Objective: to report the experience of the use of Analyse Lexicale par Contexte d’un Ensemble de Segments de Texte/ALCESTE in two qualitative researches in Nursing. Method: it is a descriptive and qualitative study, experience report, which demonstrates the use of the computational program ALCESTE in the analysis of a set of textual data in two dissertations. Results: the lexical analysis performed by the ALCESTE groups the semantic roots and defines them by means of classes, taking into account the function of the word within a given text. Both dissertations made use of the resource of categorization of classes generated by the program in question. In the building and analysis of their results, one of them has emphasized the use of the Descendent Hierarchical Analysis (AHD), and the other stressed the Factorial Analyses of Correspondence (AFC). Conclusion: it should be recognized the complexity of the ALCESTE, although it is versatile and transparent as it casts its statistical processes and the presented results. Descriptors: Methodological Research in Nursing; Qualitative Research; Nursing.
INTRODUCTION

One of the most harrowing stages for researchers, especially in health-related issues, is the definition of analysis strategies in qualitative investigations. At first, the choice is delineated from the cutting of the study object and the establishment of objectives in response to questioning/research problem, supported by theoretical assumptions and adopted benchmarks. Nonetheless, this choice does not end the concerns and neither represents a simple pathway towards the investigative process.

The increasing development of computer sciences and the consequent acceptance by the international scientific community of statistical softwares have fostered the choice of the researcher by means of alternative of qualitative researches. There is a set of computational programs that helps with the statistical analysis of textual and linguistic data, which are grouped into four areas: Analysis of Similarities; Multidimensional Scales; Factorial Analyses of Correspondence; Hierarchical Analysis. This organization allows the researcher to have an expansion of its sample in qualitative studies. Furthermore, it subsidizes a time saving in the analytical process, which would be held in days or weeks in a traditional perspective.

Communication is a space of combination of words, an infinite resource, since the speeches, conversations and human interactions are open systems, whose elements are the words and the movements are an undefined set of possible sequences. Accordingly, the act of working with textual data in qualitative researches requires the preparation of a corpus (body). The corpus is a complete collection of writings, or even a complete set of literature about a given subject. Therefore, it is a finite set of materials, which is determined in advance by the analyst, with inevitable arbitrariness, according to the adopted methodological purpose.

Considered as a pioneer in the use of computer sciences in content analysis, the program “Analyse Lexicale par Contexte d’un Ensemble de Segments de Texte” (ALCESTE) was elaborated in France in the 1970s, being that it is a computerized technique and also a methodology for analyzing texts. While technique, the ALCESTE investigates the distribution of vocabularies in a text written in transcriptions of oral texts. As methodology, it integrates a large amount of sophisticated statistical methods through segmentation, hierarchical classification, correspondence analysis, among other resources, configuring itself as a method of exploration and description.

For ALCESTE, the elaboration of the corpus is based on criteria required by the program itself. Firstly, the textual data should present a certain coherence to ensure a measurable logic of statistical analysis of texts. Secondly, it is useful for large texts and/or significant number of subjects. It satisfactorily meets the condition of responses coming from open questions, interviews, oral narratives, media data, papers or book chapters from a common focus.

This program performs a lexical analysis of the words of the set of texts, in four procedural steps, regardless of the origin of its production, verbal and/or written communication. It groups the semantic roots, defining them through classes, taking into account the function of the word within a given text. Thus, both it is possible to quantify and infer on the division of classes, which are defined in function of the occurrence and of the co-occurrence of words, in addition to their textual function, categorizing them just like the content analysis. Therefore, a statement is considered an expression of a point of view, that is to say, a frame of reference uttered by a research subject, being that it might be syntactic, pragmatic, semantic or cognitive. Unlike a single word, the statement is about an object from the point of view of the subject who speaks or writes.

The ALCESTE produces a bulky record of results, offering multiple sources of interpretation. It computes, for each class, a list of words that are characteristic of that same class, generated from the Chi-square Test, which reveals the associative strength between each word and its class. The lists are the results to be interpreted, considering the domain of the researcher and its knowledge of the drawing conceptual field, for achieving the empirically justified theoretical interpretation.

It is assumed that the investigative process focused on data analysis often is ended in methodological confusion, triggering statements by means of a misunderstanding of the research design. Accordingly, it is intended to minimize the effects of this choice, at the same time that one dares to speak of analysis procedure of qualitative researches, from the ALCESTE.

It aims at disseminating the use of ALCESTE in qualitative researches in Nursing, from two methodological experiences adopted in the analysis of textual data. Its construction is...
intended to contribute with researchers interested in the use of informational programs that assist in the analysis of qualitative data in the health scope.

**METHOD**

It is a descriptive, qualitative study, experience report, which demonstrates the use of the computational program ALCESTE, from two dissertations produced in the Post-Graduate Program in Nursing (PGENF) from the Federal University of Rio Grande do Norte (UFRN), in the years 2007 and 2008.

We deal with two examples of field researches, with distinct study objects, same research instrument (interview), and data worked from the ALCESTE. The first study\(^{6}\) emphasizes the Descendent Hierarchical Analysis (AHD), and the second\(^{7}\) highlights the Factorial Analyses of Correspondence (AFC).

The description of the lived experience in the first study was conducted in order to emphasize the use of the program associated with the Thematic Content Analysis, treatment and analysis of collected information. In the second research, the experience is described in order to contemplate the use of ALCESTE along with the Theory of Social Representations (TRS) with sights to analyze the set of textual data.

The methodological design and its exposure, difficulties and facilities of use, as well as the results obtained with the application of ALCESTE, were considered in both descriptions. Next, in a sense of complementarity, it is intended to demonstrate from these studies the features of ALCESTE, its use in practice and the methodological possibilities of data analysis.

**RESULTS AND DISCUSSION**

The first research experience\(^{6}\) has examined the process of nursing training and its performance in the Brazilian Unified Health System (SUS), according to the perception of students and professors from the Faculty of Nursing at the State University of Rio Grande do Norte (FAEN/UERN), using triangulation of techniques in data collection: documentary research for the text of the Pedagogical Project of Course (PPC); the closed questionnaire for students of eighth and ninth periods of the course; and the interview to professors. Only the interview has received treatment from the ALCESTE.

The data obtained from the interview were submitted to two analysis procedures: Thematic Content Analysis and Lexical Analysis of the ALCESTE. Both resources are based on the frequencies and the Chi-square Test, being that they might be used for the same set of texts.\(^{8}\)

In the preparation of the corpus for the ALCESTE, it should be highlighted the attention to the materials used in their preparation, which should have only one thematic focus, one specific theme, due to it is the only part handled by the researcher, as a guarantee of relevance, homogeneity and synchronism.\(^{2}\) In this perspective, the variables called Units of Initial Context are defined (UCIs) that serve to individualize the indicative text of each interview.

Its formatting does not allow the use of quotation marks, apostrophes, percentages, asterisks, because they are non-specific universal characters, which should be replaced, taking into account the limitations of dictionaries.\(^{1,9}\) Accordingly, the preparation of the corpus is recommended not to use slangs, conjunctions, articles, numerals, pronouns, auxiliary verbs and adverbs, as it substantially improves the result to be obtained.

The ALCESTE works with calculations performed on the co-occurrence of words in textual segments, seeking to distinguish classes of words that represent distinct forms of speech on the topic of interest of the investigation. In the content analysis, it identifies the thematic plurality present in a set of texts and considers the frequency of these themes in the entirety, allowing, via comparison between the elements of the corpus (words or sentences), the composition of groupings of elements of closer meanings, allowing the formation of general categories of content.\(^{8}\)

In the first research, the UCIs were considered from 14 interviews conducted with professors, each one containing three questions. They were separated by command lines, containing asterisks that represent the categorical variables given by the researcher, just like in the example below:

```
**** *cgd_1 (*cgc_1) *gn_2 *tgr_1 *trd_1 *cpt_1 *prg_1 *id_1 *cgr_1 *dil_3 *rgt_2 *vep_2
```

The coding has attended the following caption: professors (cgd), coordinators (cgc), gender (gn) length of graduation (tgr), working time in teaching (trd), education level (cpt), questions of the interview script (prg), age (id), graduation course (cgr),
Use of the ALCESTE in the analysis of... Azevedo DM de, Costa RKS, Miranda FAN de.

After its preparation, the corpus was processed by the ALCESTE in 20 seconds. Among the results generated by the ALCESTE, it should be highlighted the AHD, as shown in Figure 1.

It should be noted the speed of processing, which becomes in the main advantage related to the use of the program in question, since the stages of the Thematic Content Analysis that involve the procedures of classification, aggregation of information and identification of empirical categories usually need a very longer time and, occasionally, exhaustive, if they are manually worked. The obtained classes are constituted by a set of Units of Elementary Context (UCEs), which are organized according to the classification and distribution of the respective vocabularies. A UCE is randomly defined by the program in question, keeping a sequential interval of three lines or from eight to ten words and/or even a piece of text. The UCEs are defined by the program in function of their vocabularies, through which the significant matrices are crossed. As the following example (Figure 2), the program defines the composition of UCEs in classes, from the intersection or crossing of the corpus and the percentage of exploitation of semantic roots of the words.

The classes are also comparatively represented with their degree of occurrence and similarity in dendograms, with basis on the Descendent Hierarchical Classification. In this first study, the conflation of the two
analytical methods followed some methodological designs.

Initially, the classes and their UCEs were submitted to a thorough process of reading and interpretation of the findings of the lexical analysis, with a view to achieving the understanding of their internal coherence. Subsequently, they were confronted with the thematic analysis. Both analyzes were permeated by the adopted theoretical design, sometimes each one separately, sometimes overridden with confrontation of the manually performed thematic axes and the classes generated by the ALCESTE.

Through the combined process, the five classes of the study were interpreted from their meanings, resulting in an equal number of thematic categories (Figure 3).

The combined method of analysis was built in the intersection of the literature review, in other words, from the: a) historical, social and political contextualization of changes in the health care model in Brazil; b) demands in the area of vocational training; c) changes in the paradigm of universal policies that are related to health, education and work; and d) presentation and review of the pedagogical project and of the curriculum of the FAEN/UEERN.

The analysis and discussion possessed as an intrinsic requirement the interpretation of each UCE not only in its articulation with the frame of reference of the study, but from other UCEs of their respective class and of other classes with which it maintains relationships, as well as in the analyzed thematic axes. Next, the second research experience will be presented.

The second research made use only of the lexical analysis of the ALCESTE, with basis on the seizure of social representations of family members of users of Psychosocial Care Centers (CAPS) of the municipality Natal-RN, Brazil, conducted with 28 family members distributed in the CAPS AD North and East and CAPS II West and East. The method of semi-structured interview was employed as a research tool, which was comprised of three parts that characterized the user and the family member, followed by an interview script with seven questions.

The criteria for analysis of discursive material have included the transcription of the interviews and the fluctuating reading. The collected material, for being extensive and bulky, required the fluctuation as a resource to define the variables in the preparation of the corpus, as well as to understand them in the results generated by the ALCESTE.

This pathway was essential to identify the meanings arising from this, without the categorical pretension, but for ensuring a general reading of the computational stuff before the contents of the interview in natura. It should be emphasized that the greatest demand for work and time took place in the preparation of the interviews and transcription to a single file, a corpus required for processing in the scope of the ALCESTE, reversing itself in new and successive fluctuations in the text.

The definition of UCIs obligatorily demanded the specific formatting and warranty of the textual unit in its homogeneity, relevance and representativeness of the results. The categorized variables were developed in the following way:

**** caps_1 *gpus_1 *pfac_1 *entr_01 *ques_1

The variables were related to the identification of the researched service (CAPS), the degree of kinship with the service user (gpus), the participation of the family member in the activities of the CAPS (pfac) and the identification of the interview question (ques), successively numbered (entr).

It is worth to remember that the corpus in the ALCESTE should be considered as a set of simple statements that reflect the points of view of the individuals, and not a set of propositions that reflect a generalist view of the world.
When performing the lexical analysis of a set of textual data, the ALCESTE groups the semantic roots and defines them by means of classes, taking into account the function of the word within a given text. Thus, it is possible to quantify and infer about the delimitation of classes, which are defined in function of the occurrence, the co-occurrence of words and their textual function.9

In the second research, the six classes generated by the ALCESTE were considered as a categorical system belonging to the discursive set of family members on the participation in the activities of the CAPS, according to affinity and greater number of UCEs, obtained in the lexical analysis. The classes might indicate social representations or imagetical fields on a given object, or just aspects of the same social representation. Such discovery will depend on its content and its relationship with factors linked to the general plan of each investigation12, suggesting a pre-categorization in defining the generated classes. Thus, it is a task of researchers to interpret and discover the relationships established in the general context, the general frame of the theoretical benchmark, the query to the dictionary, generated by ALCESTE; the lists of present grammatical categories, the reduced forms, the Descendent Hierarchical Classification and the Factorial Analyses of Correspondence.

The findings of the second study were analyzed in the light of the Theory of Social Representations (TSR)13 and the Theory of Central Core15, and the definition of the categories followed the sequence of semantic contents, in descending order, through a system of nominal correspondence between the obtained classes and the developed categories (Figure 4).

The nominal definition of categories followed the implicit order of the theoretical-conceptual building designed for the study, in other words, it has prioritized the importance recommended by the literature for obtaining qualitative results from the ALCESTE.

The established correspondence between classes and categories is not mandatory in representational researches, since it related to the choice and the ability of the researcher to give support in the seizure and in the reading of the representational field, from the images offered in the Factorial Analyses of Correspondence9, shown in Figure 3.

Figure 4. Correspondence between the Descendent Hierarchical Classification and the Categorical System. Natal-RN, Brazil, 2008.
It should be emphasized that the “logic” of the Factorial Analyses of Correspondence is the following: the more distant the elements are arranged in the plan, the less they “speak” of the same things. Regarding the arrangement of groupings in opposite poles in the plane of the axes, it does not necessarily indicate a relationship of semantic opposition of the aforementioned groupings. As seen in Figure 2, a lexical field is only possible from the availability of others, and this relationship between fields might also be held in a complementary manner.

As a result, the ALCESTE provides the two axes, x and y, to the disposal of the reduced forms and classes in the plan, and the information about which of the axes comprises more strongly to the disposal of the elements. From this factorial arrangement, it should be started the reading of the results in which the researcher, remaining itself true to its theoretical-conceptual design, begins his/her interpretation.

CONCLUSION

The qualitative studies prioritize the discursive productions, translated into a collection of texts that requires a preparation and a rationale for their selection, in accordance with epistemological arguments, since the sampling and inclusion criteria of the subjects need further clarifications, in addition to ensuring a logical, valid, reliable and trustworthy basis for the study.

It is recognized the complexity of the ALCESTE, although it is versatile and transparent as it casts its statistical processes and the presented results in word lists, configuration of classes, figures and charts resulting from the Descendent Hierarchical Analysis and the Factorial Analyses of Correspondence.

From the examples shown in this paper, it is assumed the feasibility in the analysis of qualitative researches with phenomena of research in Health and Nursing. Nevertheless, new experiences and methodological discussions need to be published to bring even more readers and researchers of this investigative possibility.

REFERENCES


Use of the ALCESTE in the analysis of...

Submission: 19/07/2012
Accepted: 01/06/2013
Published: 15/07/2013

Correspondence Address
Dulcian Medeiros de Azevedo
Universidade do Estado do Rio Grande do Norte
Curso de Graduação em Enfermagem / Campus Caicó (Prédio do CAIC)
Rua André Sales, 667
Bairro Paulo XI
CEP: 59 300 – 000 – Caicó (RN), Brazil