

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

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

### RESUMO

Thiago Silva Carvalho<sup>1,\*</sup> 

Daniel Marques Mota<sup>1</sup> 

Flávio Saab<sup>1,II</sup> 

**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 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

### 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

<sup>1</sup> Gerência-geral de Regulamentação e Boas Práticas Regulatórias, Agência Nacional de Vigilância Sanitária (Anvisa), Brasília, DF, Brasil

<sup>II</sup> Programa de Pós-graduação em Administração, Universidade de Brasília, Brasília, DF, Brasil

\* E-mail: thscarvalho@gmail.com



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

## INTRODUÇÃO

A Análise de Impacto Regulatório (AIR) é definida como “processo sistemático de análise, baseado em evidências, que busca avaliar, a partir da definição de um problema regulatório, os possíveis impactos das opções de ação disponíveis para o alcance dos objetivos pretendidos, tendo como finalidade orientar e subsidiar a tomada de decisão”<sup>1</sup>. Lançada pela primeira vez nos Estados Unidos, em 1981, tem sido fortemente promovida por organizações internacionais, como a Organização para a Cooperação e Desenvolvimento Econômico e o Banco Mundial, obtendo resultados positivos de implementação em vários países desenvolvidos e, também, em desenvolvimento<sup>2</sup>. Requer o envolvimento amplo e transparente dos segmentos da sociedade interessados nos processos de regulação governamental<sup>3,4</sup>.

No Brasil, particularmente na Agência Nacional de Vigilância Sanitária (Anvisa), a internalização da AIR, como prática para promover a eficiência, a transparência e a melhoria da capacidade de respostas aos seus processos regulatórios, foi fortalecida com a publicação da Portaria nº 1.741, de 12 de dezembro de 2018 (doravante, Portaria)<sup>1</sup>. Esta Portaria prevê o envolvimento da sociedade em diferentes momentos participativos de realização da AIR<sup>1</sup>, objetivando, sobretudo, ampliar o debate público sobre propostas de intervenção regulatória que protejam e promovam a saúde da população sem, contudo, onerar desnecessariamente e excessivamente o cidadão e o setor produtivo do país.

Um dos momentos participativos previstos para a conclusão de AIR conduzida pela Anvisa é a submissão do relatório preliminar às contribuições da sociedade. Este documento apresenta conteúdo narrativo, principalmente, sobre a análise e definição do problema regulatório a ser enfrentado, justificativa de atuação da Anvisa, experiências internacionais de enfrentamento de problema similar e identificação e comparação das intervenções regulatórias para tratamento do problema. Conforme, previsto na Portaria, a Tomada Pública de Subsídios (TPS) é o mecanismo de participação social, aberto ao público, realizado em prazo definido, para coletar dados e informações, por escrito, sobre o relatório preliminar de AIR<sup>1</sup>.

Se por um lado, a Portaria consolida o entendimento da Agência de que a participação da sociedade na AIR é fundamental para promover a transparência e assegurar à sociedade um espaço para exercício da cidadania, por outro, apresenta um importante desafio para a Anvisa: como tornar viável e responsiva a análise das contribuições discursivas oriundas da sociedade recebidas pela Anvisa nos seus processos regulatórios participativos?

O uso de *software* para análise de textos advindos, por exemplo, de perguntas abertas contidas em formulários eletrônicos tem sido recomendado na literatura científica por diversas razões. Algumas delas são: i) necessidade de lidar com grandes volumes de respostas textuais; ii) viabilizar o aprimoramento das análises, sobretudo, tornando possível a integração dos domínios quantitativos e qualitativos na análise; e iii) trazer maior objetividade e avanços às interpretações dos dados pelo pesquisador<sup>5,6</sup>.

Não foram identificados estudos na literatura que realizaram análise textual de contribuições da sociedade sobre temas ligados à regulação governamental no Brasil com o uso do *Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* (IRaMuTeQ), um *software* muito utilizado na análise de dados qualitativos em programas de pós-graduação nas áreas da saúde<sup>7</sup>.

Este estudo teve como objetivo demonstrar a utilização do *software* IRaMuTeQ, como ferramenta de apoio à análise de contribuições sociais oriundas de processo regulatório conduzido pela Anvisa. A escolha pelo *software* IRaMuTeQ deveu-se a sua relevância em processos mais rigorosos e sistemáticos de análise textual e que ainda é pouco difundido no campo da regulação sanitária no Brasil, bem como a facilidade no manejo das suas funcionalidades<sup>8,9</sup>.

## MÉTODO

### Desenho do estudo

Estudo de caso que utilizou o *corpus* textual composto pelas respostas da sociedade feitas à pergunta “O problema principal



apresentado foi identificado corretamente?”, do relatório preliminar da AIR da Rotulagem Nutricional de Alimentos, que foi submetido à TPS pela Anvisa. O problema regulatório definido nessa AIR foi: “Dificuldade de uso da rotulagem nutricional pelos consumidores brasileiros”<sup>10</sup>. Com a TPS, a Anvisa objetivou coletar dados, informações ou evidências sobre o relatório preliminar de AIR, a fim de auxiliar na tomada de decisão.

As justificativas para a escolha da pergunta foram: a análise e a definição do problema regulatório são os aspectos mais importante de uma AIR<sup>11</sup>; e o uso de uma única pergunta facilitaria a análise estatística pelo software IRaMuTeQ<sup>12</sup>.

A TPS do relatório preliminar da AIR da Rotulagem Nutricional de Alimentos teve como objetivos apresentar à sociedade propostas de soluções regulatórias, que facilitassem a compreensão das principais propriedades nutricionais dos alimentos comercializados no país pelos consumidores e reduzissem as situações que geram engano quanto a sua composição. Além disso, a TPS visou coletar contribuições que ajudassem no aprimoramento de tais soluções e/ou na proposição de outras, bem como que respaldassem outros pilares importantes da AIR, como o problema regulatório<sup>13</sup>.

### O software IRaMuTeQ

O IRaMuTeQ, versão 0.7 alfa 2, é um software gratuito, ancorado no ambiente estatístico do software R e na linguagem Python, que oferece diferentes tipos de análise estatística de dados qualitativos, os quais são processados e transformados em *corpus* textual<sup>5,12</sup>. Os resultados deste estudo foram baseados em ferramentas oferecidas pelo IRaMuTeQ para viabilizar a interpretação dos dados mais amplamente descritos pelas palavras do *corpus* textual.

### Coleta das informações

A coleta das contribuições da sociedade ocorreu por meio de formulário eletrônico disponibilizado no portal da Anvisa por 60 dias, compreendendo o período de 25 de maio a 24 de julho de 2018. O formulário foi composto por 22 perguntas distribuídas em quatro seções: 1) Percepção da sociedade (duas perguntas); 2) Análise do Impacto Regulatório (nove perguntas); 3) Design gráfico e comunicação (dez perguntas); e 4) Prazo de adequação (uma pergunta). A pergunta “O problema principal apresentado foi identificado corretamente?”, utilizada na presente análise, constava na 2ª seção do formulário.

### Modelagem das informações

As contribuições da sociedade foram organizadas em planilha de Excel® e, em seguida, a preparação para serem processadas no IRaMuTeQ foi realizada em quatro etapas: 1) a inspeção visual e o uso de funcionalidades do Excel® para identificação e exclusão das contribuições repetidas e de elementos textuais que não agregavam qualquer tipo de valor, como conectores e caracteres especiais. Um único texto, representando as contribuições repetidas, foi mantido para fins de montagem do *corpus* textual a ser estudado. A utilização de texto único, neste caso, visou evitar vies estatístico e assegurar o correto resultado da análise. Além disso,

as indicações de citações bibliográficas foram desconsideradas para a formação do *corpus* textual; 2) as contribuições apresentadas em inglês passaram por tradução eletrônica, via Google Tradutor, e ajustes feitos por um dos pesquisadores, possibilitando sua incorporação na análise; 3) as contribuições textuais com menos de 40 palavras foram agrupadas em textos maiores, de modo a permitir que o software realizasse a análise estatística dos dados; e 4) o resultado de todo esse processo de tratamento das contribuições formou o *corpus* textual final para análise no IRaMuTeQ.

O *corpus* textual final incluiu as formas ativas, ou seja, substantivos, verbos (suplementares ou não), adjetivos (suplementares ou não), advérbios (suplementares ou não) e formas não reconhecidas e as formas suplementares que, neste caso, incluíram apenas substantivos suplementares e adjetivos numéricos. Neste estudo, as formas ativas foram denominadas de palavras ativas.

### Análise das informações

O *corpus* textual final foi dividido em dois grupos sociais de interesse - 12 *stakeholders* e público em geral - na tentativa de demonstrar a potencialidade de discriminação das análises fornecidas pelo IRaMuTeQ, no que tange a identificação de argumentos com características similares e distintas, segundo os posicionamentos dos referidos grupos relativos à pergunta estudada.

As contribuições de 12 *stakeholders* foram agrupadas em um *corpus* textual, representando o primeiro grupo social de interesse. Os 12 *stakeholders* selecionados foram: comunidade acadêmica nacional (participante A), comunidade acadêmica internacional (participante B), entidade de defesa do consumidor (participante C), setor produtivo de alimentos em nível nacional (participantes D, E e F), setor produtivo de alimentos em nível internacional (participante G), órgão internacional de saúde (participantes H e I), entidade de profissional da saúde (participantes J e L) e órgão do governo federal (participante M). Nessa seleção buscou-se priorizar segmentos da sociedade que tivessem potencial de apresentarem posições concordantes e antagônicas sobre a pergunta estudada. Este grupo social funcionou como um controle do outro grupo estudado.

O restante das contribuições foi agrupado em um segundo *corpus* textual, o qual foi considerado representar a visão do público em geral.

Os dois *corpora* textuais foram submetidos a três análises fornecidas pelo IRaMuTeQ<sup>14</sup> e descritas a seguir:

- i. Análise lexicográfica: envolve a medição da frequência e distribuição das palavras no *corpus* textual. A quantificação das palavras ativas e suplementares nos *corpora* textuais foi realizada após o processo de lematização, que consiste na substituição de uma palavra por seu termo raiz (por exemplo: “problematização” por “problema”). Esse processo elimina o final flexionado da palavra para normalizar o texto, diminuindo sua complexidade sem comprometer a precisão<sup>14</sup>;
- ii. Classificação hierárquica descendente (CHD): conforme descrita por Mendes et al.<sup>14</sup>, trata-se de um tipo de análise de conglomerado que categoriza as palavras ativas em classes



lexicais. A análise considera a frequência e as posições das palavras ativas que estão no texto usando os dados das tabelas de contingência das palavras. O algoritmo começa a assumir que todas as palavras ativas pertencem inicialmente à mesma classe e a divide sequencialmente, maximizando a inércia interclasses e minimizando a inércia intraclasses. A inércia é definida como uma medida de variância entre os perfis individuais entorno do perfil médio. Portanto, quanto maiores as diferenças entre as palavras, maior será a inércia entre elas. Esse processo iterativo é interrompido quando uma nova divisão de palavras não melhora a inércia interclasses. Nesta perspectiva, o número final de classes é *a priori* desconhecido. Dendrogramas são elaborados com a finalidade de ilustrar a divisão entre as classes. Para medir a força de associação entre as palavras ativas e sua respectiva classe foi utilizado o qui-quadrado de Pearson. Quanto maior o qui-quadrado de Pearson, mais provável é a hipótese de dependência entre palavra ativa e classe. No entanto, este teste estatístico não mostra como essas duas variáveis (palavra ativa e classe) estão relacionadas, por isso também foi realizada uma análise fatorial de correspondência; e

- iii. Análise fatorial de correspondência (AFC): permite, por meio de gráficos, visualizar a proximidade das palavras e das classes oriundas da CHD; não se trata de contar palavras, mas de relações entre palavras. AAFC foi interpretada em termos de oposição entre os eixos X e Y<sup>14</sup>.

A robustez das análises do IRaMuTeQ foi avaliada considerando: (i) o aproveitamento mínimo de 70% dos segmentos de texto na análise lexicográfica; (ii) valor do qui-quadrado de Pearson acima de 3,85 nas análises de CHD, representando uma separação entre as classes satisfatória; e (iii) soma dos fatores dos eixos dos gráficos da AFC próxima de 100%. Os valores foram baseados no estudo feito por Mendes et al.<sup>14</sup>.

As classes lexicais foram nomeadas subjetivamente por consenso entre os autores, com base na composição das palavras mais predominantes na tentativa de representar as ideias de cada classe<sup>14,15</sup>.

### Considerações éticas

As contribuições da sociedade analisadas foram obtidas no contexto de ações de vigilância sanitária, situação na qual é dispensada a apreciação por Comitê de Ética em Pesquisa. Foram observados aspectos éticos da Resolução do Conselho Nacional de Saúde (CNS) nº 510, de 7 de abril de 2016<sup>16</sup>. Os resultados apresentados garantem o sigilo individual dos participantes.

## RESULTADOS

Foram apresentadas 564 contribuições à pergunta “O problema principal apresentado foi identificado corretamente?”, cujos segmentos da sociedade que se identificaram como: setor produtivo (n = 196; 34,7%), consumidores (n = 154; 27,3%), profissionais da saúde (n = 68; 12,0%) e instituições de ensino (n = 68; 12,0%) foram os principais atores registrados.

Foram identificadas 221 (39,2%) contribuições repetidas, sendo que 177 (80,0%) foram advindas de uma única instituição pertencente ao setor produtivo, seguida de 29 (13,1%) de duas universidades estrangeiras e 15 (6,8%) de uma instituição representante do consumidor. Apenas uma resposta de cada um desses participantes foi somada às outras contribuições únicas (n = 343), resultando em um *corpus* textual final composto por 346 contribuições.

### Análise lexical dos 12 stakeholders

A análise lexicográfica do *corpus* textual formado pelas 12 contribuições (12 textos) produziu 3.244 ocorrências (palavras e formas); destas 357 (11,0%) são palavras que aparecerem uma única vez no *corpus* textual. Após a lematização do *corpus* textual, obteve-se 678 palavras ativas e nove suplementares. Das palavras ativas, as cinco mais frequentes foram: ser (n = 66 vezes), consumidor (n = 35), problema (n = 31), informação (n = 29) e nutricional (n = 28 vezes). Obtiveram-se 93 segmentos de texto, dos quais 66 (71,0%) foram aproveitados pelo IRaMuTeQ.

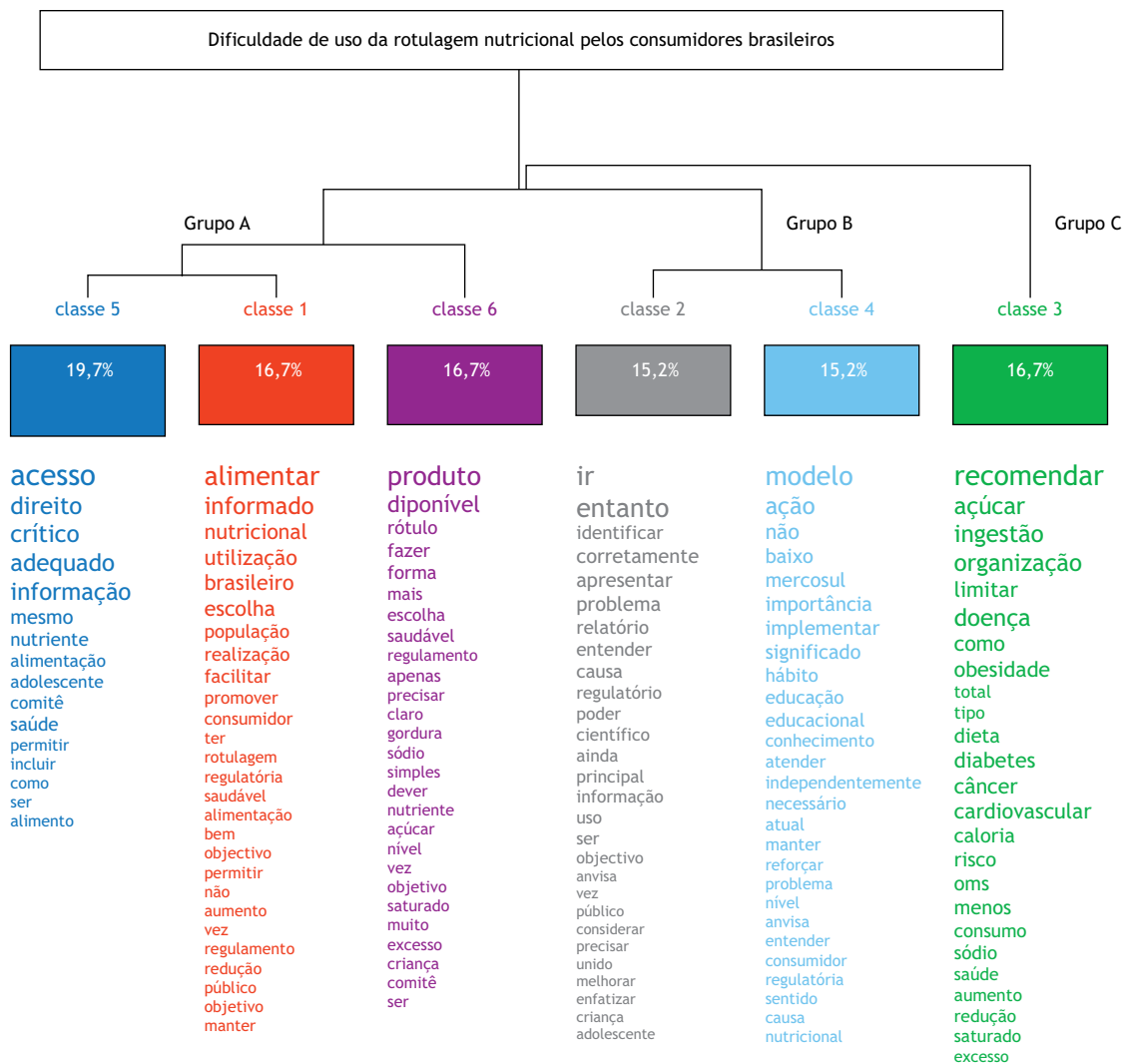
A análise de CHD das palavras ativas produziu cinco classes lexicais divididas em três grandes grupos. O Grupo A formado pelas classes 1, 5 e 6, o Grupo B representado pelas classes 2 e 4 e o Grupo C formado isoladamente pela classe 3. Este resultado está demonstrado no dendrograma, o qual contém, também, a lista das principais palavras ativas por classe (Figura 1).

Conforme observado na Figura 1, o Grupo A foi o mais representativo, totalizando 53,1% dos segmentos de texto. Já a classe 5, compreendendo 13 (19,7%) segmentos de texto, resultou na mais representativa, seguida das classes 1, 3 e 6, todas com 11 (16,7%) segmentos de texto. Em todas as palavras de cada classe foram obtidos qui-quadrados de Pearson superiores a 3,86.

As classes lexicais receberam as seguintes nomeações: a) classe 1: uso da rotulagem pelos consumidores brasileiros em suas escolhas alimentares; b) classe 2: debate sobre a identificação do problema regulatório; c) classe 3: relação entre hábitos alimentares e risco de aparecimento de doenças; d) classe 4: importância da educação nutricional e da harmonização com o Mercado Comum do Sul (Mercosul); e) classe 5: direito e acesso à informação; e f) classe 6: papel dos rótulos claros para promover escolhas saudáveis entre os produtos disponíveis.

A AFC resultou em cinco fatores que explicam 32,3%, 22,1%, 17,7%, 14,6% e 13,3% do modelo. Os dois fatores com percentuais maiores explicam um acumulado de 54,4% do modelo, os quais estão representados nos eixos X e Y do plano cartesiano (Figura 2).

No eixo X, representando 32,3% de distribuição no *corpus* textual, a classe 3 (eixo X negativo) é claramente separada da classe 4 (eixo X positivo). Há maior predomínio de posicionamento central das classes 1, 5 e 6 nesse eixo. No eixo Y, representando 22,1% de distribuição no *corpus* textual, as classes 3 e 4, embora estejam do lado positivo do eixo Y, estão claramente posicionadas em quadrantes diferentes. Já as classes 5 e 6 estão posicionadas completamente no eixo Y negativo. Não há posicionamento de palavras no cruzamento dos eixos X e Y.



\* As palavras mais no topo da lista e maior tamanho possuem mais influência na classe.  
Fonte: Elaborada pelos autores, 2019.

Figura 1. Dendrograma das seis classes lexicais obtidas a partir da classificação hierárquica descendente das palavras ativas advindas das contribuições dos 12 stakeholders (Total de segmentos de texto = 66)\*.

A combinação dos dois eixos (X e Y), que oferece uma visão bidimensional, distingue o *corpus* textual em três mundos lexicais. O primeiro, um mundo de eixos X e Y positivos, no quadrante superior direito, mostrando a classe 4, que corresponde ao tema “importância da educação nutricional e da harmonização com o Mercosul”; o segundo, um mundo de coordenadas com eixo X negativo e eixo Y positivo no quadrante superior esquerdo, mostrando a classe 3, correspondendo ao tema “relação entre hábitos alimentares e risco de aparecimento de doenças”; e finalmente, um mundo lexical com posicionamento central em relação ao eixo X e negativo para o eixo Y. Este mundo lexical está concentrado principalmente no quadrante inferior, mostrando as demais classes.

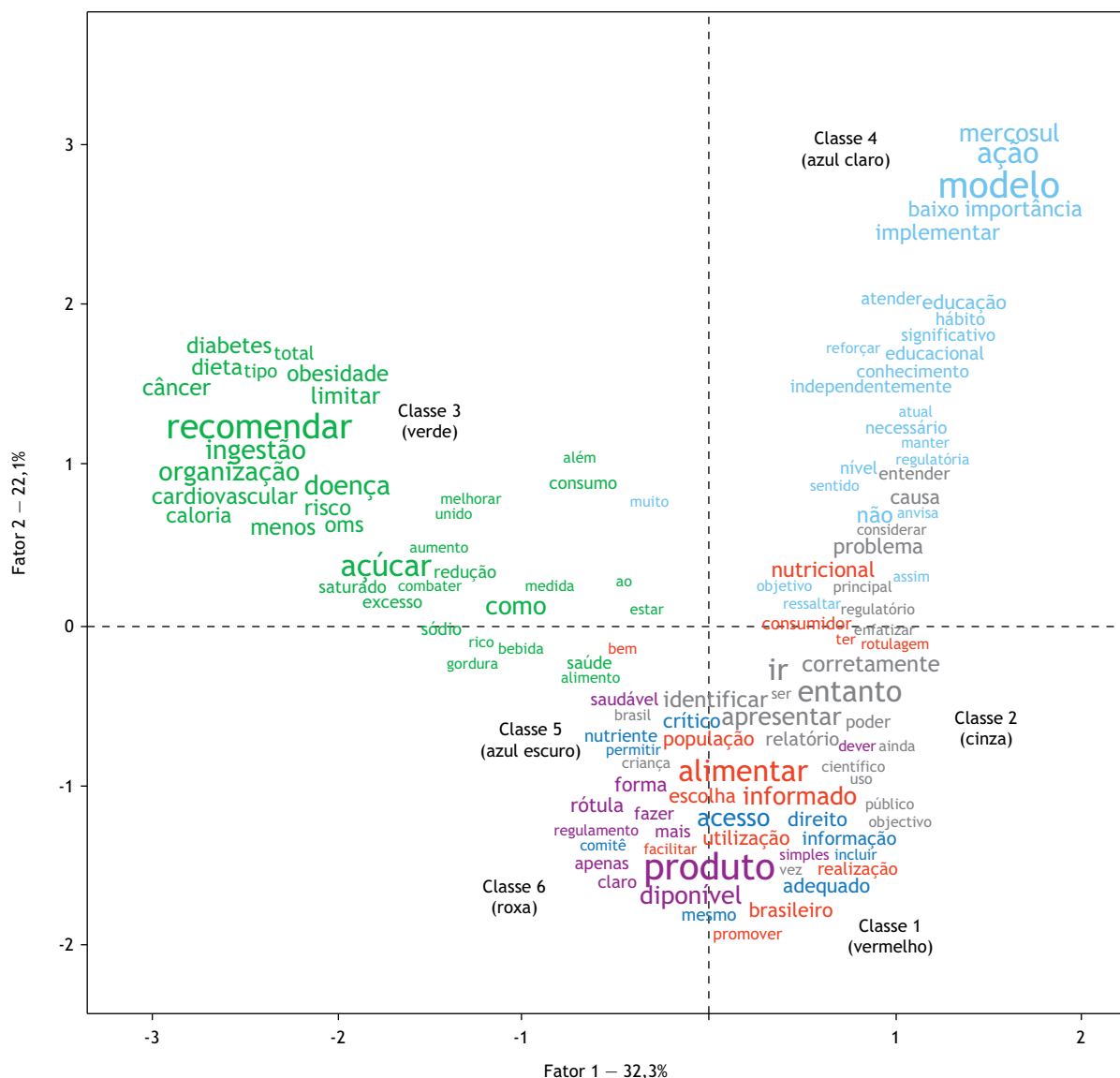
A distribuição gráfica dos argumentos apresentados pelos stakeholders indica um distanciamento entre aqueles apresentados pelos participantes B (comunidade acadêmica internacional) e C (entidade de defesa do consumidor) (classe 3) e pelos participantes G (setor produtivo de alimentos em nível internacional)

e D (setor produtivo de alimentos em nível nacional) (classe 4), estando ambos os respondentes afastados dos argumentos apresentados pelos demais stakeholders, representados pela comunidade acadêmica nacional, entidade de profissional de saúde e órgão do governo federal (Figura 3).

#### Análise lexical do público em geral

Além de representantes dos segmentos da sociedade mencionados acima, as contribuições de outros atores formaram o *corpus* textual deste grupo social. Foram eles: sociedade civil (n = 21), organismos internacionais (n = 12), especialista em comunicação (n = 8), profissionais do Sistema Nacional de Vigilância Sanitária (n = 7) e consultorias (n = 4). Um total de 25 participantes identificaram-se como Outros. O *corpus* textual do público em geral não contou com contribuições advindas de representantes de órgãos do governo federal, uma vez que houve apenas uma contribuição registrada, a qual fez parte do grupo social anterior.





Fonte: Elaborada pelos autores, 2019.

**Figura 2.** Análise fatorial de correspondência das palavras ativas mais frequentes em cada uma das classes lexicais obtidas na classificação hierárquica descendente das contribuições dos 12 stakeholders.

Um total de 218 contribuições com menos de 40 palavras foram agrupadas, formando um *corpus* textual com 116 textos analisados pelo *software*.

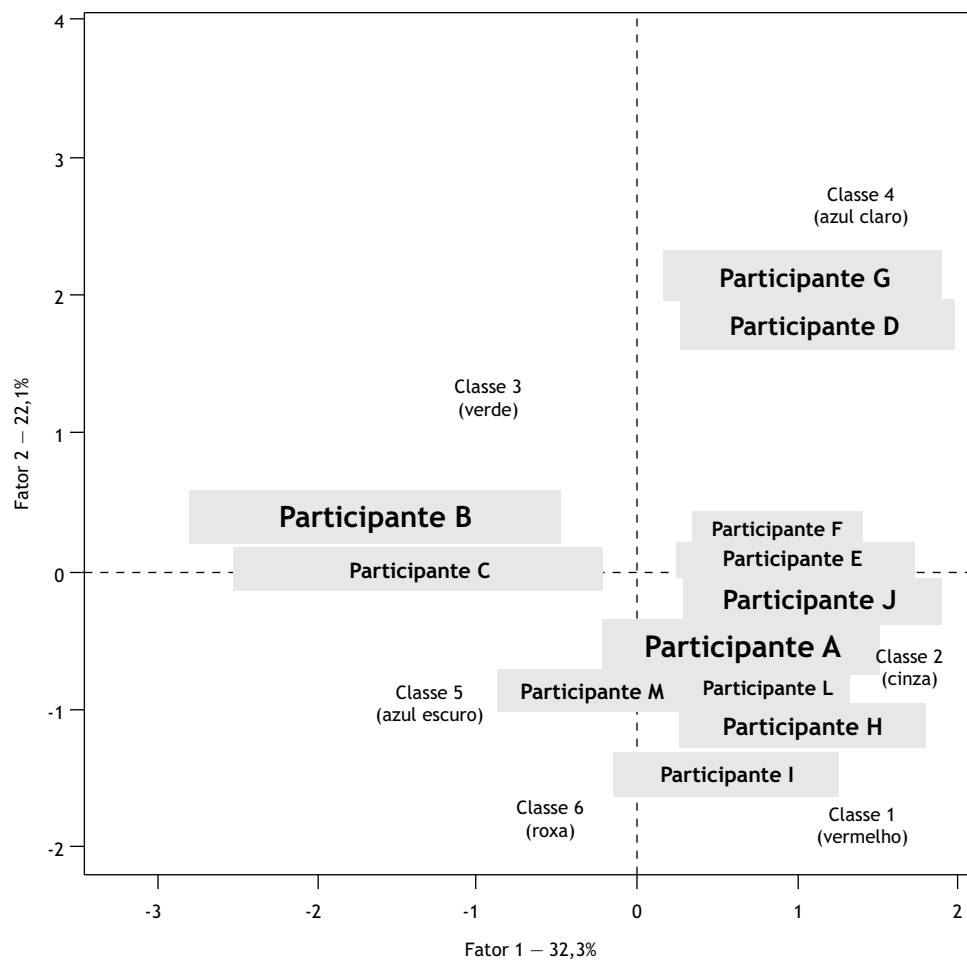
A análise lexicográfica desse *corpus* textual gerou 15.084 ocorrências (palavras e formas); destas, 1.619 (10,7%) são hápax, ou seja, ocorrências que aparecem apenas uma vez no texto. Após lematização do *corpus* textual, obteve-se 1.846 palavras ativas e 25 suplementares. Das palavras ativas, as cinco mais frequentes foram: ser (n = 431 vezes), não (n = 233 vezes), nutricional (n = 204 vezes), consumidor (n = 184 vezes) e informação (n = 184 vezes). Foram obtidos 430 segmentos de texto, dos quais 390 (90,7%) foram aproveitados pelo IRaMuTeQ.

A análise de CHD das palavras ativas resultou em cinco classes lexicais divididas em três grandes grupos. Os Grupo A e C são

formados, respectivamente, pelas classes 5 e 4, enquanto o Grupo B foi composto pelas classes 1, 2 e 3. A Figura 4 apresenta este resultado em formato de dendrograma, o qual incluiu a lista das principais palavras ativas por classe.

Conforme pode ser observado na Figura 4, o Grupo B foi o mais representativo, totalizando 68,0% dos segmentos de texto. Já a classe 3, compreendendo 124 (31,8%) segmentos de texto, resultou na mais representativa, seguida da classe 2 com 88 (22,5%) e classe 5 com 75 (19,2%). Em todas as palavras de cada classe foram obtidos qui-quadrados de Pearson superiores a 3,91.

As classes lexicais receberam as seguintes nomeações: a) classe 1: potencial dos rótulos confundirem e apresentarem informações falsas; b) classe 2: importância da educação nutricional; c) classe 3: debate sobre a identificação do problema regulatório;



Fonte: Elaborada pelos autores, 2019.

Figura 3. Distância lexical entre os argumentos apresentados pelos stakeholders selecionados.

d) classe 4: necessidade de intervenção da Anvisa e de harmonização com o Mercosul; e e) classe 5: relação entre hábitos alimentares e risco de aparecimento de doenças.

A AFC resultou em quatro fatores que explicam 33,8%, 29,7%, 19,3% e 17,2% do modelo. Os dois fatores com percentuais maiores explicam um acumulado de 63,5% do modelo, os quais estão representados nos eixos X e Y do plano cartesiano (Figura 5).

No eixo X, representando 33,8% de distribuição no *corpus* textual, a classe 5 está completamente posicionada no lado positivo, enquanto a classe 2, na sua maior parte, está posicionada no lado negativo do eixo. No eixo Y, representando 29,7% de distribuição no *corpus* textual, a classe 4 está completamente posicionada no lado negativo, enquanto as classes 1, 2 e 5 estão, na sua maior parte, distribuídas no lado positivo do eixo. Há várias palavras posicionadas no cruzamento dos eixos X e Y (Figura 5).

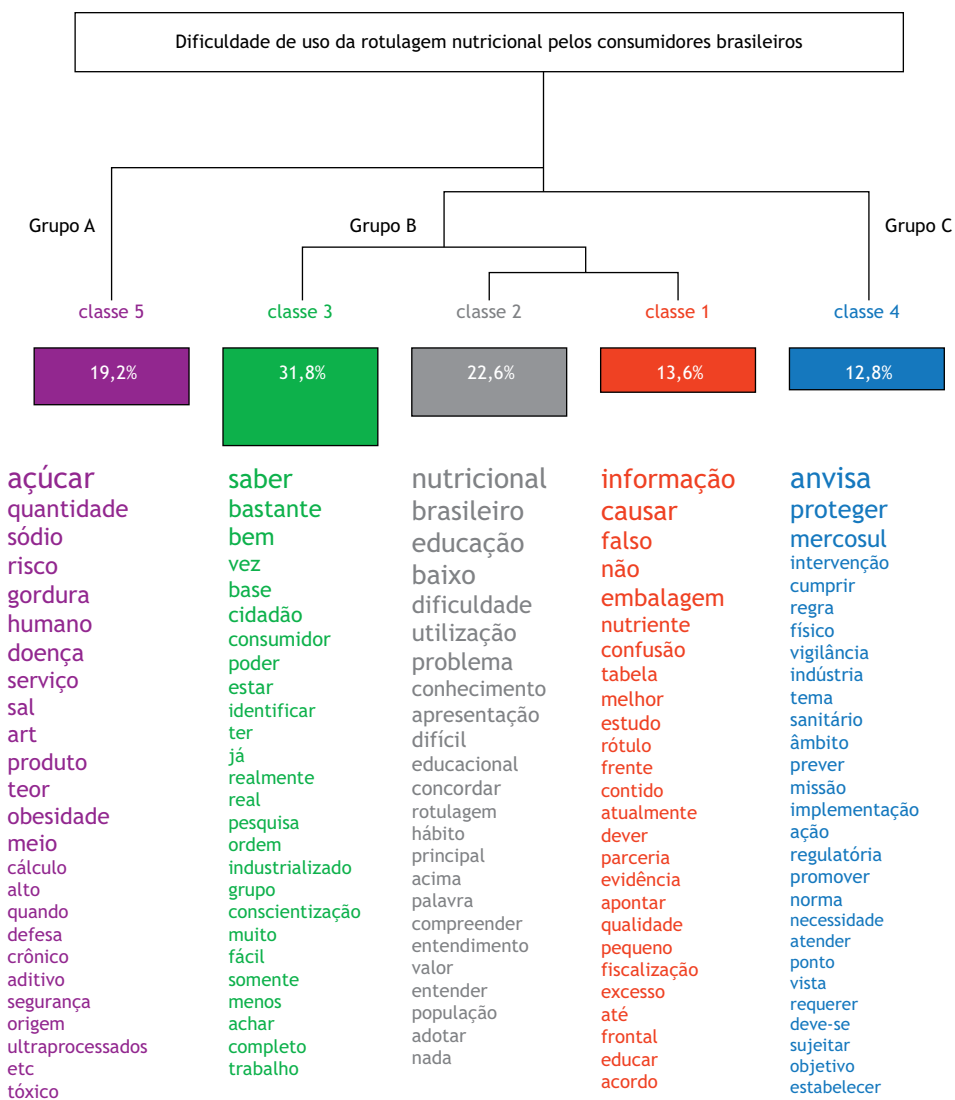
A combinação dos dois eixos (X e Y), que oferece uma visão bidimensional, distingue o *corpus* textual em três mundos lexicais (Figura 5). O primeiro, um mundo de eixos X e Y positivos, no quadrante superior direito, mostrando predominantemente a classe 5, que corresponde ao tema “relação entre hábitos

alimentares e risco de aparecimento de doenças”; o segundo, um mundo com eixo X negativo e eixo Y positivo no quadrante superior esquerdo, mostrando predominantemente a classe 2, correspondendo ao tema “importância da educação nutricional”; e finalmente, um mundo lexical com posicionamento central no eixo X e negativo no eixo Y distribuído pelos quadrantes inferiores, mostrando predominantemente a classe 4 (necessidade de intervenção da Anvisa e de harmonização com o Mercosul).

A classe 5 é formada por argumentos oriundos da sociedade civil, instituições de ensino, Sistema Nacional de Vigilância Sanitária e instituição governamental. A classe 4 os argumentos são formados por organismos internacionais, pessoas/entidades que se identificaram como “Outros” e, principalmente, pelo setor produtivo. As classes 1 e 2 são formadas por argumentos advindos de consultorias e especialistas em comunicação, respectivamente. A classe 3 é composta pelas contribuições dos consumidores e profissionais da saúde (Figura 6).

## DISCUSSÃO

As análises lexicais, utilizando o IRaMuTeQ, permitiram discriminar as tendências argumentativas predominantes sobre a



\* As palavras mais no topo da lista e maior tamanho possuem mais influência na classe.  
Fonte: Elaborada pelos autores, 2019.

Figura 4. Dendrograma das seis classes lexicais obtidas a partir da classificação hierárquica descendente das palavras ativas advindas das contribuições do público em geral (Total de segmentos de texto = 390)\*.

pergunta “O problema principal apresentado foi identificado corretamente?”, da AIR da Rotulagem Nutricional de Alimentos, oriundos dos dois grupos sociais estudados, bem como analisar as inter-relações entre elas. Todas as análises lexicais foram consideradas robustas, pois cumpriram com os três requisitos metodológicos estabelecidos previamente.

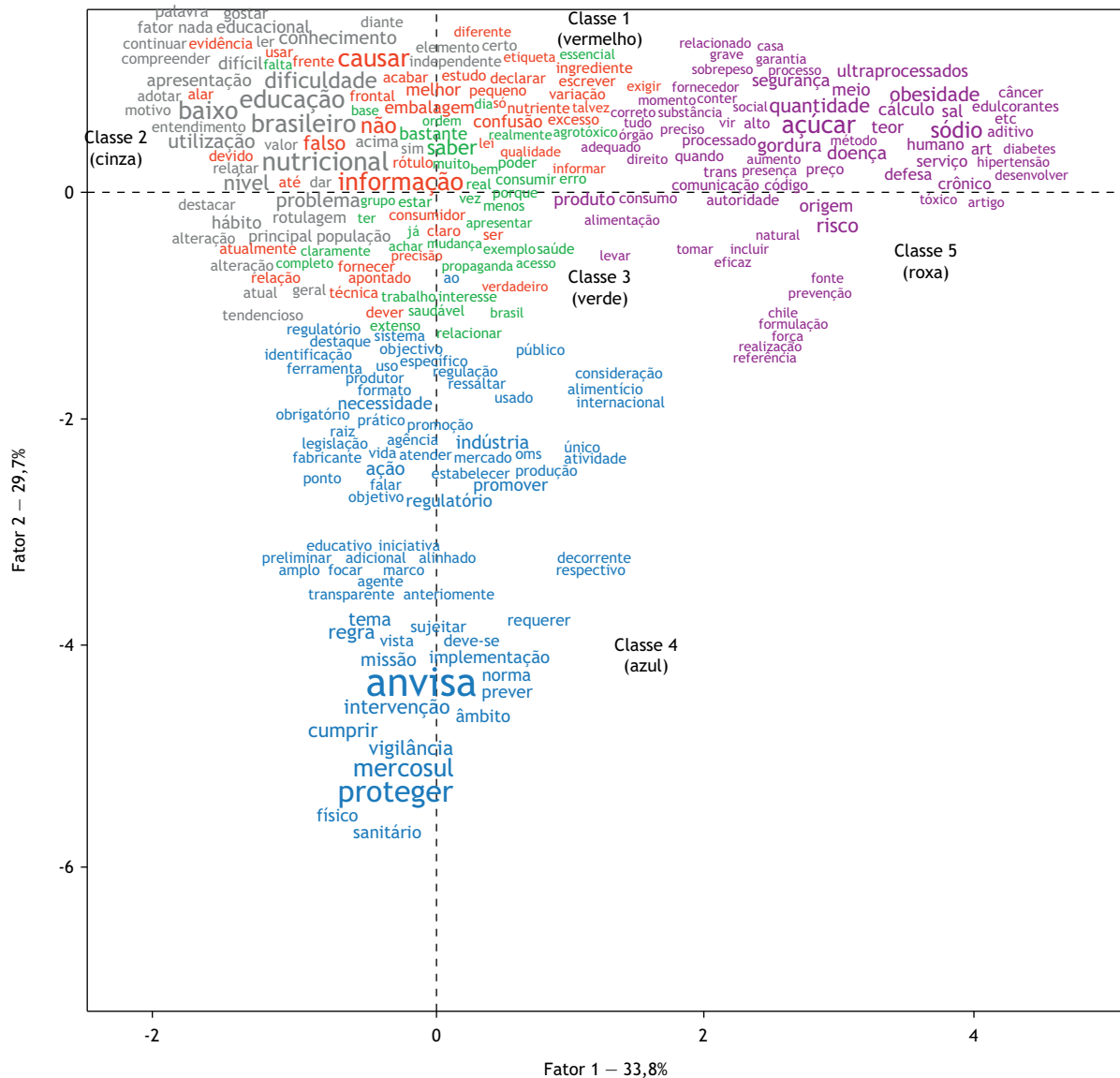
As análises de CHD dos *corpora* textuais dos grupos sociais foram equivalentes quanto à divisão em grandes grupos, à quantidade de classes lexicais e aos argumentos/palavras predominantes que resultaram na nomeação das classes, indicando um elevado poder de discriminação do software IRaMuTeQ em relação a representação dos conjuntos de ideias e seus respectivos atores sociais.

A diferença no número de classes ficou por conta da classe 5 (direito e acesso à informação) cuja contribuição provém do representante do órgão público federal e que constava apenas

no *corpus* textual dos 12 *stakeholders*. Essa semelhança também ocorreu tanto em relação às palavras predominantes (quatro de cinco - ser, consumidor, informação e nutricional) na análise lexicográfica quanto nas análises provenientes da AFC, em que tange a divisão do conjunto de palavras em três mundos lexicais, envolvendo classes com argumentações parecidas.

De maneira geral, para ambos os grupos sociais, as classes lexicais obtidas têm bastante relação com o problema definido na AIR “Dificuldade de uso da rotulagem nutricional pelos consumidores brasileiros”<sup>5</sup>. Por exemplo: as contribuições representadas pelas classes 2 (12 *stakeholders*) e 3 (público em geral), que receberam a mesma denominação, retratam o debate sobre a identificação do problema regulatório enquanto que as contribuições da classe 1 do público em geral apontam para o potencial dos rótulos confundirem e apresentarem informações falsas





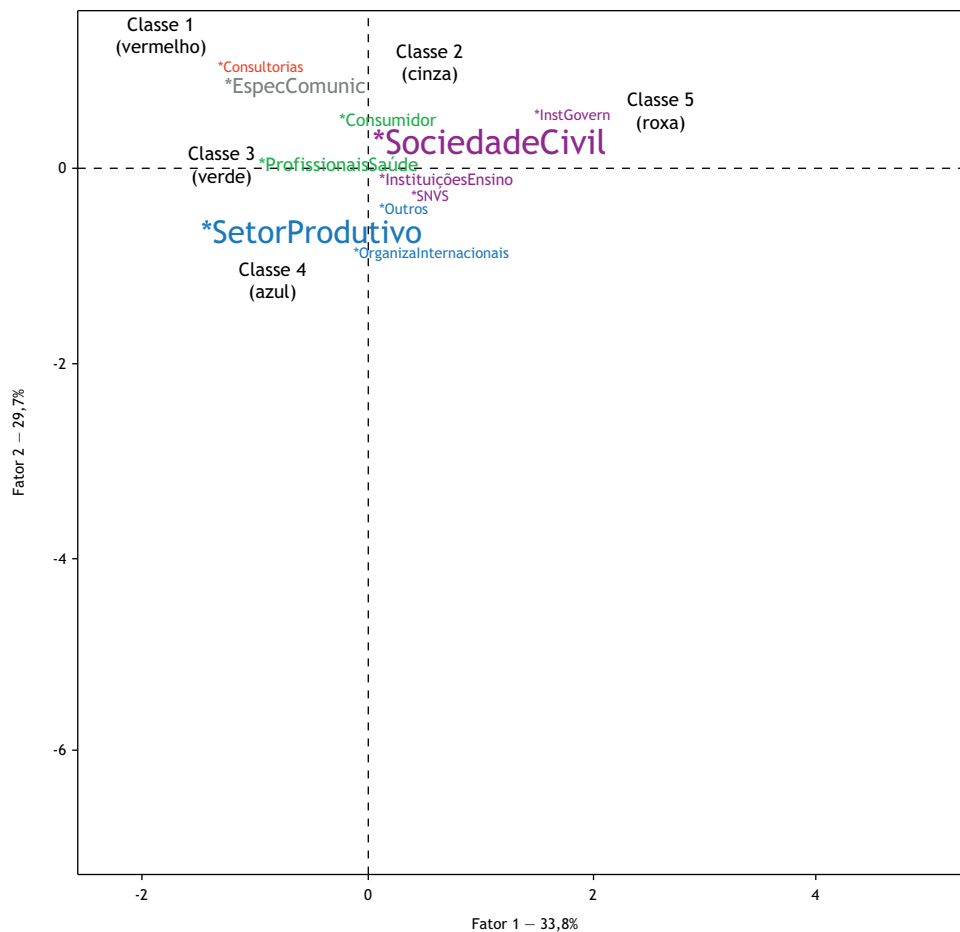
Fonte: Elaborada pelos autores, 2019.

Figura 5. Análise fatorial de correspondência das palavras ativas mais frequentes em cada uma das classes lexicais obtidas na classificação hierárquica descendente das contribuições do público em geral.

nas escolhas dos alimentos. Verificou-se, ainda, uma nítida relação de palavras presentes nos *corpora* textuais com o problema regulatório. Três das palavras mais frequentes nos dois *corpora* textuais “consumidor” e “informação”, ambas morfemas substantivos, e “nutricional”, morfema adjetivo, exemplificam essa relação. O conjunto das cinco principais palavras em cada *corpus* textual sugere a seguinte frase como discurso simbólico representativo das contribuições sociais: “A informação nutricional [nos alimentos] é (não é) um problema para o consumidor”.

O estudo das repetições ou da frequência das palavras em um texto permite identificar mundos lexicais que podem estar relacionados a tendências ideológicas semelhantes, conflitos, rupturas, reaproximações ou oposições argumentativas sobre determinado tema<sup>15</sup>.

Neste estudo com base na riqueza das palavras observadas em cada classe, identificaram-se mundos lexicais relacionados com reaproximações de discursos em relação ao problema regulatório. Por exemplo: as classes 1, 2, 5 e 6 aparecem entrelaçadas, formando um único mundo lexical no *corpus* textual dos 12 *stakeholders*, ou seja, há palavras comuns nessas classes, indicando maior afinidade contextual. Tais classes representam as contribuições de sete segmentos da sociedade: dois órgãos internacionais de saúde, um da comunidade acadêmica nacional, dois das entidades de profissionais da saúde, um do setor produtivo de alimentos em nível nacional e o representante do órgão do governo federal. Em relação ao *corpus* textual do público em geral, esse entrelaçamento ocorreu com as classes 1, 2 e 3, as quais são representadas pelos seguintes atores, respectivamente: consultorias, especialistas em comunicação e consumidores e profissionais da saúde.



Fonte: Elaborada pelos autores, 2019.

Figura 6. Distância lexical entre os argumentos apresentados pelo público em geral.

A identificação de oposições argumentativas visualizadas entre as classes 3 e 4, no grupo dos 12 *stakeholders* e entre as classes 4 e 5 no grupo do público em geral, também, foram preservadas nas análises do IRaMuTeQ. As classes 3 e 4 representam as contribuições da comunidade acadêmica internacional e entidade de defesa do consumidor e setor produtivo de alimentos em nível nacional e internacional, respectivamente. Já as classes 4 e 5 são contribuições advindas, principalmente, do setor produtivo e da sociedade civil, respectivamente.

Foi observada uma clara diferença na distribuição das palavras ao longo do cruzamento dos eixos X e Y entre os *corpora* textuais estudados. No caso do *corpus* textual do público em geral, diferentemente do que ocorre para o *corpus* textual dos 12 *stakeholders*, observou-se várias palavras que estão no cruzamento dos eixos X e Y, indicando que são comuns em todas as classes. Uma explicação para esse acontecimento é a presença de maior diversidade dos participantes no grupo do público em geral.

Vários estudos têm utilizado o *software* IRaMuTeQ no processamento de textos obtidos por diferentes formas, a exemplo de questionários/formulários com questões abertas<sup>17</sup>. No entanto, não foram identificados estudos na literatura que realizaram análise lexical de contribuições da sociedade sobre temas ligados à regulação sanitária com o uso do *software* IRaMuTeQ. Este

estudo, ao que parece, é o primeiro a explorar esse *software* como ferramenta viável e responsiva para possibilitar a análise das contribuições discursivas oriundas dos processos de participação da sociedade estabelecidos pela Anvisa.

Algumas das principais contribuições do IRaMuTeQ para análise textual discursiva incluem agilidade nas análises, reduzindo o gasto de energia do pesquisador na extração de dados passíveis de interpretação de um grande volume de informação e ampliação na oferta de novas possibilidades de interpretações e relações, que poderiam passar despercebidas no trabalho manual<sup>18</sup>. Outras três vantagens do *software* são: i) ajudar na redução de possíveis vieses na análise de consultas públicas por parte dos profissionais envolvidos com a tarefa; ii) promover maior rigor estatístico as análises qualitativas; e iii) possibilitar a realização de diferentes tipos de análises, inclusive multivariadas, com o apoio de gráficos<sup>7,9,19</sup>.

Três tipos de análises do *corpus* textual, de um total de cinco, executadas pelo *software* foram realizadas neste estudo. A análise de similitude e a nuvem de palavras, que compõem as análises lexicais do IRaMuTeQ, não fizeram parte do estudo, podendo, isso ser caracterizado como uma limitação. No entanto, em revisão de escopo que buscou identificar o uso do IRaMuTeQ nas pesquisas qualitativas na área da saúde no Brasil predominaram as pesquisas que utilizaram apenas um tipo de análise fornecida



pelo *software*, sendo que a CHD foi a mais utilizada<sup>7</sup>. Ademais, a análise de similitude e a nuvem de palavras apresentam uma contribuição menor para as análises de consultas públicas quando comparadas com as análises lexicográficas, CHD e AFC<sup>20</sup>. Ressalta-se que as análises disponibilizadas pelo IRaMuTeQ não excluem o uso de outras análises, como a análise de conteúdo, sendo, também, indispensável o repertório conceitual do pesquisador para a interpretação das informações<sup>20</sup>.

Outra possível limitação do estudo foi a não utilização de *software* editor de texto para revisões ortográficas e textuais na etapa de modelagem das informações, as quais ocorreram por meio de inspeção visual feita por um dos pesquisadores. Alguns autores recomendam o uso dessa ferramenta, principalmente, quando da existência de volumosos textos para serem preparados e submetidos ao IRaMuTeQ<sup>8</sup>. Neste estudo, a preparação do *corpus* textual para submeter as análises lexicais oferecidas pelo IRaMuTeQ constituiu-se na atividade que mais demandou tempo e atenção por parte de um dos pesquisadores.

Uma parte das contribuições da sociedade analisada pelo IRaMuTeQ não foi alocada nas classes lexicais, haja vista que o aproveitamento do *corpus* textual, em ambos os casos, não foi 100%. No entanto, os percentuais de aproveitamento foram considerados suficientes para a realização da CHD, uma vez que uma retenção mínima de 70% dos segmentos de textos é defendida por alguns autores para este tipo de análise<sup>14</sup>.

Um estudo que investigou o uso do IRaMuTeQ para análise de contribuições da sociedade a consulta pública sobre a incorporação

do medicamento Trastuzumab para o tratamento inicial do câncer de mama no sistema público de saúde no Brasil, realizada em 2012 pelo Comitê Nacional de Incorporação de Tecnologia em Saúde, concluiu que, apesar dos resultados promissores obtidos, são necessários mais estudos para a validação do *software* no uso em análises de contribuições públicas<sup>9</sup>.

Este estudo, além de constatar a funcionalidade e potencialidade do IRaMuTeQ, contribui, também, na validação de seu uso em análises de contribuições sociais do tipo discursivas, com grande volume textual de difícil operacionalização por parte dos envolvidos. Para Krug<sup>8</sup>, a validação corresponde à capacidade de um instrumento produzir os resultados adequados e precisos que comporão as conclusões corretas.

## CONCLUSÕES

A discriminação e a compreensão das relações, resultantes das análises lexicais do IRaMuTeQ, mostraram-se satisfatórias, uma vez que os resultados foram bastante semelhantes, indicando certa exatidão na comparação dos *corpora* textuais dos dois grupos estudados. Conforme indicam os resultados deste estudo, 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.

## REFERÊNCIAS

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](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. Florianópolis: 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-Llamos 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 N° 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 N° 8.080, de 19 de setembro de 1990, pela Lei N° 8.142, de 28 de dezembro de 1990, pelo decreto N° 5.839, de 11 de julho de 2006. *Diário Oficial União.* 7 abr 2016.
17. Gutierrez 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.

---

#### Contribuição dos Autores

Carvalho TS - Concepção, planejamento (desenho do estudo), análise, interpretação dos dados e redação do trabalho. Mota DM - Análise, interpretação dos dados e redação do trabalho. Saab F - Planejamento (desenho do estudo), análise, interpretação dos dados e redação do trabalho. Todos os autores aprovaram a versão final do trabalho.

#### Conflito de Interesse

Os autores informam não haver qualquer potencial conflito de interesse com pares e instituições, políticos ou financeiros deste estudo.



Esta publicação está sob a licença Creative Commons Atribuição 3.0 não Adaptada.

Para ver uma cópia desta licença, visite [http://creativecommons.org/licenses/by/3.0/deed.pt\\_BR](http://creativecommons.org/licenses/by/3.0/deed.pt_BR).