ARTICLE
https://doi.org/10.22239/2317-269x.01429

Use of the IRaMuTeQ software to analyze society's contributions in a regulatory process conducted by the National Health Regulatory Agency

Utilização do *software* IRaMuTeQ na análise de contribuições da sociedade em processo regulatório conduzido pela Agência Nacional de Vigilância Sanitária

Thiago Silva Carvalho[I,*]
Daniel Marques Mota[I]
Flávio Saab[I,II]

ABSTRACT

Introduction: The use of software for analyzing texts arising, for example, from open questions contained in electronic forms has been recommended in the scientific literature. **Objective:** To demonstrate the use of IRaMuTeQ software, as a support tool for the analysis of social contributions of regulatory process conducted by the Brazilian Health Regulatory Agency (Anvisa). **Method:** Case study that used a textual *corpus* composed of the answers of the society about the preliminary report of the Regulatory Impact Analysis of the Nutrition Labeling that was submitted to the Public Subsidy Taking (TPS) by Anvisa. Contributions were organized into two *corpus*, according to two groups (12 stakeholders and the general public), which underwent the following analyses in IRaMuTeQ: lexicographic analysis to determine the frequency and distribution of active words; descending hierarchical classification (CHD) to categorize active words in classes; and correspondence factor analysis (CFA) to visualize the proximity of words and classes from CHD. **Results:** 346 contributions were analyzed, of which 12 were part of the textual *corpus* of the 12 stakeholders. Four words (being, consumer, information and nutritional) of the five most frequent were equal in both textual *corpora*. CHD analyses of both *corpora* resulted in three major groups, with a slight difference in the number of lexical classes produced. CFA resulted in three lexical worlds for both textual *corpora*. **Conclusions:** The lexical analyses, using IRaMuTeQ, allowed to discriminate the argumentative tendencies as well as to understand the relations between words and classes. The IRaMuTeQ can be considered a useful tool to support the routine analysis of open questions provided in forms submitted to social participation mechanisms promoted by Anvisa. The use of software can make decision-making more agile and reliable, since it allows the public authority to know and consider the contributions of society presented in the participatory process.

KEYWORDS: Brazilian Health Regulatory Agency; Social Participation; Regulatory Impact Assessment; Software IRaMuTeQ; Food Labeling

RESUMO

Introdução: O uso de *software* para análise de textos advindos, por exemplo, de perguntas abertas contidas em formulários eletrônicos tem sido recomendado na literatura científica. **Objetivo:** Demonstrar a utilização do *software Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* (IRaMuTeQ) como ferramenta de apoio à análise de contribuições sociais de processo regulatório conduzido pela Agência Nacional de Vigilância Sanitária (Anvisa). **Método:** Estudo de caso que utilizou um corpo textual formado pelas contribuições da sociedade ao relatório preliminar da Análise de Impacto Regulatório da Rotulagem Nutricional de Alimentos submetido à Tomada Pública de Subsídios (TPS) pela Anvisa. As contribuições

[I] Gerência-geral de Regulamentação e Boas Práticas Regulatórias, Agência Nacional de Vigilância Sanitária (Anvisa), Brasília, DF, Brasil

[II] Programa de Pós-graduação em Administração, Universidade de Brasília, Brasília, DF, Brasil

* **E-mail:** thscarvalho@gmail.com

Received: Nov 20, 2019
Approved: Feb 11, 2020

foram organizadas em dois *corpora*, segundo dois grupos (12 *stakeholders* e público em geral), os quais foram submetidos as seguintes análises no IRaMuTeQ: análise lexicográfica para determinar a frequência e distribuição das palavras ativas, classificação hierárquica descendente (CHD) para categorizar palavras ativas em classes e análise fatorial de correspondência (AFC) para visualizar a proximidade das palavras e das classes oriundas da CHD. **Resultados:** Um total de 346 contribuições foi analisado, das quais 12 compuseram o *corpus* textual dos *12 stakeholders*. Quatro palavras (ser, consumidor, informação e nutricional) das cinco mais frequentes foram iguais nos dois *corpora* textuais. As análises da CHD de ambos os *corpora* resultaram em três grandes grupos, com uma pequena diferença no número de classes lexicais produzidas. As AFC resultaram em três mundos lexicais para ambos os *corpora* textuais. **Conclusões:** As análises lexicais, utilizando o IRaMuTeQ, permitiram discriminar as tendências argumentativas, bem como compreender as relações entre palavras e classes. O IRaMuTeQ pode ser considerado uma ferramenta útil para apoiar a rotina de análise de questões abertas previstas em formulários submetidos aos mecanismos de participação social promovidos pela Anvisa. A utilização do *software* pode tornar mais ágil e confiável a tomada de decisão, uma vez que permite que a autoridade pública conheça e considere as contribuições da sociedade apresentadas no processo participativo.

PALAVRAS-CHAVE: Agência Nacional de Vigilância Sanitária; Participação Social; Análise de Impacto Regulatório; *Software* IRaMuTeQ; Rotulagem de Alimentos

INTRODUCTION

Regulatory Impact Analysis (RIA) is defined as an "evidence-based systematic process of analysis that seeks to assess, based on the definition of a regulatory problem, the possible impact of the options available to achieve the intended objectives, with the purpose of guiding and informing decision-making processes"[1]. First launched in the United States in 1981, it has been strongly promoted by international organizations, like the Organization for Economic Cooperation and Development and the World Bank, with positive results after implementation in several developed and also developing countries[2]. It requires the broad and transparent engagement of all society segments that hold a stake in government regulation processes[3,4].

In Brazil, particularly at the National Health Surveillance Agency (Anvisa), the internalization of RIA as a practice to drive efficiency, transparency and greater ability to respond to regulatory processes gained momentum after the publication of Ordinance n. 1.741, in December 12, 2018 (hereinafter, Ordinance)[1]. This Ordinance provides for the society's engagement in several participatory moments in the conduction of RIAs[1], aiming, above all, to encourage the public debate on proposals for regulatory intervention that protect and promote the health of the population, without, however, burdening citizens and the country's productive sector unnecessarily and excessively.

One of the participatory moments foreseen for the conclusion of RIAs conducted by Anvisa is the submission of the preliminary report to inputs from the society. This document presents narrative content, especially about the analysis and definition of the regulatory problem to be addressed, justification for Anvisa's initiative, international experiences with similar problems and identification and comparison of regulatory interventions to deal with the problem. As provided for in the Ordinance, the Public Input Survey (TPS) is a social participation mechanism open to the population, carried out within a predetermined period, to collect data and information, in writing, on the preliminary report of the RIA[1].

If, on the one hand, the Ordinance consolidates the Agency's understanding that society's participation in RIAs is fundamental to promote transparency and ensure a space for exercising citizenship, on the other hand, it poses an important challenge for Anvisa: how to go about the analysis of the written inputs received by Anvisa in its participatory regulatory processes?

The use of software to analyze text from open-ended questions in electronic forms has been recommended in the scientific literature for several reasons. Some of them are: i) the need to deal with large volumes of written answers; ii) it enables the improvement of the analyses, especially because of the integration of quantitative and qualitative domains; and iii) it enables more objective and deeper data interpretation by the researcher[5,6].

We found no studies in the literature that did the text analysis of society's inputs to topics related to government regulation in Brazil using the *Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* (IRaMuTeQ) software, although this software has been widely used in the analysis of qualitative data in postgraduate programs in health-related areas[7].

This study aimed to demonstrate the use of IRaMuTeQ as a support tool for the analysis of social inputs coming from the regulatory process conducted by Anvisa. The choice of the IRaMuTeQ software was due to its relevance in more rigorous and systematic processes of text analysis, which is not yet widespread in the field of health regulation in Brazil, and because of its user-friendly features[8,9].

METHOD

Study design

Case study that used the text *corpus* of the society's answers to the following question: "Was the main problem presented

correctly identified?", from the preliminary report of RIA of the Nutritional Labeling of Food, submitted to a TPS by Anvisa. The regulatory problem defined in this RIA was: "Difficulty in using nutrition labeling by Brazilian consumers"[10]. With the TPS, Anvisa aimed to survey data, information or evidence on the preliminary report of RIA in order to support its decision-making process.

The justifications for choosing the question were: the analysis and definition of the regulatory problem are the most important aspects of an RIA[11]; and the use of a single question would facilitate statistical analysis on the IRaMuTeQ software[12].

The TPS of the preliminary report of Nutritional Labeling for Food RIA aimed to present the society with proposals for regulatory solutions to improve the consumers' understanding of the main nutrition facts of food marketed in Brazil and prevent situations that lead to misunderstanding of these nutrition facts. Additionally, the TPS aimed to collect inputs to help improve these solutions and/or propose others, as well as support other important cornerstones of the RIA, such as the regulatory problem[13].

The IRaMuTeQ software

IRaMuTeQ, version 0.7 alpha 2, is free software, anchored in the statistical environment of R software and in the Python language, which offers different types of statistical analysis of qualitative data, which are processed and combined into a text *corpus*[5,12]. The results of this study were based on tools offered by IRaMuTeQ to enable the interpretation of the data most commonly described by the words of the text *corpus*.

Collection of information

Society's inputs were collected via electronic form available on the Anvisa portal for 60 days, from May 25 to July 24, 2018. The form had 22 questions distributed in four sections: 1) Society's perception (two questions); 2) Regulatory Impact Analysis (nine questions); 3) Graphic design and communication (ten questions); and 4) Adjustment period (one question). The question "Was the main problem presented correctly identified?", used in this analysis, was in the 2nd section of the form.

Information modeling

Society's inputs were organized in an Excel® spreadsheet and then prepared to be processed within IRaMuTeQ in four stages: 1) visual inspection and use of Excel® features to identify and exclude repeated inputs and text elements that did not add any value, like connectors and special characters. A single text, representing the repeated inputs, was kept for the purpose of assembling the text *corpus* to be studied. The use of a single text, in this case, aimed to avoid statistical bias and ensure the correct result of the analysis. Additionally, references to bibliographic citations were not considered in the assembly of the text *corpus*; 2) inputs made in English were electronically translated via Google Translate, and the adjustments were made by one of the researchers to enable their incorporation in the analysis; 3) written inputs with less than 40 words were grouped into longer texts to enable the software to perform the statistical analysis of the data; and 4) the result of this whole process of treatment of inputs made up the final text *corpus* for analysis within IRaMuTeQ.

The final text *corpus* included the active forms, that is, nouns, verbs (supplementary or not), adjectives (supplementary or not), adverbs (supplementary or not) and unrecognized forms and supplementary forms that, in this case, included only supplementary nouns and numeric adjectives. In this study, the active forms were called active words.

Information analysis

The final text *corpus* was divided into two social groups of interest – 12 *stakeholders* and the general public – in an attempt to demonstrate the potential for discrimination in the analyses enabled by IRaMuTeQ, with regard to the identification of arguments with similar and distinct characteristics, according to the positions of the referred groups about the studied question.

The inputs from 12 stakeholders were grouped into a text *corpus* representing the first social interest group. The 12 selected stakeholders were: Brazilian academic community (participant A), international academic community (participant B), consumer protection body (participant C), Brazilian food industry (participants D, E and F), international food industry (participant G), international health agency (participants H and I), healthcare professional association (participants J and L) and federal government agency (participant M). This selection sought to include segments of the society with the potential to present agreeing and disagreeing positions about the question. This social group functioned as a control of the other studied group.

The rest of the inputs were grouped into a second text *corpus*, which was considered to represent the view of the general public.

The two text *corpora* were subjected to three analyses by IRaMuTeQ[14] and described below:

1. Lexicographic analysis: involves measuring the frequency and distribution of words in the text *corpus*. The quantification of active and supplementary words in the text *corpora* was carried out after the stemming process, which consists of replacing a word with its root term (for example: "problematization" by "problem"). This process eliminates the inflected end of the word to normalize the text, reducing its complexity without compromising accuracy[14];

2. Descending hierarchical classification (DHC): as described by Mendes et al.[14], this is a type of conglomerate analysis that sorts active words into lexical classes. The analysis considers the frequency and positions of the active words in the text using data from the word contingency tables. The algorithm starts to assume that all active words initially belong

to the same class and divides it sequentially, maximizing interclass inertia and minimizing intraclass inertia. Inertia is defined as a measure of variance between the individual profiles around the average profile. Therefore, the greater the difference between the words, the greater the inertia between them. This iterative process is interrupted when a new division of words does not improve interclass inertia. In this perspective, the final number of classes is unknown *a priori*. Dendrograms are designed to illustrate the division between classes. Pearson's chi-square test was used to measure the strength of association between the active words and their respective class. The greater the Pearson's chi-square, the more likely the hypothesis of dependence between active word and class. However, this statistical test does not show how these two variables (active word and class) are related, so a correspondence factor analysis was also performed; and

3. Correspondence factor analysis (CFA): enables the graphic visualization of the proximity of words and classes from the DHC; it is not a matter of counting words, but of relationships between words. CFA was interpreted in terms of opposition between the X and Y axes[14].

The robustness of the IRaMuTeQ analyses was assessed considering: (i) the minimum use of 70% of the text segments in the lexicographic analysis; (ii) Pearson's chi-square value above 3.85 in DHC analyses, representing a satisfactory separation between classes; and (iii) sum of the factors of the axes of the CFA graphs close to 100%. The values were based on a study by Mendes et al.[14].

The lexical classes were subjectively named by consensus among the authors, based on the composition of the most prevalent words in an attempt to represent the ideas of each class[14,15].

Ethical considerations

The inputs from the society were obtained in the context of health surveillance initiatives, a situation in which consideration by the Research Ethics Committee is not required. Ethical aspects of the National Health Council (CNS) Resolution n. 510, of April 7, 2016, were observed[16]. The results presented here ensure the individual confidentiality of the participants.

RESULTS

564 inputs were presented to answer the following question "Was the main problem presented correctly identified?". Most inputs came from the industry (n = 196; 34.7%), consumers (n = 154; 27.3%), healthcare professionals (n = 68; 12.0%) and educational institutions (n = 68; 12.0%).

221 (39.2%) repeated inputs were identified, 177 (80.0%) of which came from a single institution belonging to the industry, followed by 29 (13.1%) from two foreign universities and 15 (6.8%) from a consumer protection body. Only one answer from each of these participants was added to the other single inputs (n = 343), resulting in a final text *corpus* composed of 346 inputs.

Lexical analysis of the 12 stakeholders

The lexicographic analysis of the text *corpus* formed by the 12 inputs (12 texts) produced 3,244 occurrences (words and forms); of which 357 (11.0%) are words that appear only once in the text *corpus*. After stemming the text *corpus*, 678 active words and nine additional words were obtained. Of the active words, the five most frequent were: (to) be (n = 66 times), consumer (n = 35), problem (n = 31), information (n = 29) and nutritional (n = 28 times). 93 text segments were obtained, of which 66 (71.0%) were used by IRaMuTeQ.

The DHC analysis of the active words produced five lexical classes divided into three large groups. Group A, formed by classes 1, 5 and 6, Group B, represented by classes 2 and 4, and Group C, formed separately by class 3. This result is shown in the dendrogram, which also contains the list of the main active words by class (Figure 1).

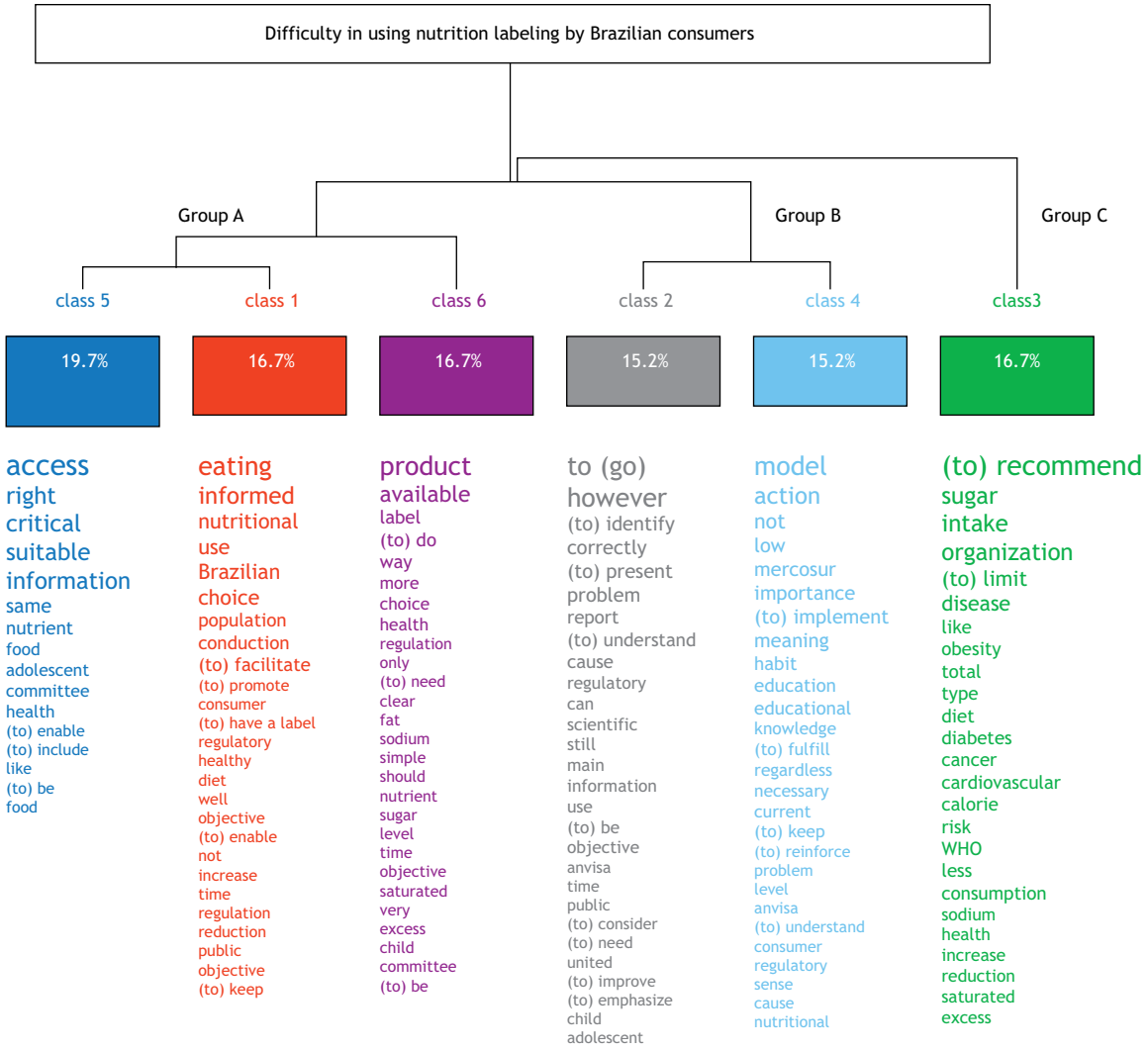
As shown in Figure 1, Group A was the most representative, totaling 53.1% of the text segments. Class 5, with 13 (19.7%) text segments, was the most representative, followed by classes 1, 3 and 6, each with 11 (16.7%) text segments. Pearson's chi-squares greater than 3.86 were obtained for all words in every class.

The lexical classes received the following names: a) class 1: use of labeling by Brazilian consumers in their food choices; b) class 2: debate on the identification of the regulatory problem; c) class 3: relationship between eating habits and risk of disease; d) class 4: importance of nutritional education and harmonization with the Southern Common Market (Mercosur); e) class 5: right and access to information; and f) class 6: the role of clear labels in promoting healthy choices among the available products.

The CFA resulted in five factors that explain 32.3%, 22.1%, 17.7%, 14.6% and 13.3% of the model. The two factors with higher percentages explain an accumulated 54.4% of the model, which are represented on the X and Y axes of the Cartesian plane (Figure 2).

On the X axis, representing 32.3% of the distribution in the text *corpus*, class 3 (negative X axis) is clearly separated from class 4 (positive X axis). There is a greater predominance of central positioning of classes 1, 5 and 6 on this axis. On the Y axis, representing 22.1% of distribution in the text *corpus*, classes 3 and 4, although on the positive side of the Y axis, are clearly positioned on different quadrants. Classes 5 and 6 are entirely positioned on the negative Y axis. There are words positioned at the intersection of the X and Y axes.

The combination of the two axes (X and Y) offers a two-dimensional view and differentiates the text *corpus* into three lexical worlds. The first, a world of positive X and Y axes, on the upper right quadrant, showing class 4, which corresponds



* The words at the top of the list and the largest size have the most influence on the class.
Source: Prepared by the authors, 2019.

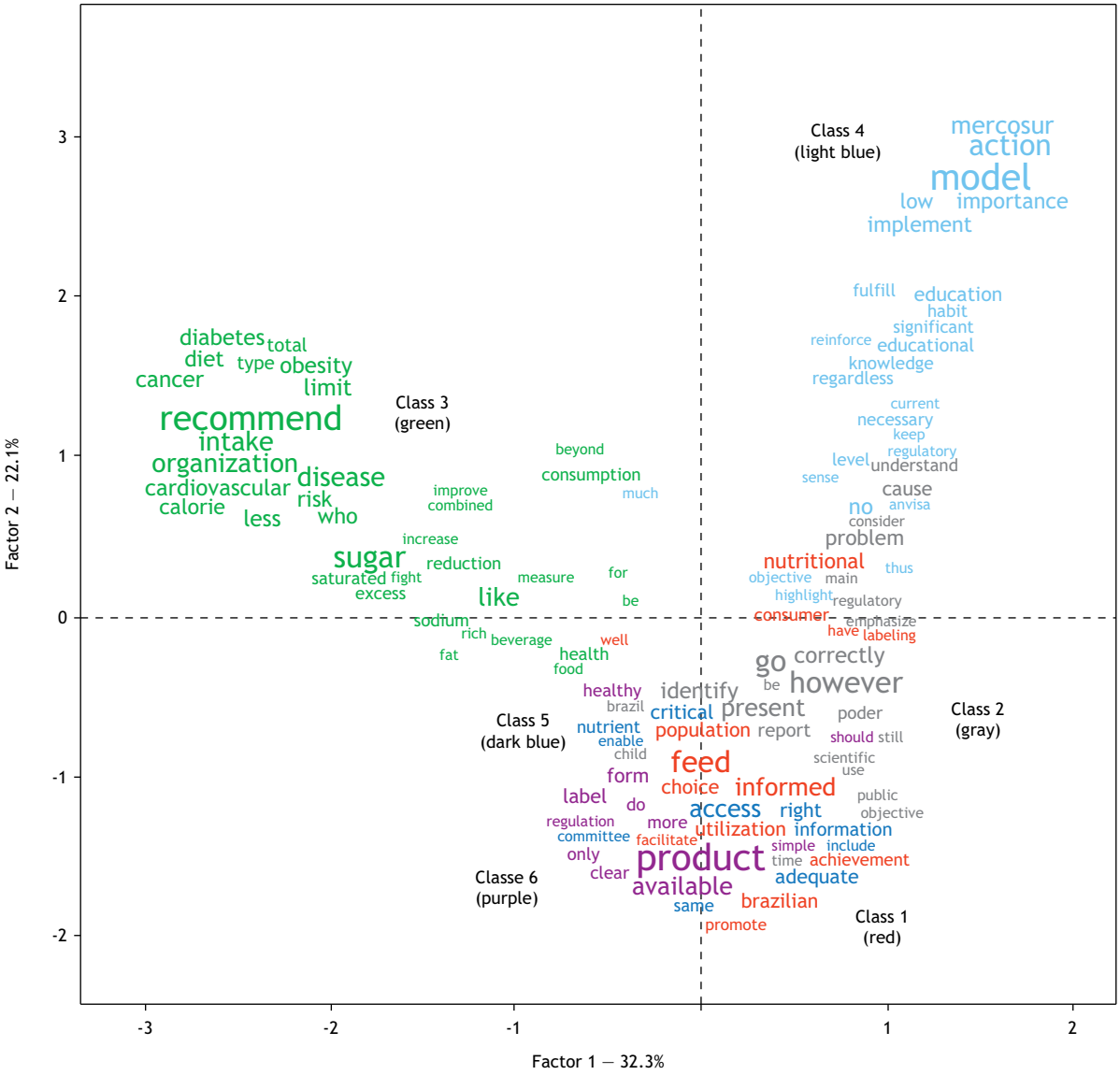
Figure 1. Dendrogram of the six lexical classes obtained from the descending hierarchical classification of the active words from the inputs from the 12 stakeholders (Total text segments = 66)*.

to the topic of "importance of nutritional education and harmonization with Mercosur"; the second, a world of coordinates with a negative X axis and a positive Y axis on the upper left quadrant, showing class 3, corresponding to the topic of "relationship between eating habits and risk of disease"; and finally, a lexical world with a central position in relation to the X axis and negative for the Y axis. This lexical world is mainly concentrated on the lower quadrant and shows the other classes.

The graphic distribution of the arguments presented by the stakeholders indicates a gap between those presented by participants B (international academic community) and C (consumer protection body) (class 3) and by participants G (international food industry) and D (Brazilian food industry) (class 4). These respondents are distant from the arguments presented by the other stakeholders, represented by the Brazilian academic community, healthcare professional association and federal government agency (Figure 3) .

Lexical analysis of the general public

In addition to representatives from the segments of society mentioned above, inputs from other players also made up the text *corpus* of this social group. They were: civil society (n = 21), international organizations (n = 12), communication specialist (n = 8), professionals from the National Health Surveillance System (n = 7) and consulting firms (n = 4). A total of 25 participants identified themselves as Others. The text *corpus* of the general public did not have inputs from representatives of federal government agencies, since there was only one registered input, which was part of the previous social group.



Source: Prepared by the authors, 2019.

Figure 2. Correspondence factor analysis of the most frequent active words in each of the lexical classes obtained in the descending hierarchical classification of the inputs from the 12 stakeholders.

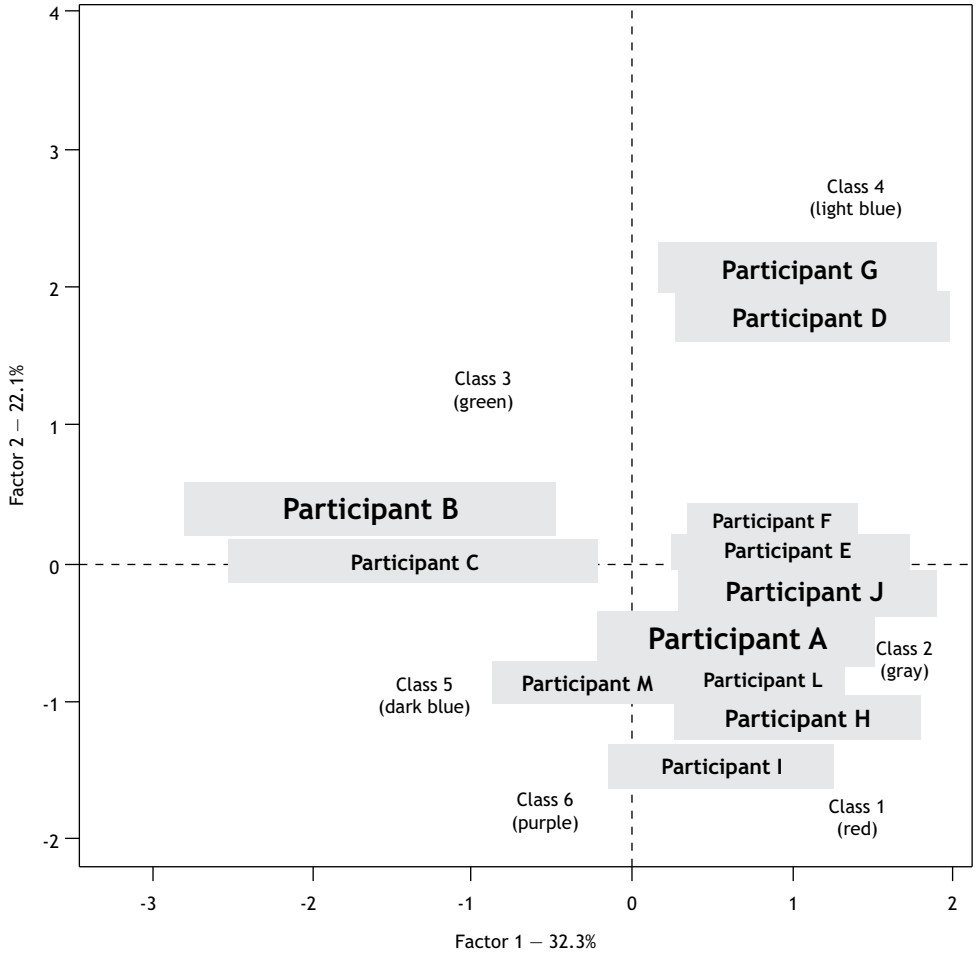
A total of 218 inputs with less than 40 words were grouped and formed a text *corpus* with 116 texts analyzed by the software.

The lexicographic analysis of this text *corpus* produced 15,084 occurrences (words and forms); of these, 1,619 (10.7%) are hapax, that is, occurrences that appear only once in the text. After stemming the text *corpus*, 1,846 active words and 25 additional words were obtained. Of the active words, the five most frequent were: (to) be (n = 431 times), not (n = 233 times), nutritional (n = 204 times), consumer (n = 184 times) and information (n = 184 times). 430 text segments were obtained, of which 390 (90.7%) were used by IRaMuTeQ.

The DHC analysis of the active words produced five lexical classes divided into three large groups. Groups A and C are made up of classes 5 and 4, respectively, while Group B is made up of classes 1, 2 and 3. Figure 4 shows this result in a dendrogram format, which included the list of the main active words by class.

As shown in Figure 4, Group B was the most representative, totaling 68.0% of the text segments. Class 3, comprising 124 (31.8%) text segments, was the most representative, followed by class 2 with 88 (22.5%) and class 5 with 75 (19.2%). Pearson's chi-squares greater than 3.91 were obtained for all words in every class.

The lexical classes received the following names: a) class 1: labels' potential to be confusing and present false information; b) class 2: importance of nutritional education; c) class 3: debate on the identification of the regulatory problem; d) class 4: need for intervention by Anvisa and harmonization with Mercosur; and e) class 5: relationship between eating habits and risk of disease.



Source: Prepared by the authors, 2019.

Figure 3. Lexical distance between the arguments presented by the selected stakeholders.

The CFA resulted in four factors that explain 33.8%, 29.7%, 19.3%, 17.2% of the model. The two factors with higher percentages explain an accumulated 63.5% of the model, which are represented on the X and Y axes of the Cartesian plane (Figure 5).

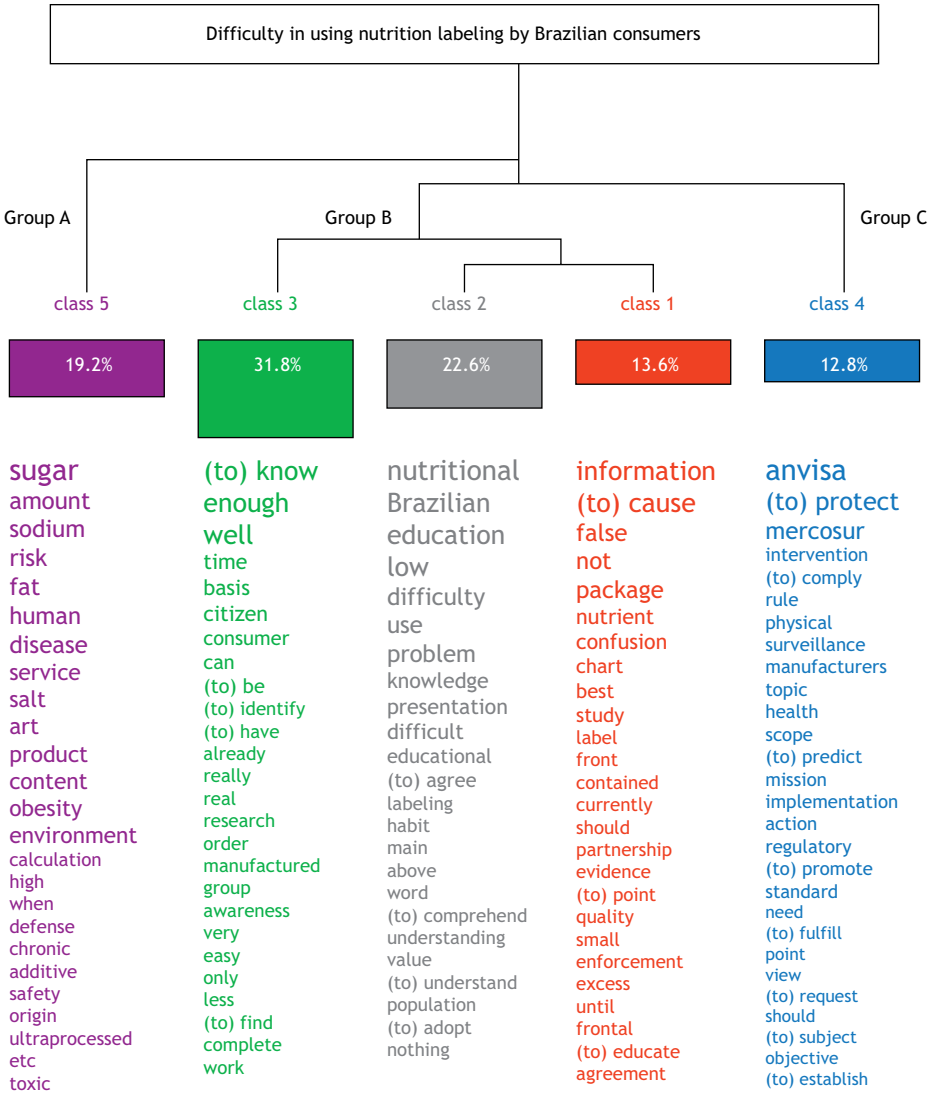
On the X axis, representing 33.8% of distribution in the text *corpus*, class 5 is entirely positioned on the positive side, whereas class 2, for the most part, is positioned on the negative side of the axis. On the Y axis, representing 29.7% of distribution in the *corpus* textual, class 4 is entirely positioned on the negative side, whereas classes 1, 2 and 5 are mostly distributed on the positive side of the axis. There are several words positioned at the intersection of the X and Y axes (Figure 5).

The combination of the two axes (X and Y) offers a two-dimensional view and differentiates the text *corpus* into three lexical worlds (Figure 5). The first, a world of positive X and Y axes, on the upper right quadrant, predominantly showing class 5, which corresponds to the topic of "relationship between eating habits and risk of disease"; the second, a world with a negative X axis and a positive Y axis on the upper left quadrant, predominantly showing class 2, corresponding to the topic of "importance of nutritional education"; and finally, a lexical world with a central position in relation to the X axis and negative for the Y axis, distributed among the lower quadrants and predominantly showing class 4 (need for intervention by Anvisa and harmonization with Mercosur).

Class 5 is made up of arguments from civil society, teaching institutions, the National Health Surveillance System and a government institution. Class 4 arguments are formed by international organizations, people/entities that identified themselves as "Others" and, mainly, by the industry. Classes 1 and 2 are formed by arguments from consultants and communication specialists, respectively. Class 3 comprises the inputs from consumers and healthcare professionals (Figure 6).

DISCUSSION

Lexical analysis using IRaMuTeQ enables us to discriminate the predominant argumentative trends about the question "Was the main problem presented correctly identified?", from the Nutritional Labeling of Food RIA, originating from the two social groups studied, as well as to analyze the



* The words at the top of the list and the largest size have the most influence on the class.
Source: Prepared by the authors, 2019.

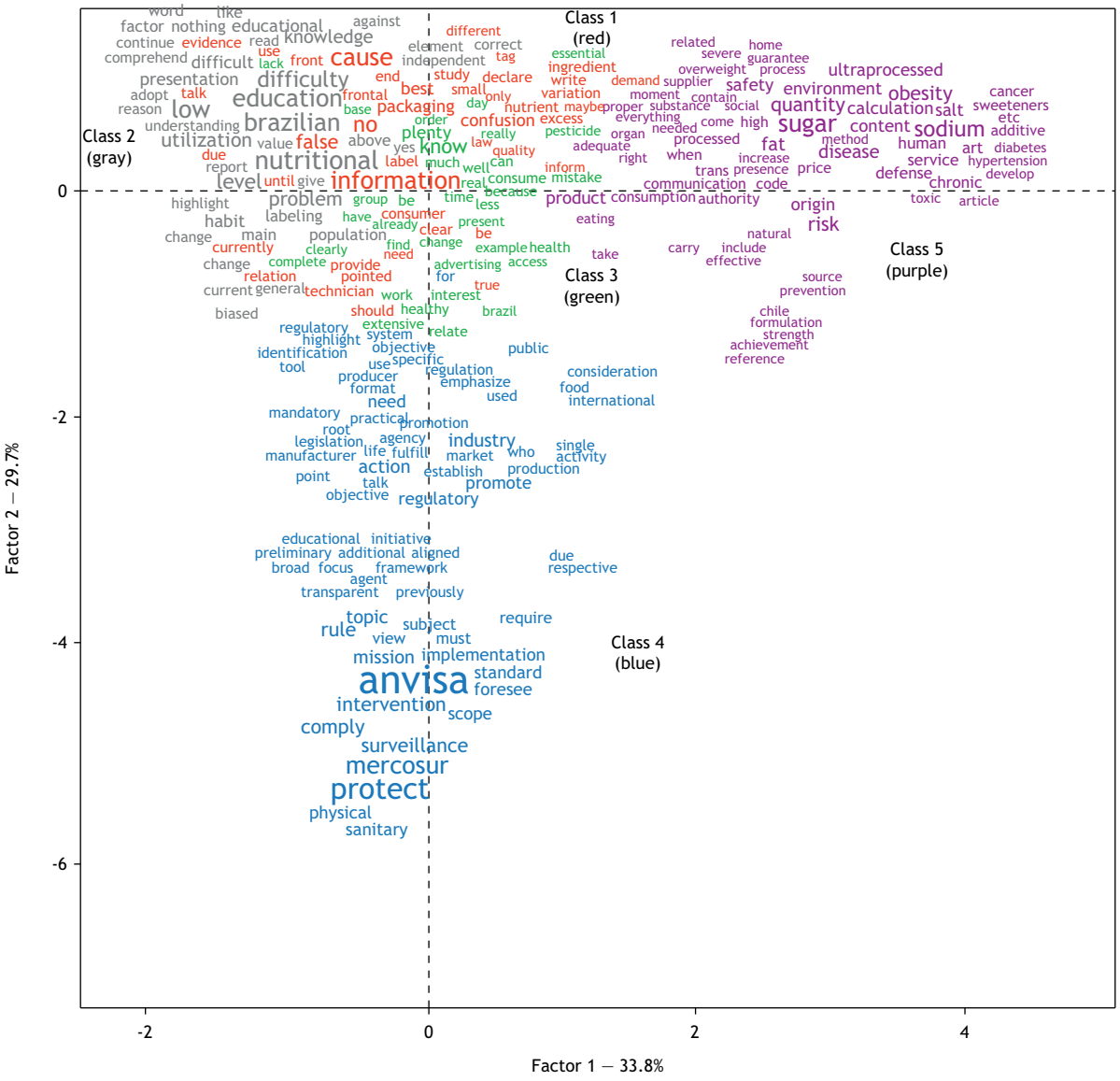
Figure 4. Dendrogram of the six lexical classes obtained from the descending hierarchical classification of the active words from the inputs from the general public (Total text segments = 390)*.

interrelationships between them. All lexical analyses were considered robust, since they met the three predetermined methodological requirements.

The DHC analyses of the text *corpora* of the social groups were equivalent in terms of division into large groups, the number of lexical classes and the predominant arguments/words that resulted in the naming of the classes. This indicates the high discrimination power of the IRaMuTeQ software in relation to the representation of the sets of ideas and their respective stakeholders.

The difference in the number of classes was due to class 5 (right and access to information), whose input came from the representative of the federal agency and was only included in the text *corpus* of the 12 stakeholders. This similarity also occurred both in relation to the predominant words (four out of five – (to) be, consumer, information and nutritional) in the lexicographic analysis and in the analyses coming from the CFA, in which the set of words is divided into three lexical worlds comprising classes with similar arguments.

In general, for both social groups, the lexical classes obtained are closely related to the problem defined in the RIA: "Difficulty in using nutrition labeling by Brazilian consumers"[5]. For example: the inputs from classes 2 (12 stakeholders) and 3 (general public), which received the same name, portray the debate on the identification of the regulatory problem, while class 1 inputs from the general public point to for the potential for labels to be confusing and present false information in food choices. There was also a clear relationship of words present in the text *corpora*



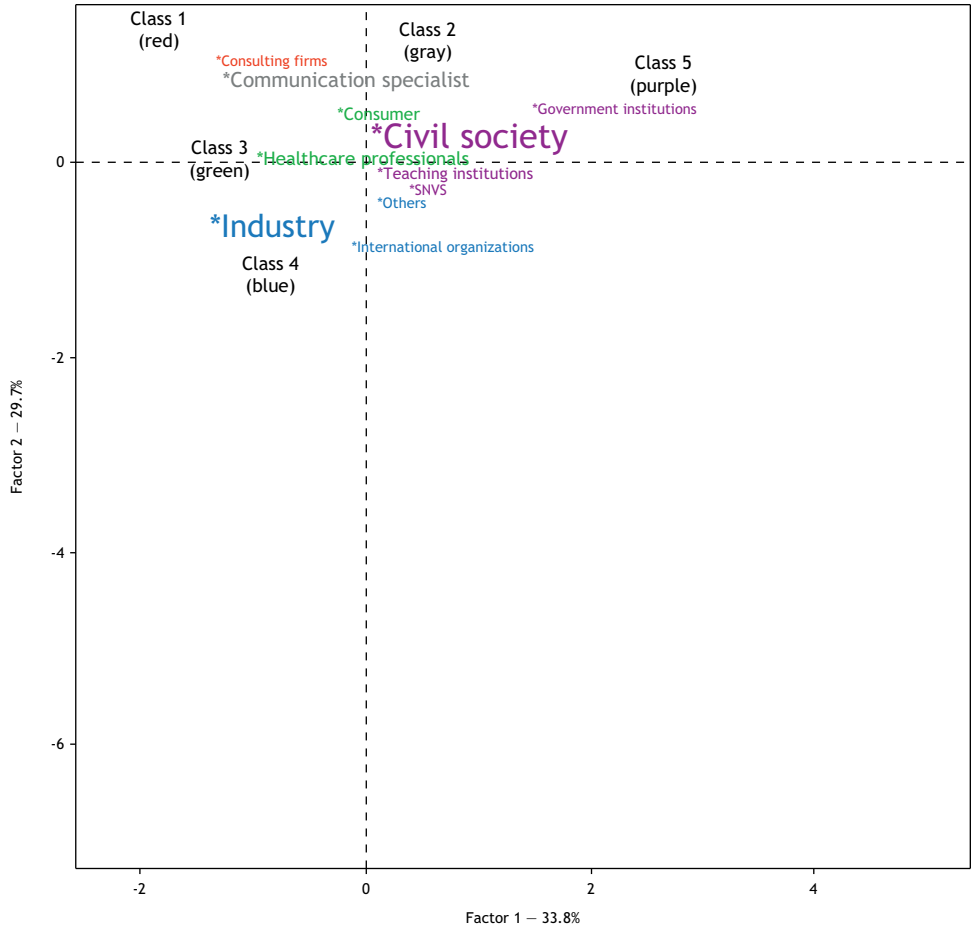
Source: Prepared by the authors, 2019.

Figure 5. Correspondence factor analysis of the most frequent active words in each of the lexical classes obtained in the descending hierarchical classification of the inputs from the general public.

with the regulatory problem. Three of the most frequent words in both *corpora*, "consumer" and "information", both substantive morphemes, and "nutritional", adjective morpheme, exemplify this relationship. The set of the five main words in each text *corpus* suggests the following sentence as a symbolic representation of the social inputs: "Nutritional information [in food] is (not) a problem for the consumer".

The study of repetitions or frequency of words in a text enables us to identify lexical worlds that may be related to similar ideological trends, conflicts, ruptures, rapprochements or argumentative oppositions about a given topic[15].

Based on the wealth of words observed in each class, this study identified lexical worlds related to rapprochement of discourses in relation to the regulatory problem. For example: classes 1, 2, 5 and 6 appear intertwined, forming a single lexical world in the text *corpus* of the 12 stakeholders, that is, there are common words in these classes, which suggests greater contextual affinity. These classes represent the inputs from seven segments of society: two international health agencies, one from the Brazilian academic community, two from the associations of healthcare professionals, one from the Brazilian food industry and a representative from a federal government agency. In relation to the text *corpus* of the general public, this intertwining occurred with classes 1, 2 and 3, which are represented by the following players, respectively: consulting firms, communication specialists, and consumers and healthcare professionals.



Source: Prepared by the authors, 2019.

Figure 6. Lexical distance between the arguments presented by the general public.

The identification of argumentative oppositions seen between classes 3 and 4, in the group of 12 stakeholders and between classes 4 and 5 in the group of the general public, were also preserved in IRaMuTeQ analyses. Classes 3 and 4 represent the inputs from the international academic community and consumer protection body and the food industry at the Brazilian and international levels, respectively. Classes 4 and 5, on the other hand, are inputs mainly from the industry and civil society, respectively.

A clear difference in the distribution of words was observed along the crossing of the X and Y axes between the *corpora*. In the case of the text *corpus* of the general public, unlike what happened for the text *corpus* of the 12 stakeholders, several words were observed at the intersection of the X and Y axes, indicating that they are common in all classes. One explanation for this is the greater diversity of participants in the general public group.

Several studies have used the IRaMuTeQ software to process texts obtained in different ways, like questionnaires/forms with open-ended questions[17]. However, we found no studies in the literature that carried out the lexical analysis of society's inputs on topics related to health regulation using the IRaMuTeQ software. Apparently, this is the first study to explore this software as a viable and responsive tool to enable the analysis of written inputs from the processes of society participation established by Anvisa.

Some of the main features of IRaMuTeQ in the analysis of written text include agile analyses, reducing the researcher's energy expenditure to extract data from a large volume of information and expanding the offer of new possibilities for interpretation and establishment of relationships, which could go unnoticed in manual work[18]. Other advantages of the software are: i) help reduce possible biases in the analysis of public inquiries by professionals involved in the task; ii) promote greater statistical rigor in qualitative analyses; and iii) enable different types of analysis, including multivariate analyses, with the support of graphs[7,9,19].

Three types of analysis of the text *corpus*, out of a total of five, performed by the software, were conducted in this study. The similarity analysis and the word cloud, which make up the lexical analyses of IRaMuTeQ, were not part of the study, and this can be considered a limitation. However, in a scope review that sought to identify the use of IRaMuTeQ in qualitative health research

in Brazil, there was a predominance of studies that used only one type of analysis done by the software, with highlights to the DHC[7]. Furthermore, similarity analysis and word cloud make a smaller contribution to public inquiry analyses when compared to lexicographic, DHC and CFA analysis[20]. It should be noted that the analyses enabled by IRaMuTeQ do not exclude other analyses, such as content analysis, and the researcher's conceptual repertoire is also critical for the interpretation of information[20].

Another possible limitation of the study is that we did not use any text editor software for spelling and textual proofreading in the information modeling stage. This proofreading occurred through visual inspection by one of the researchers. Some authors recommend the use of this tool, especially when there are bulky texts to be prepared and submitted to IRaMuTeQ[8]. In this study, the preparation of the text *corpus* for lexical analyses on IRaMuTeQ was the activity that demanded the most time and attention from one of the researchers.

Some of the society's inputs analyzed by IRaMuTeQ were not allocated to the lexical classes, given that the use of the text *corpus*, in both cases, was not 100%. However, the percentages of use were considered sufficient to perform the DHC, since a minimum retention of 70% of the text segments is defended by some authors for this type of analysis[14].

A study that investigated the use of IRaMuTeQ in the analysis of society's inputs to a public inquiry on the incorporation of Trastuzumab for the early treatment of breast cancer in the Brazilian public health system, done in 2012 by the National Committee for the Incorporation of Technology in Health, concluded that, despite the promising results, further studies were necessary for the validation of this software in analyses of public inquiries[9].

This study, in addition to verifying the functionality and potential of IRaMuTeQ, also contributes to the validation of its use in the analysis of written social inputs with a large volume of text that is difficult to manage. For Krug[8], validation corresponds to the ability of an instrument to produce adequate and accurate results that will give rise to correct conclusions.

CONCLUSIONS

The discrimination and understanding of the relationships resulting from the lexical analyses of IRaMuTeQ proved to be satisfactory, since the results were quite similar, which suggests some accuracy in the comparison of the text *corpora* of the two groups we studied. As the results of this study indicate, IRaMuTeQ can be considered a useful tool to support the routine analysis of open-ended questions provided for in forms submitted to the social participation mechanisms promoted by Anvisa. The use of the software can make the decision-making process more agile and reliable, since it enables public managers to learn and consider the inputs of society presented in the participatory process.

REFERENCES

1. Agência Nacional de Vigilância Sanitária – Anvisa. Portaria Nº 1.741, de 12 de dezembro de 2018. Dispõe sobre as diretrizes e os procedimentos para melhoria da qualidade regulatória na Agência Nacional de Vigilância Sanitária. Diário Oficial União. 14 dez 2018.
2. International Telecommunication Union – ITU. Using regulatory impact analysis to improve decision making in the ICT sector. Geneva: International Telecommunication Union; 2014[acesso 19 jul 2019]. Disponível em: https://www.itu.int/dms_pub/itu-d/opb/pref/D-PREF-BB.RPT5-2014-PDF-E.pdf
3. Ballantine B, Devonald B. Modern regulatory impact analysis: the experience of the European Union. Regul Toxicol Pharmacol. 2006;44(1):57-68. https://doi.org/10.1016/j.yrtph.2005.06.016
4. Centro de Estudos de Direito Econômico e Social – Cedes. Diretrizes gerais e guia de análise de impacto regulatório. Brasília: Centro de Estudos de Direito Econômico e Social; 2018.
5. Camargo BV, Justo AM. Iramuteq: um software gratuito para análise de dados textuais. Temas Psicol. 2013;21(2):513-8. https://doi.org/10.9788/TP2013.2-16
6. Cope DG. Computer-assisted qualitative data analysis software. Oncol Nurs Forum. 2014;41(3):322-3. https://doi.org/10.1188/14.ONF.322-323
7. Salvador PTCO, Gomes ATL, Rodrigues CCFM, Chiavone FBT, Alves KYA et al. Uso do software iramuteq nas pesquisas brasileiras da área da saúde: uma scoping review. Rev Bras Promo Saude. 2018;31(Supl):1-9.
8. Krug FS. Iramuteq em um acervo literário: amostra de um trabalho possível [Dissertação]. Passo Fundo: Universidade de Passo Fundo; 2017.
9. Carvalho VKS, Sousa MAS, Barreto JOM, Silva EM. Public engagement in health technology assessment in Brazil: the case of the trastuzumab public consultation. BMC Health Serv Res. 2019;19(1):762. https://doi.org/10.1186/s12913-019-4555-6
10. Agência Nacional de Vigilância Sanitária – Anvisa. Relatório preliminar de análise de impacto regulatório sobre rotulagem nutricional. Brasília: Agência Nacional de Vigilância Sanitária; 2018[acesso 19 jul 2019]. Disponível em: http://portal.anvisa.gov.br/tomada-publica-de-subsidios
11. Australia. Best practice regulation handbook. Canberra: Commonwealth of Australia; 2013.
12. Camargo BV, Justo AM. Tutorial para uso do software Iramuteq. Florianopolis: Universidade Federal de Santa Catarina; 2018[acesso 18 jul 2019]. Disponível em: http://iramuteq.org/documentation/fichiers/tutoriel-portugais-22-11-2018
13. Agência Nacional de Vigilância Sanitária – Anvisa. Começou: opine na discussão de rotulagem de alimentos. Brasília: Agência Nacional de Vigilância Sanitária; 2018[acesso 19 jul 2019]. Disponível em: http://portal.anvisa.gov.br/tomada-publica-de-subsidios

14. Mendes AM, Tonin FS, Buzzi MF, Pontarolo R, Fernandez-Llimos F. Mapping pharmacy journals: a lexicographic analysis. Res Social Adm Pharm. 2019;15(2):1464-71. https://doi.org/10.1016/j.sapharm.2019.01.011
15. Safaa L, Housni KE, Bédarg F. Authenticity and tourism: what TripAdvisor reviews reveal about authentic travel to Marrakech. In: Schegg R, Stangl B. Information and communication technologies in tourism 2017. Rome: Springer; 2017. p. 595-606.
16. Ministério da Saúde (BR). Resolução Nº 510, de 7 de abril de 2016. O Plenário do conselho nacional de saúde em sua quinquagésima nona reunião extraordinária, realizada nos dias 6 e 7 de abril de 2016, no uso de suas competências regimentais e atribuições conferidas pela lei Nº 8.080, de 19 de setembro de 1990, pela Lei Nº 8.142, de 28 de dezembro de 1990, pelo decreto Nº 5.839, de 11 de julho de 2006. Diário Oficial União. 7 abr 2016.
17. Gutierres LS, Santos JLG, Peiter CC, Menegon FHA, Sebold LF et al. Boas práticas para segurança do paciente em centro cirúrgico: recomendações de enfermeiros. Rev Bras Enferm. 2018;71(Supl. 6):2940-7. https://doi.org/10.1590/0034-7167-2018-0449
18. Ramos MG, Lima VMR, Amaral-Rosa MP. Contribuições do software Iramuteq para a análise textual discursiva. In: Atas do 7º Congresso Ibero-Americano em Investigação Qualitativa em Educação. Fortaleza: Universidade de Fortaleza; 2018. p. 505-14.
19. Vieira FS. A produção em saúde do IPEA: contribuição à promoção do desenvolvimento socioeconômico brasileiro. Brasília: Instituto de Pesquisa Econômica Aplicada; 2019.
20. Carvalho VKS. Engajamento público na avaliação de tecnologias em saúde no Brasil: o caso da consulta pública sobre o transtuzumabe [Dissertação]. Brasília: Universidade de Brasília; 2018.

Authors' Contribution

Carvalho TS - Conception, planning (study design), analysis, interpretation of data and writing of the paper. Mota DM - Analysis, interpretation of data and writing of the paper. Saab F - Planning (study design), analysis, interpretation of data and writing of the paper. All authors approved the final draft of the paper.

Conflict of interest

The authors report that there is no potential conflict of interest with peers and institutions, nor political or financial conflicts in this study.