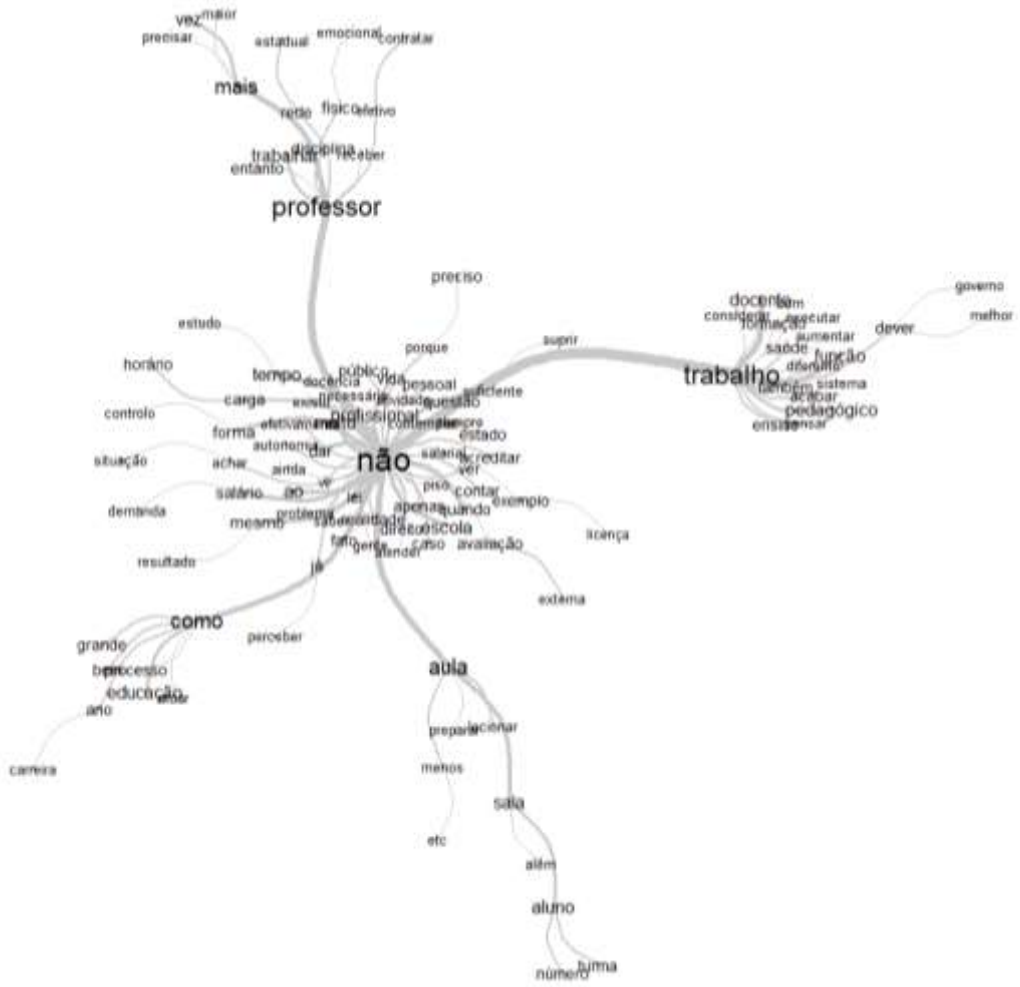
The factorial plan above shows the predominance of sociodemographic variables in Classes 1, 2 and 3. It is observed that teachers 02, 07 and 08 were categorized in Class 3 ("Intensificação (Intensification)"). According to the software, they are predominantly male and have an effective employment relationship. In Class 1 ("Fragilização de direitos (Weakening of rights)"), the predominance of teachers 05, 06 and 09 is evident, effective employment relationship, marital status: married, workload: 40 hours or more, career stage: examination. Teachers 03, 04 and 10 were categorized in Class 2 ("Flexibilização (Flexibility)"), predominantly female and career stage: variation.

e-ISSN: 2595-4881

- Similarity analysis

In figure 07, where the theory of graphs was applied, there is an emphasis on the words "não (no)", "aula (class)", "trabalho (work)" and "professor (teacher)", which creates a branch of other terms that have a certain degree of connectivity.

Figure 07: Similarity Analysis



Source: Research data (IRAMUTEQ) (2020)

Based on the image generated by the software, we organized figures 08, 09, 10 and 11, in which the expressions that are more congruent with the terms "não (no)", "aula (class)", "trabalho (work)" and "professor (teacher)".

e-ISSN: 2595-4881

Figure 08: Terms related to the expression "no"

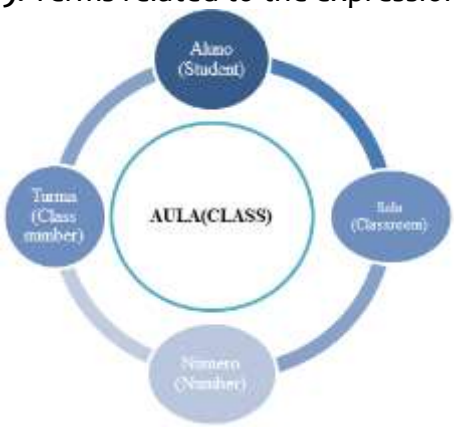
Source: Authors (2020)

Figure 08, by bringing the expression "no", as the central axis, and its ramifications "autonomy", "law", "salary" and "floor ", denotes the denial of these aspects in the work of teachers. Negative expressed in the transcripts below:

> ***DO NOT*** *consider my salary sufficient to meet my personal and family demands, so I have another "job"* ***(Prof. 02).*** (SALARY)
>
> *Unfortunately, most municipalities and states (with allegations that they do not have resources) have* ***NOT*** *fulfilled any of these rights: they were the result of a lot of union struggle. We've been struggling for 30 years* ***(Prof. 02).*** (FLOOR)
>
> *Unfortunately, in many municipalities it is* ***NOT*** *fulfilled* ***(Prof. 04).*** (FLOOR)
>
> *This law determines the salary floor national professional for public teaching professionals in basic education. But, it is* ***NOT*** *fulfilled everywhere* ***(Prof. 05).*** (FLOOR)
>
> *I* ***DON'T*** *believe that teaching is free. [...] External indexes and evaluations are a form of market domination, they take away from the teacher the freedom to act according to the reality where he works and standardize all individuals* ***(Prof. 08).*** (AUTONOMY)
>
> *[...] labor rights are* ***NOT*** *effectively respected, as there is always something that the system cheats* ***(Prof. 02).*** (RIGHTS)

e-ISSN: 2595-4881

Figure 09: Terms related to the expression "class"



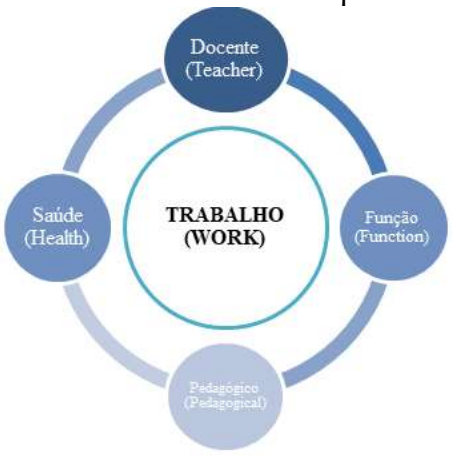
Source: Authors (2020)

Figure 09, when bringing the expression "class", as the central axis, and its ramifications "student", "room", "number" and "class" denotes the phenomenon of intensification/overload of teachers' work. Phenomenon expressed in the transcripts below:

> *If the salary is not enough, the alternative is to increase the workload and look for more **CLASSES** and other schools to teach **(Prof. 06).***
>
> *The low salary has led the teacher to work with more **CLASSES**, more **CLASSES**, in all shifts and in more than one school **(Prof. 07).***
>
> *[...] the excess of **CLASSES** and in different schools effectively contributes to the process of exhaustion/dissatisfaction/illness of the teacher. Also factors such as a large number of **CLASSES** and different disciplines to administer, rooms with large **NUMBER** OF **STUDENTS** [...] **(Prof. 02).***

See Figures 10 and 11:

Figure 10: Terms related to the expression "work"



Source: Own authorship (2020)

e-ISSN: 2595-4881

Figura 11: Terms related to the expression "professor"

Source: Own authorship (2020)

In Figure 10, the terms that have the greatest connection with the "work" axis were "teacher", "function", "pedagogical" and "health". In figure 11, the branches "work", "more", "physical" and "emotional" are connected with the central term "teacher". This denotes that the multiplicity of functions performed by teachers in their pedagogical practice and, above all, how these are established/organized in the capitalist mode of production, greatly affects the health/disease markers of these professionals. As stated by the interviewees:

> *All of this, together with the growing pace of social demands in which we live, does harm the **MENTAL, PHYSICAL AND EMOTIONAL HEALTH** of us, teachers, who are increasingly limited in our service. In the end, we are reduced to numbers, whether for the State to save money or even to reach the desired indexes by external evaluations [...]* ***(Prof. 08).***
>
> *And the results from this process are alienation and **HEALTH** weakened* ***(Prof. 06).***
>
> *It is noticeable that the school is also moving towards a business model. This can really compromise the work and the **HEALTH of the** professor* ***(Prof. 03).***
>
> *We certainly are and will be harmed in some way, especially our **HEALTH**, due to internal and external pressure* ***(Prof. 03).***
>
> *The world of work today is very competitive. This competitive climate affects our psychological, with our **HEALTH PHYSICS**, with our certainties, significantly harming our **HEALTH**. [...] I agree that this subjection to government demands to obtain necessary benefits is an important factor in harming our **HEALTH (Prof. 01).***

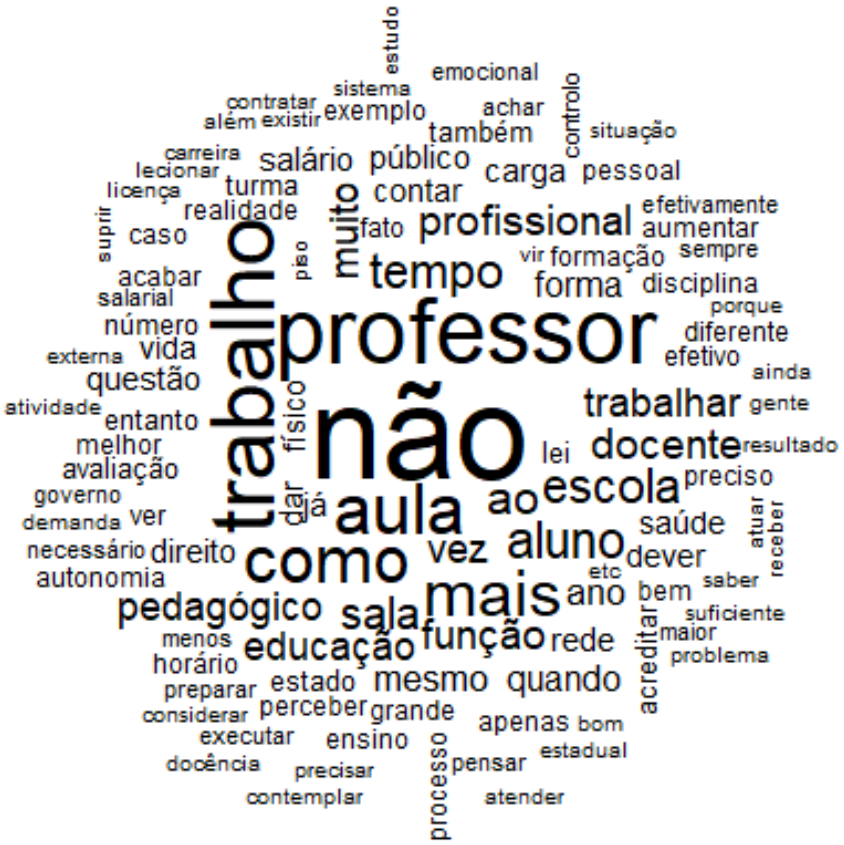
e-ISSN: 2595-4881

> *The damage of all this is the superhuman effort of the teacher, which has immediate effects on their* ***PHYSICAL/MENTAL HEALTH*** *(**Prof. 02**).*
>
> *The precariousness of teaching work is the great villain of the illness of education professionals, as the accumulation of work, as well as uncertainty, favors the* ***EMOTIONAL*** *breakdown of professionals* ***(Prof. 04).***

- Word Cloud

The Word Cloud groups words and graphically organizes them according to their frequency (CAMARGO; JUSTO, 2013). In this perspective, according to the word cloud indicated in Figure 12, a descending sequence of number of occurrences can be observed in the expressions "no", "teacher", "work", "school, "student", "class" , “room”, “function”, “education”, “pedagogical”, “teacher”, “work”, among others with lower incidence.

Figure 12: Word cloud



Source: Research data (IRAMUTEQ) (2020)

e-ISSN: 2595-4881

The image generated by the software, from the Word Cloud, denotes, among other issues, the precariousness of the work of the teaching working class, materialized, above all, based on the way in which each work is organized along the lines of contemporary capitalism. Teachers' speeches show that the phenomenon of precariousness considerably affects their locus of work. Notwithstanding, marked by flexibilization, intensification, non-compliance with educational legislation, flexibilization of contractual forms, loss of autonomy over the work process, accountability, competitiveness, deprofessionalization, degradation, education and labor of the working class service of capital, psychic suffering, illness and alienation of this professional category.

Conclusion

In line with the objective of this study, it was possible to describe the use of the IRAMUTEQ software as a tool to support data processing in qualitative research (and/or quali-quanti). Thus, although IRAMUTEQ guarantees accurate results, the use of the software does not replace the researcher's protagonism. It is noteworthy, from this perspective, that all stages of the research require seriousness and knowledge of the object under investigation, which permeates the choice of approach, the coherent and consistent discussion with the literature and the quality of the instrument to carry out the interview in the collection of data, so that the preparation of the corpus allows a better processing and presentation of the results generated by the software. In this sense, the researcher continues to lead the research, and his role is valued by IRAMUTEQ, which enables the interpretation of results already processed with scientific rigor.

From the approach of the IRAMUTEQ software as a tool in data processing in qualitative research (and/or quali-quanti), finally, the need to deepen and broaden the discussion around the use of the software, and also, the development of skills in its use (development of technique). We face this study with the content of continuity, since the use of software in qualitative research has been growing considerably. What we consider here, therefore, are the possibilities of IRAMUTEQ as a data processing tool. Finally, it is expected

that the discussions derived from this investigative process will stimulate future studies in the same direction.

References

AQUINO, Jairo Santos. **As representações sociais acerca do ensino superior a distância da UFES: a percepção de egressos dos cursos de administração e física**. 2015. 91 f. Dissertação (Mestrado em Gestão Pública) – Programa de Pós-graduação em Gestão Pública do Centro de Ciências Jurídicas e Econômicas da Universidade Federal do Espírito Santo, Vitória, Espírito Santo, 2015.

CAMARGO, B. V. ALCESTE: Um programa informático de análise quantitativa de dados textuais. In: Moreira, A. S. P.; Camargo, B. V.; Jesuíno, J. C.; Nóbrega, S. M. (Eds.) **Perspectivas teórico-metodológicas em representações sociais**. João Pessoa: Editora da UFPB, 2005. (p. 511-539)

CAMARGO, Brigido Vizeu; JUSTO, Ana Maria. IRAMUTEQ: um software gratuito para análise de dados textuais. **Temas em Psicologia**, v. 21, n. 2, p. 513-518, 2013.

MARCHAND, P.; P. RATINAUD. L'analyse de similitude appliqueé aux corpus textueles: les primaires socialistes pour l'election présidentielle française. In: **Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles**. JADT 2012. (687–699).

PIOVEZAN, Patricia Regina. **As políticas educacionais e a precarização do trabalho docente no Brasil e em Portugal**. 2017. 225 f. Tese (Doutorado em Educação) – Programa de Pós-Graduação em Educação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista, Marília, 2017.

REINERT, M. ALCESTE, une méthodologie d'analyse des données textuelles et une application: Aurélia de G. de Nerval. **Bulletin de méthodologie sociologique**, n.28, p.24-54, 1990.

SOUZA, M. *et al*. O uso do software IRAMUTEQ na análise de dados em pesquisas qualitativas. **Rev Esc Enferm USP**, v.52, e03353, 2018. DOI: http://dx.doi.org/10.1590/S1980-220X2017015003353.