

A linguagem probabilística empregada por estudantes do quinto ano do ensino fundamental: identificando fenômenos ou experimentos aleatórios

The probability language used by fifth-year elementary school students: identifying random phenomena or experiments

Ailton Paulo de Oliveira Júnior¹
Fátima Aparecida Kian²
Maria Manuel da Silva Nascimento³

Resumo: Neste estudo demonstra-se o processo de apresentação da linguagem utilizada para indicar o significado atribuído por alunos de 10 e 11 anos referente à identificação do conceito de fenômenos ou experimentos aleatórios. Realizou-se análises textuais multivariadas por meio do *software* IRaMuTeQ, empregando a Classificação Hierárquica Descendente - CHD e a Análise de Similitude para identificar e elucidar a variedade de expressões escritas, símbolos e representações utilizadas no processo gradual de aprendizado e aquisição dos conceitos básicos em probabilidade. Os resultados revelam a prevalência de palavras e expressões verbais da linguagem comum, especialmente relacionadas ao significado intuitivo de probabilidade, mas também se observa uma linguagem mais formal, essencial para a compreensão dos conceitos relacionados à probabilidade.

Palavras-chave: Linguagem probabilística. Letramento probabilístico. Fenômenos ou experimentos aleatórios. Ensino Fundamental. Análise Textual Multivariada

Abstract: This study demonstrates the process of presenting the language used to indicate the meaning attributed by 10- and 11-year-old students regarding the identification of the concept of phenomena or random experiments. Multivariate textual analyzes were carried out using the IRaMuTeQ software, using the Descending Hierarchical Classification – DHC and Similarity Analysis to identify and elucidate the variety of written expressions, symbols and representations used in the gradual process of learning and acquiring basic concepts in probability. The results reveal the prevalence of words and verbal expressions from common language, especially related to the intuitive meaning of probability, but a more formal language is also observed, essential for understanding concepts related to probability.

Keywords: Probability language. Probability literacy. Random phenomena or experiments. Fundamental School. Multivariate Textual Analysis

1 Introdução

Acredita-se que a abordagem ou linguagem intuitiva da probabilidade representa um componente central e fundamental nas fases iniciais na formação de qualquer cidadão, pois utiliza-se no dia a dia termos de uso comum para abordar a incerteza e expressar quantificação e grau por meio de frases coloquiais, bem como crenças em relação a eventos incertos.

É essencial que sejam assimilados os conceitos probabilísticos desde os primeiros anos do ensino fundamental que se inicia de maneira não formal, introduzindo terminologia

¹ Universidade Federal do ABC • Santo André, SP — Brasil • ✉ ailton.junior@ufabc.edu.br • ORCID <https://orcid.org/0000-0002-2721-7192>

² Universidade Federal do ABC • Santo André, SP — Brasil • ✉ fatima.kian@ufabc.edu.br • ORCID <https://orcid.org/0000-0003-0105-7335>

³ Universidade de Trás-os-Montes e Alto Douro • Quinta de Prados, Vila Real — Portugal • ✉ manematik@gmail.com • ORCID <https://orcid.org/0000-0002-3913-4845>

associada às ideias de probabilidade por meio de atividades ou situações problemáticas centradas nas avaliações que os alunos fazem com base em suas próprias vivências.

Assim, parte-se de Batanero (2013) ao indicar que na formação probabilística deve-se ter compreensão adequada da probabilidade, sendo necessário que os alunos sejam capazes de diferenciar entre situações aleatórias e determinísticas, ou seja, apreciar algumas características básicas da aleatoriedade.

Ainda se tem apoio em Vásquez (2018) quando indica a necessidade de apresentar um conjunto de situações em que o aluno possa identificar situações do cotidiano que estejam associados a experimento ou fenômeno aleatório. Batanero e Godino (2002) dizem que, em probabilidade, chama-se de experimento tanto aqueles que podemos provocar, quanto os que podem ser observados no mundo real.

Morgado, Pitombeira, Carvalho e Fernandez (2004) reforçam que os fenômenos aleatórios acontecem constantemente em nossa vida diária e o que os diferencia de um experimento determinístico é que os primeiros são experimentos que, repetidos sob as mesmas condições, produzem resultados diferentes, já o segundo é um experimento que, quando repetido em condições semelhantes, conduz a resultados essencialmente idênticos.

Dessa forma, apresenta-se duas situações-problema para que os alunos possam refletir sobre o conceito de fenômeno aleatório, solicitando que o aluno escreva uma palavra ou algumas palavras sobre a possibilidade da situação proposta ocorrer. Portanto, o propósito global deste estudo foi elucidar qual é a linguagem empregada na identificação do conceito de fenômenos ou experimentos aleatórios por alunos do quinto do ensino fundamental de uma escola municipal em Barueri, São Paulo e buscar associá-la à linguagem probabilística delineada pela Base Nacional Comum Curricular - BNCC (Brasil, 2018) e que estejam relacionadas a situações de seu cotidiano.

2 Marco teórico

Quanto à linguagem relacionada à probabilidade, Vásquez (2018) argumenta que essa está intrinsicamente ligada à linguagem utilizada no cotidiano, destacando que os primeiros elementos linguísticos incorporados à língua desempenham um papel fundamental nas pesquisas relacionadas ao ensino de probabilidade, especialmente nos anos iniciais do ensino fundamental.

Vásquez (2014) argumenta que a conexão entre as expressões de uso comum e a linguagem probabilística faz dos primeiros elementos linguísticos presentes na fala cotidiana um componente crucial para os estudos probabilísticos. Um das possibilidades é explorar expressões ou termos específicos da probabilidade que fazem parte da linguagem cotidiana, estando presentes no contexto diário ou muito próximo a ele.

Além disso, embasados em Batanero (2015), por exemplo, o conceito de aleatoriedade não é simples, tendo ao longo da história, diferentes significados e associado a discussões filosóficas. Também pode-se encontrar definições diferentes do que é utilizado no dia a dia e na sala de aula, geralmente partindo do uso de diferentes termos ou expressões como "imprevisibilidade", "possibilidade de vários resultados", "incontrolável", etc..

Oliveira Júnior, Datori Barbosa, Souza e Cardoso (2019), consideram que a partir das questões (problemas) associados aos conceitos de experimento aleatório é possível estabelecer relações de equivalência entre as diferentes formas de apresentação de problemas probabilísticos, sendo relevante para melhorar o desempenho de alunos do Ensino Fundamental na resolução de problemas, uma vez que a progressiva apreensão de conceitos básicos da

probabilidade reduz o efeito de variáveis que tornam maior a complexidade da sucessão de habilidades necessárias para o domínio do conceito de aleatoriedade.

Segundo Bryant e Nunes (2012) a compreensão da probabilidade de resultados incertos desempenha um papel extremamente importante em nossas vidas. Todas as pessoas dependem disso para decidir sobre, por exemplo, o tratamento médico que devem seguir, o seguro de vida ou do carro segundo suas necessidades, etc. Essas e muitas outras decisões dependem do nosso conhecimento dos possíveis eventos que podem acontecer e da compreensão individual sobre a probabilidade desses eventos.

Complementando, Gal (2005) concebe o letramento probabilístico partindo de duas perspectivas, os elementos cognitivos e da disposição. Os elementos cognitivos têm relação com os cálculos de probabilidade, a linguagem utilizada para se comunicar probabilisticamente, diferentes contextos de uso da probabilidade e perguntas críticas sobre o conceito. Os da disposição estão ligados às crenças pessoais, atitudes, hábitos e a sentimentos pessoais em relação à incerteza e ao risco no contexto da probabilidade.

Reforçando um dos elementos cognitivos, a linguagem, segundo Gal (2005), essa consiste nos termos utilizados para comunicar sobre acaso, aleatoriedade, etc. Por exemplo, palavras tais como: possível, provável, improvável, evento certo, evento impossível, chance, probabilidade, acaso, aleatório, incerteza, pode acontecer, não pode acontecer, experimento, evento, espaço amostral, sorte, azar, a esmo etc., seriam termos utilizados para comunicar o que queremos descobrir ou analisar sobre a probabilidade.

Além disso, de acordo com Gal (2005), outro elemento cognitivo seria o contexto, ou seja, ser letrado probabilisticamente requer que o sujeito possa desenvolver elementos das grandes ideias, calcular probabilidades, linguagem probabilística, a partir de um conhecimento de mundo ou de contexto, considerando que as pessoas apresentam expectativas específicas sobre o papel e impacto sobre as ideias de chance e aleatoriedade em diferentes processos e situações em que são requeridas.

Gal (2005) aponta que as grandes ideias podem ser entendidas como: a aleatoriedade, independência, variação e outros elementos, como a capacidade dos alunos compreenderem a derivação, representação, interpretação e implicação de declarações probabilísticas. Assim, torna-se imperativo que as pessoas passem por um letramento probabilístico para lidar com a ampla variedade de situações do mundo real que incluem a geração ou interpretação de mensagens probabilísticas, assim como a tomada de decisões diante de fenômenos aleatórios que se apresentem.

Em trabalhos realizados no Brasil voltados a estudos aderentes ao tema dessa pesquisa, Oliveira Júnior e Kian (2022) e Oliveira Júnior, Kian e Santos (2023), indicam que as crianças têm formalizado, mesmo que intuitivamente, desde os anos iniciais da educação básica, noções probabilísticas como, por exemplo, determinar o significado das palavras acaso, aleatório, incerteza e identificando fenômenos ou experimentos aleatórios, devendo haver a preocupação em trabalhar esses conceitos. Esses conhecimentos se constituem como um elemento básico inicial que os alunos apresentam para então, construir uma conexão com a linguagem probabilística, permitindo a esses começar a usar uma linguagem precisa e especializada.

Ainda se destaca Oliveira Júnior, Silva Santos e Kian (2023), que indicam que a predominância de palavras e expressões verbais da linguagem estatística e probabilísticas comum estão relacionadas principalmente ao significado intuitivo e seu uso no cotidiano.

3 Procedimentos metodológicos

Para a análise dos resultados obtidos no estudo, utilizou-se a análise textual, um tipo específico de análise de dados, na qual trata-se de material verbal transcrito, ou seja, de textos (Nascimento-Schulze & Camargo, 2000). Essa análise tem várias finalidades, sendo possível analisar textos, entrevistas, documentos, redações etc. No caso desse estudo, analisou-se as respostas indicadas pelos alunos ao instrumento disponibilizado via *Google Forms*.

Considerando Nascimento-Schulze e Camargo (2000), tendo como base a compreensão da análise textual, descreve-se no Quadro 1 conceitos importantes e associados ao contexto metodológico desse trabalho.

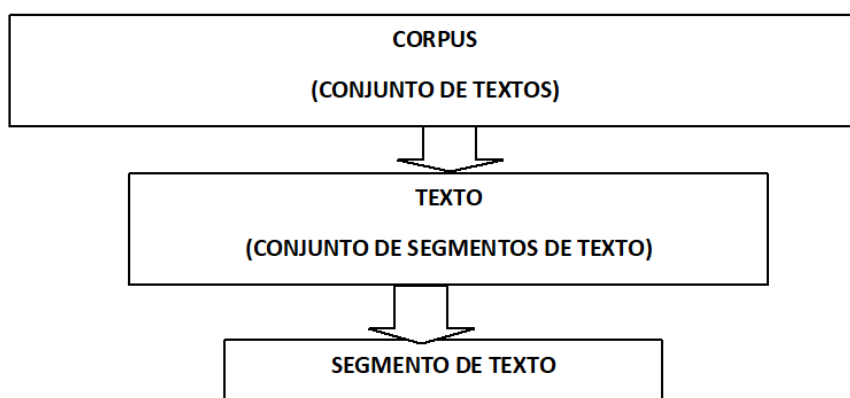
Quadro 1: Conceitos importante para a análise textual

Conceito	Descrição do conceito
<i>Corpus</i>	Construído pelo pesquisador, é o conjunto texto que se pretende analisar. No caso dessa trabalho, por exemplo, um <i>corpus</i> será composto das respostas a questões abertas, que faz parte de um questionário empregado como instrumento de uma pesquisa do tipo enquete.
Textos	Em se tratando de respostas a questões abertas de um questionário, cada texto será composto da adição dos trechos obtidos das respostas somente quando elas se referirem a um mesmo tema (uma mesma questão ou pergunta). Caso as questões referiram-se a temas ou aspectos diferentes, é necessário realizar uma análise para cada questão. Como mencionado anteriormente, a análise é sensível à estruturação do estímulo que produz o material textual, e isto é uma importante fonte de invalidação das conclusões. Quando as respostas apresentarem uma média em torno de três à cinco linhas, é necessário um número bem maior de respostas para a constituição de um corpus de análise.
Segmentos de texto	São excertos de texto, na maior parte das vezes, do tamanho de até três linhas, dimensionadas pelo próprio <i>software</i> em função do tamanho do <i>corpus</i> . Os segmentos de textos são considerados o ambiente das palavras.

Fonte: Elaborados pelos autores segundo Nascimento-Schulze e Camargo (2000).

A Figura 1 indica a relação sequencial dos conceitos que se fazem importantes para a realização de uma análise textual.

Figura 1: Noções de *corpus*, texto, segmento de texto.



Fonte: Elaborados pelos autores segundo Nascimento-Schulze e Camargo (2000).

A definição das unidades de análise será realizada pelo pesquisador e dependerá da natureza da pesquisa, ou seja, as respostas do instrumento dizem respeito às respostas de "n" participantes a uma questão aberta e cada resposta será um texto e teremos, portanto, "n" textos.

O conjunto de textos constituirá um corpus de análise. O corpus adequado à análise do tipo Classificação Hierárquica Descendente – CHD, deve constituir-se num conjunto textual centrado em um tema. O material textual deve ser monotemático, pois a análise de textos sobre vários itens previamente estruturados ou diversos temas resulta na reprodução da estruturação prévia dos mesmos (Camargo, 2005).

Destaca-se que as análises do tipo CHD, para serem úteis à classificação de qualquer material textual, requerem uma retenção mínima de 75% dos segmentos de texto, sendo que para alguns autores há de se considerar a possibilidade de se considerar o aproveitamento de 70% dos segmentos de textos (Camargo & Justo, 2013).

A análise de similitude é baseada na teoria dos grafos cujos resultados auxiliam no estudo das relações entre objetos (Marchand & Ratinaud, 2012; Salviati, 2017). Marchand e Ratinaud (2012) indicam que as análises de similitude são utilizadas frequentemente por pesquisadores das representações sociais (cognição social), possibilitando identificar as coocorrências entre as palavras e seu resultado traz indicações da conexidade entre essas, auxiliando na identificação da estrutura da representação.

Para Tavares (2019), o objetivo destas análises é identificar como foi realizada a construção do discurso que compõe o corpus textual, identificando a estrutura base que relaciona as formas assim como, os temas por grau de relevância que conectam as partes importantes que caracterizam os textos contidos na base de dados.

Assim, para a análise dos dados, utilizou-se o *software* IRaMuTeQ (Interface R para Texto Multidimensional e Análise de Questionário), que foi desenvolvido como ferramenta auxiliar do processo de codificação dos elementos trazidos por meio da coleta de dados (Mutombo, 2013).

O referido software foi utilizado para realizar uma análise lexical quantitativa que considera a palavra como unidade, também oferecendo a sua contextualização no *corpus* ou no instrumento de pesquisa ou questionário. Cada pergunta do instrumento é composta por conteúdos semânticos, que formaram o banco de dados ou *corpus* analisado pelo *software*.

Portanto, o trabalho utilizou de forma conjugada um *software* desenvolvido para a análise textual e de um procedimento de análise de conteúdo clássica. O IRaMuTeQ apoia-se em cálculos efetuados sobre a coocorrência de palavras em segmentos de texto, buscando distinguir classes de palavras que representem formas distintas de discurso sobre o tópico de interesse da investigação.

Complementando esses aspectos, a análise de conteúdo identifica a pluralidade temática presente num conjunto de textos, e pondera a frequência desses temas no conjunto, permitindo, via comparação entre os elementos do corpus (palavras ou sentenças), a constituição de agrupamentos de elementos de significados mais próximos, viabilizando a formação de categorias mais gerais de conteúdo.

Destacando a análise de conteúdo, Bardin (2015) diz que pode ser aplicada, desde a análise de comunicações individuais até coletivas, bem como a mensagens de diferentes naturezas e suportes. Ou seja, utiliza-se de códigos linguísticos (escritas ou orais), icônicos (sinais, grafismos, fotografia, filmes, etc.) e outros códigos semióticos (tudo que não sendo linguístico pode ser portador de significações como gestos, comportamentos, sintomas patológicos ou espaços) que podem ser passíveis de análise de conteúdo. Estes códigos possuem significações, tornando-se passível de inferência e técnicas para identificá-la.

Segundo Bauer (2002), a análise de conteúdo pode reconstruir mapas de conhecimento à medida que eles estão corporificados em textos, considerando que podem ser utilizadas diferentes linguagens para representar o mundo como conhecimento e autoconhecimento. Dessa forma, para reconstruir esse conhecimento, é necessário ir além da classificação das unidades do texto e orientar-se na construção de redes de unidades de análise para representar o conhecimento não apenas por elementos, mas também em suas relações.

4 Resultados e discussões

Os chamados fenômenos ou experimentos aleatórios⁴ são processos reais ou hipotéticos nos quais os possíveis resultados podem ser identificados antecipadamente, mas não podemos prever com certeza o que vai acontecer, pois o acaso representa um papel preponderante. Portanto, um experimento aleatório possui duas características fundamentais: (1) O resultado não pode ser previsto com certeza; (2) Embora não seja possível afirmar que resultado em particular ocorrerá, é possível descrever o conjunto de todos os resultados possíveis do experimento.

Assim, dentre as situações apresentadas aos alunos como fenômenos ou experimentos aleatórios, a primeira é a observada na Figura 2.

Figura 2: Situação 1 indicada na *Google Forms* e que está associada a um fenômeno ou experimento aleatório

1 - Considere a seguinte situação: “Sair na rua e encontrar um amigo da sua escola”. Escreva uma palavra ou algumas palavras sobre a possibilidade dessa situação ocorrer.



Fonte: Elaborado pelos autores

No resumo das respostas de 60 alunos (98,4% do total) à primeira questão que indica um fenômeno ou experimento aleatório, observa-se que as palavras de maior frequência, ou ainda, aquelas que apresentam o número de formas ativas com frequência maior ou igual a 3 são indicadas na Tabela 1. Lembra-se que as formas ativas e suplementares são as palavras consideradas ativas (adjetivos, nomes, verbos e advérbios) e suplementares (artigos e pronomes), exceto os artigos e as preposições que foram eliminados.

Tabela 1: Indicação das formas ativas com três ou mais indicações na análise referente às respostas dos alunos à primeira questão que apresenta um fenômeno ou experimento aleatório

Palavra (função lexical)	Frequência	Palavra (função lexical)	Frequência
Bem (advérbio)	9	Estar (verbo)	4
Amigo (nome)	9	Chance (nome)	3
Acontecer (verbo)	9	Bairro (nome)	3
Não (advérbio)	7	Mesmo (adjetivo)	3
Possível (adjetivo)	6	Achar (verbo)	3
Provável (adjetivo)	6	Certeza (nome)	3
Encontrar (verbo)	5	Coincidência (nome)	3
Morar (verbo)	4	Ocorrer (verbo)	3
Rua (nome)	4		

Fonte: Organizado a partir da saída do IRaMuTeQ

⁴ <http://clubes.obmep.org.br/blog/sala-de-estudo-probabilidades/>

Observa-se na Tabela 1 que a palavra com maior frequência é o advérbio “Bem”, sendo que pelo dicionário online de português (<https://www.dicio.com.br/acontecer/>), seu significado voltado à aspectos probabilísticos é aquilo em que há correção, de modo justo, de modo exato.

Identifica-se ainda que o verbo “Acontecer” se refere a processo que está diretamente relacionado com o significado dado pelos alunos à noção de acaso, ratificado pelo sistema periférico da análise de similitude, no qual, contém as seguintes palavras: acaso, algo, motivo, coisa, esperar, nada e imprevisível.

Outras duas palavras merecem ser destacadas que são: 1) O adjetivo “Possível” que tem (todas) as condições essenciais para se desenvolver, realizar ou existir; ou ainda, algo que talvez aconteça, que em uma grande possibilidade para que se efetive; 2) O adjetivo “Provável” que reflete algo esperável ou que provavelmente acontecerá ou ainda com grande possibilidade para que aconteça ou se realize.

Para iniciar a apresentação do que os alunos indicaram referente às respostas dos alunos à primeira questão que indica um fenômeno ou experimento aleatório, utilizou-se o método de Reinert que propõe uma classificação hierárquica descendente segundo o método descrito por Reinert (1990), visando obter classes de segmentos de texto (ST) que, ao mesmo tempo, observa-se vocabulário semelhante entre si e vocabulário diferente das ST das outras classes.

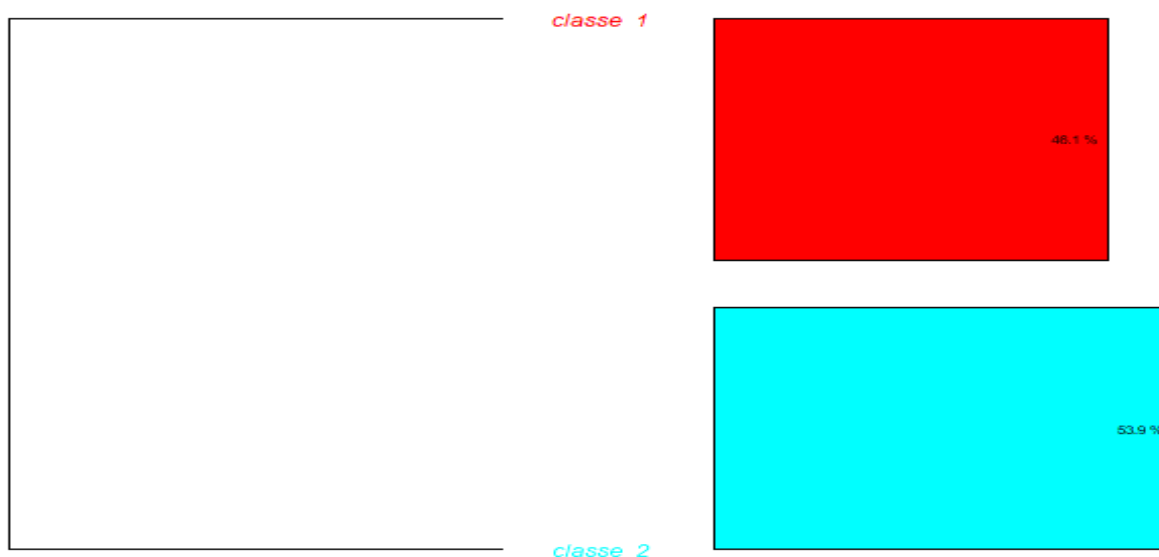
Enfatiza-se que a escolha pela utilização de uma ou outra técnica de análise depende das características do problema e dos objetivos da pesquisa (Leblanc, 2015). Nessa direção, o referencial teórico-metodológico do pesquisador, acrescido do suporte de *softwares* de análise lexicométrica, podem conferir maior confiabilidade às inferências realizadas em pesquisas qualitativas (Camargo & Justo, 2013).

Portanto, esta análise é baseada na proximidade léxica e na ideia de que palavras usadas em contexto similar que estão associadas ao mesmo mundo léxico e são parte de mundos mentais específicos ou sistemas de representação. Nessa análise, os segmentos de texto são classificados de acordo com seu respectivo vocabulário e o conjunto de termos é particionado de acordo com a frequência das raízes das palavras. O sistema procura obter classes formadas por palavras que são significantemente associadas com aquela classe (a significância começa com o teste de quiquadrado - χ^2).

Segundo Oliveira (2015) é uma das análises mais importantes do IRamuteq, pois, o software utiliza a lógica de correlação, partindo de segmentações do corpus textual, juntamente com a lista de formas reduzidas e o dicionário (no caso em português) disponibilizado no mesmo para apresentar um esquema hierárquico de classes. Dessa forma é processado o texto de modo que possam ser identificadas classes de vocabulário, sendo possível inferir quais ideias o corpus textual deseja transmitir, ou seja, a análise é feita a partir de uma lógica estatística processada por computador e aplicada de forma lexical.

Assim, na Figura 3 apresenta-se o dendrograma gerado na CHD dos resultados do IRaMuTeQ, que indica as partições que foram feitas no *corpus* até que se chegasse às duas classes finais.

Figura 3: Resultado da Classificação pelo Método de Reinert referente às respostas dos alunos à primeira questão que apresenta um fenômeno ou experimento aleatório: Dendrograma



Fonte: Saída do *software* IRaMuTeQ.

Assim, no resultado da Classificação pelo Método de Reinert: Dendrograma, Figura 3, o *corpus* “Corpo” foi dividido em dois *subcorpus*, sendo que a Classe 1 representa 46,1% do corpus total e a Classe 2 representa 53,9%.

Portanto, as duas classes contêm as formas ativas ou palavras organizadas que apresentaram maior frequência, em ordem decrescente, e que são significativas para representar cada um dos subcorpus por meio do teste de associação qui-quadrado gerado nos relatórios do Iramuteq, ou seja, a maior aderência delas na classe e entre as classes e que pode ser observado na Tabela 2.

Tabela 2: Indicação das formas ativas com três ou mais indicações que formam as duas classes na análise referente às respostas dos alunos à primeira questão que apresenta um fenômeno ou experimento aleatório.

Classe 1	Frequência	Classe 2	Frequência
Amigo (Nome)	9	Bem (Advérbio)	9
Acontecer (Verbo)	9	Provável (Adjetivo)	6
Não (Advérbio)	7		
Encontrar (Verbo)	5		
Rua (Nome)	4		

Fonte: Organizado a partir da saída do IRaMuTeQ

Tomando o dendrograma (Figura 3) e as palavras de maior frequência e que apresentaram relação significativa entre elas (Tabela 2), buscou-se identificar as respostas dos alunos à primeira questão referente a um fenômeno ou experimento aleatório.

Assim, na Classe 1 que se denominou por “Apresentar situações em que vivenciaram ou poderia vivenciar a situação proposta”, tomando as combinações das palavras que indicam relação significativa, destaca-se os seguintes enxertos:

**** *_n_02 *_Id_2 *_Gen_1 *_GM_1

Já aconteceu alguns dias eu falo um oi ele para um lado e eu para o outro ou nós conversamos.

**** *_n_03 *_Id_3 *_Gen_2 *_GM_1

Encontrei um amigo na rua.

**** *_n_06 *_Id_1 *_Gen_1 *_GM_1

Você **pode encontrar um amigo da sua escola**, se ele estiver no parque, na praça ou no mercado.

**** *_n_10 *_Id_1 *_Gen_1 *_GM_1

Pode acontecer. Já aconteceu comigo.

**** *_n_19 *_Id_2 *_Gen_2 *_GM_1

Muitas, se o seu amigo ou amiga morar no mesmo bairro ou cidade que você, poucas se não.

**** *_n_20 *_Id_3 *_Gen_2 *_GM_1

Coincidência eu encontrar meu amigo na rua.

**** *_n_22 *_Id_3 *_Gen_2 *_GM_1

Pode acontecer, mas não tem muita probabilidade.

**** *_n_25 *_Id_2 *_Gen_2 *_GM_1

Cinquenta por cento, pois todos os meus amigos moram no mesmo bairro que eu.

**** *_n_45 *_Id_3 *_Gen_2 *_GM_1

Eu já encontrei vários amigos na rua, a possibilidade de isso acontecer novamente é bem alta.

**** *_n_47 *_Id_3 *_Gen_1 *_GM_1

Não achar ninguém, achar crianças que você não conheça ou nem ir pra rua.

**** *_n_53 *_Id_2 *_Gen_1 *_GM_1

Acho que isso **raramente vai acontecer.**

**** *_n_56 *_Id_2 *_Gen_2 *_GM_1

Muita chance de isso acontecer.

Na Classe 2 na qual se denominou por “Utilizar termos probabilísticos para explicar a situação proposta”, tomou-se as combinações das palavras que apresentam relação significativa, destacando os seguintes enxertos:

**** *_n_04 *_Id_2 *_Gen_2 *_GM_1

Pouco provável.

**** *_n_05 *_Id_2 *_Gen_1 *_GM_1

Provável.

**** *_n_12 *_Id_3 *_Gen_2 *_GM_2

É bem difícil ocorrer isso.

**** *_n_15 *_Id_2 *_Gen_2 *_GM_1

Bem provável porque várias pessoas moram perto de mim.

**** *_n_17 *_Id_2 *_Gen_1 *_GM_1

Improvável, aleatório.

**** *_n_18 *_Id_3 *_Gen_2 *_GM_1

É provável que aconteça.

**** *_n_24 *_Id_2 *_Gen_2 *_GM_2

Bem provável que isso aconteça.

**** *_n_35 *_Id_2 *_Gen_2 *_GM_1

Bem pouco. Cinco por cento em cem.

**** *_n_38 *_Id_2 *_Gen_1 *_GM_1

Possível, improvável, aleatório, surpresa.

**** *_n_39 *_Id_3 *_Gen_2 *_GM_1

Muito provável. Moro muito perto da escola.

**** *_n_50 *_Id_2 *_Gen_1 *_GM_1

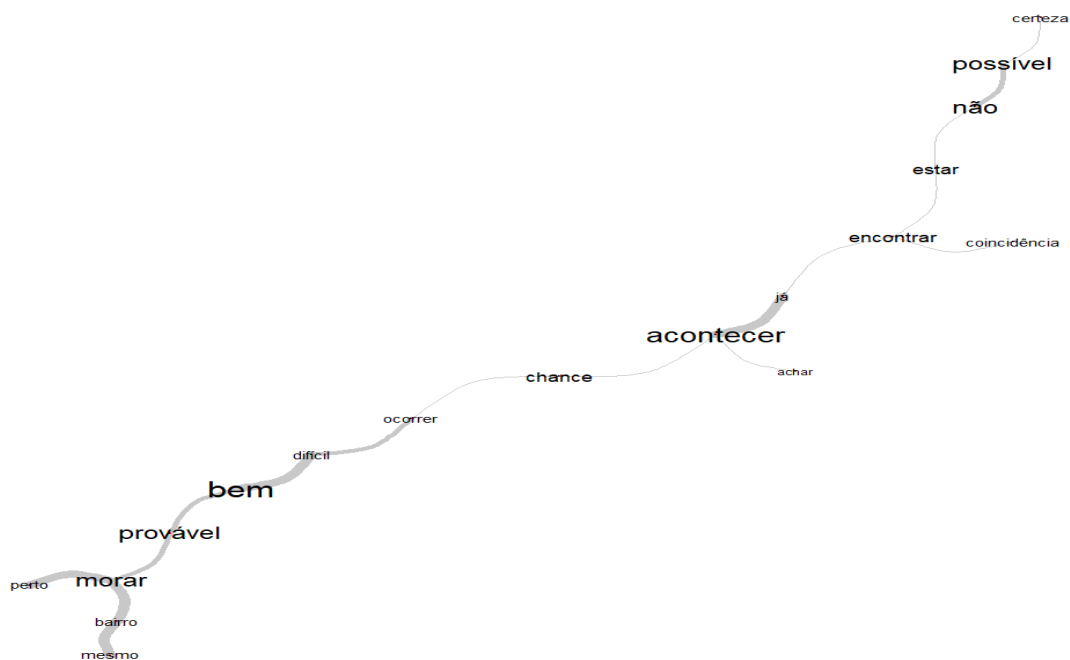
Bem difícil dessa situação ocorrer.

**** *_n_60 *_Id_3 *_Gen_2 *_GM_1

Bem difícil.

Para complementar as análises, apresenta-se uma análise de similitude utilizando o software do IRaMuTeQ que por meio de um dendrograma (Figura 3) indica a quantidade maior encontrada de termos agrupados demonstrando objetivamente as mais frequentes.

Figura 3: Grafo de Similitude referente às respostas dos alunos à primeira questão que apresenta um fenômeno ou experimento aleatório.



Fonte: Dendrograma gerado pelo software IRaMuTeQ

Para entendermos e analisarmos foi preciso identificar como foi feita a construção do discurso que compõem o corpus do texto, e assim identificarmos a estrutura inicial que relaciona as formas lexicais e o contexto presente na base de dados.

Observa-se, da mesma forma, que a CHD que o substantivo masculino “Bem” tem destaque no dendrograma associado ao verbo acontecer e outras formas lexicais e bem próximo o adjetivo provável que se configura como um outro aspectos que os alunos avaliaram a situação proposta.

Para Marocci e Nacarato (2013) é necessário inicialmente compreender o que os alunos entendem sobre os termos mais frequentes, para depois ampliar os significados já construídos e progredir na formação de conceitos probabilísticos, pressupondo processos de comunicação e reflexão pelos alunos.

Com isso percebemos em nossos estudo que as situações diárias provocam na criança a utilização de linguagem probabilística apresentando indícios do entendimento sobre conceitos probabilísticos.

Apresenta-se uma segunda situação que também partiu de situação associado a um fenômeno ou experimento aleatório (Figura 4).

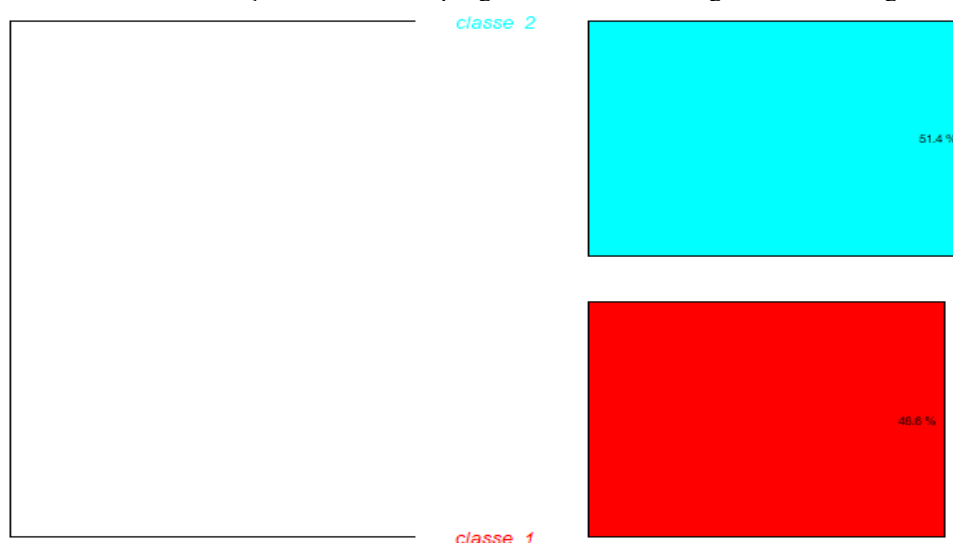
Figura 4: Situação 4 apresentada no *Google Forms* e que está associada a um fenômeno ou experimento aleatório

2) Considere a seguinte situação: “Encontrar uma criança em sua escola que goste de verduras e legumes”. Escreva uma palavra ou algumas palavras sobre a possibilidade dessa situação ocorrer.



Fonte: Elaborado pelos autores

No caso dessa questão realizou-se uma CHD tomando os resultados pelo IRaMuTeQ, tendo acesso ao dendrograma (Figura 5), que apresenta as partições que foram feitas no *corpus* até que se chegasse às duas classes finais.

Figura 5: Resultado da Classificação pelo Método de Reinert referente ao experimento ou fenômeno aleatório “Encontrar uma criança em sua escola que goste de verduras e legumes”: Dendrograma

Fonte: Saída do software IRaMuTeQ

Assim, no resultado da Classificação pelo Método de Reinert: Dendrograma, Figura 5, o *corpus* “Corpo” foi dividido em dois *subcorpus*, sendo que a Classe 1 representa 48,6% do *corpus* total e a Classe 2 representa 51,4%.

Portanto, as duas classes contêm as formas ativas ou palavras organizadas que apresentaram maior frequência, em ordem decrescente, e que são significativas para representar cada um dos *subcorpus* por meio do teste de associação quiquadrado gerado nos relatórios do IRaMuTeQ, ou seja, a maior aderência delas na classe e entre elas (Tabela 3).

Tabela 3: Indicação das formas ativas com três ou mais indicações que formam as duas classes na análise referente ao experimento ou fenômeno aleatório “Encontrar uma criança em sua escola que goste de verduras e legumes”

Classe 1	Frequência	Classe 2	Frequência
Legume (Nome)	7	Difícil (Adjetivo)	6
Comer (Verbo)	5	Acontecer (Verbo)	5
Só (Adjetivo)	5	Gostar (Verbo)	5
Gosto (Nome)	5	Possibilidade (Nome)	4
Saudável (Adjetivo)	4		

Fonte: Organizado a partir da saída do IRaMuTeQ

Tomando o dendrograma (Figura 5) e as palavras de maior frequência e que apresentaram relação significativa entre elas (Tabela 3), buscou-se identificar as definições indicadas referente a opinião dos alunos para o experimento ou fenômeno aleatório “Encontrar uma criança em sua escola que goste de verduras e legumes”.

Indica-se a seguir, respostas de alguns alunos que geraram a Classe 1 na qual se chamou de “Indicar situações específicas cotidianas bem como linguagem probabilística específica que justificam a possibilidade, ou não, de gostar de legumes e verduras”, indicando as associações do substantivo masculino “Legume” (núcleo central dessa classe) associado com as seguintes palavras periféricas, nessa ordem: (1) Verbo “Comer”; (2) Adjetivo “Só”; (3) Substantivo Masculino “Gosto”; (4) Adjetivo “Saudável”; e que destacou-se em negrito, quais sejam:

**** *_n_06 *_Id_1 *_Gen_1 *_GM_1

Existem crianças que são vegetarianos, eles só comem legumes e verduras.

**** *_n_17 *_Id_2 *_Gen_1 *_GM_1

A mãe dela deve ter ensinado sobre legumes desde bebê, ou seja, é possível.

**** *_n_18 *_Id_3 *_Gen_2 *_GM_1

Eu gosto de legumes, só não gosto de brócolis e couve-flor.

**** *_n_20 *_Id_3 *_Gen_2 *_GM_1

Muita, pois a maioria das crianças come verduras e legumes.

**** *_n_25 *_Id_2 *_Gen_2 *_GM_1

Oitenta por cento. Até eu amo muito verduras e legumes.

**** *_n_37 *_Id_2 *_Gen_1 *_GM_1

Parabéns. Você come alimentos saudáveis.

**** *_n_41 *_Id_3 *_Gen_1 *_GM_1

Criança saudável.

**** *_n_47 *_Id_3 *_Gen_1 *_GM_1

Sou eu. Só não gosto de milho, na verdade ele só que come carne.

**** *_n_49 *_Id_3 *_Gen_1 *_GM_1

Eu também gosto de legumes, são saudáveis.

**** *_n_52 *_Id_3 *_Gen_2 *_GM_1

Porque a mãe ensinou desde bebê comer legumes e vegetais.

**** *_n_58 *_Id_2 *_Gen_1 *_GM_1

Ótimo. Alimentação saudável.

Em relação à Classe 2 que chamaremos de “Listar fatos que indiquem dificuldade, ou não, em comer legumes e verduras, além de indicar aspectos intuitivos e específicos em relação à probabilidade”, indicando as associações do adjetivo “Difícil” (núcleo central dessa classe) associado com as seguintes palavras periféricas, nessa ordem: (1) Verbo “Acontecer”; (2) Verbo “Gostar”; (3) Substantivo masculino “Possibilidade”; destacado em negrito, quais sejam:

**** *_n_03 *_Id_3 *_Gen_2 *_GM_1

Difícil.

**** *_n_04 *_Id_2 *_Gen_2 *_GM_1

Provável, pode acontecer.

**** *_n_07 *_Id_3 *_Gen_1 *_GM_1

Muito difícil, pois criança gosta de bolacha salgadinho.

**** *_n_12 *_Id_3 *_Gen_2 *_GM_2

Bem difícil, eu acho.

**** *_n_15 *_Id_2 *_Gen_2 *_GM_1

Pode acontecer, eu mesmo amo.

**** *_n_16 *_Id_3 *_Gen_2 *_GM_1

Essa **possibilidade é bem difícil de acontecer**.

**** *_n_22 *_Id_3 *_Gen_2 *_GM_1

Pode sim acontecer, pois muitas crianças no mundo gostam.

**** *_n_27 *_Id_3 *_Gen_2 *_GM_1

Possivelmente, mas difícil.

**** *_n_30 *_Id_2 *_Gen_2 *_GM_2

Possibilidade.

**** *_n_36 *_Id_2 *_Gen_1 *_GM_1

Difícil.

**** *_n_40 *_Id_2 *_Gen_2 *_GM_1

Total possibilidade.

**** *_n_45 *_Id_3 *_Gen_2 *_GM_1

A possibilidade é média, pois tem crianças que gostam e outras que não gostam.

**** *_n_50 *_Id_2 *_Gen_1 *_GM_1

Isso **não é tão fácil de acontecer**.

Na Classe 2 encontrou-se o adjetivo “Difícil” em sua grande maioria seguida do verbo “Acontecer”, e o verbo “Gostar” e o substantivo feminino “Possibilidade”, que se associados e destacados nestas falas, indicam a possibilidade de indicar aspectos da vida real entre alunos desta faixa etária, portanto este contexto representa de forma muito simples o entendimento dos alunos e o seu cotidiano que levam para a resposta do experimento aleatório em que se encaixa. Considera-se que estes dois grupos fazem o uso tanto da linguagem probabilística, quanto da linguagem menos formal (intuitiva) encaminhando-se para uma linguagem formal.

Por fim, apresenta-se as seguintes definições do que seja experimento ou fenômeno aleatório (probabilístico), quais sejam:

1. Fenômeno aleatório é a situação ou acontecimento cujos resultados não podem ser previstos com certeza (Magalhães & Lima, 2004);
2. Experimento de probabilidade é uma ação ou tentativa pela qual resultados (contagens, medições ou respostas) são obtidos (Larson & Farber, 2010);
3. Experimento aleatório é um experimento no qual podemos descrever o conjunto de todos os resultados possíveis, mas não podemos dizer, a priori, qual desses resultados vai acontecer (Pinheiro, Cunha, Carvajal & Gomes, 2015).

Por fim, Marocci e Nacarato (2013), expressam que o sentido atribuído a expressões ou palavras a situações do cotidiano é subjetivo, ou seja, a compreensão de certa situação envolvendo a probabilidade será diferente para cada pessoa. A esse respeito, Konold (1991) indica que os alunos possuem uma compreensão coerente, proveniente do contato com circunstâncias que envolvem incerteza e, a partir disso, procuram dar sentido às situações vivenciadas.

5 Considerações finais

O estudo sugere que esses alunos do quinto ano do Ensino Fundamental de escola no município de Barueri, São Paulo, possuem conhecimentos prévios e experiências do seu contexto cotidiano que permitem a exploração da probabilidade desde as primeiras fases do seu aprendizado.

Partindo dessa premissa, o estudo indica que o ensino dos conceitos probabilísticos deve

envolver o considerar o desenvolvimento das primeiras noções e elementos de aproximação a essa área do conhecimento para a aquisição e crescimento da linguagem probabilística.

Considerando a complexidade e o elevado grau de abstração dos conceitos probabilísticos é imperativo empreender o processo gradualmente, permitindo que os alunos adquiram um entendimento apropriado da linguagem específica associada aos termos associados à linguagem probabilística.

Além disso, os resultados dessa pesquisa convergem para os estudos de Vásquez (2014, 2018) e Vásquez e Alsina (2017) ao afirmarem que se deve fomentar a evolução da alfabetização probabilística, acreditando ser crucial contemplar diversas fases nesse processo de aquisição, incluindo a assimilação dos conceitos fundamentais para o desenvolvimento do conhecimento probabilístico.

Um ponto central, focalizado neste estudo, é a linguagem da probabilidade, que utiliza vocabulário comum ao dia a dia das crianças para referenciar a situações-problema associados a fatos incertos.

Outro aspecto apresentado por esse estudo converge para as afirmações indicadas por Gal (2005), ou seja, propor atividades que estejam próximas ao cotidiano do aluno, possibilitando que os alunos sejam letrados probabilisticamente, requerendo, por exemplo, a linguagem probabilística, permitindo a partir de um conhecimento de mundo ou de contexto que os alunos apresentem expectativas específicas sobre o papel e impacto sobre as ideias de chance e aleatoriedade em diferentes processos e situações em que são requeridas.

Esse estudo destaca a importância em focar na linguagem probabilística que está ligada ao desenvolvimento conceitual, ou seja, o significado que as crianças atribuem às palavras que são utilizadas para nos referirmos ao acaso ou à probabilidade. Reforça-se a necessidade em analisar os termos e expressões orais e escritas das crianças, buscando a compreensão sobre esse vocabulário, uma vez que as crianças utilizam frequentemente vocabulário impreciso ou incorreto.

Por fim, partindo desse estudo, reforça-se a necessidade de elaboração de mais e diferentes tarefas probabilísticas que se destinem a explorar a linguagem probabilística utilizada pelos alunos. Essas tarefas devem destacar situações de incerteza seguindo a proposta de ensino de estatística e probabilidade segundo a BNCC (Brasil, 2018), além dos documentos norte-americanos denominado Diretrizes para Avaliação e Instrução em Educação Estatística – GAISE I e II (Franklin *et al.*, 2007; Bargagliotti *et al.*, 2020), alinhando-se às recomendações para a educação básica no ensino de probabilidade.

Referências

- Bardin, L. (2015). *Análise de conteúdo*. Lisboa: Edições 70.
- Bargagliotti, A.; Franklin, C.; Arnold, P.; Gold, R.; Johnson, S.; Perez, L. & Spangler, D. A. (2020). *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II): a framework for Statistics and Data Science Education*. Alexandria, VA. American Statistical Association.
- Batanero, C. (2013). Sentido Estadístico: componentes y desarrollo. In: J. M. Contreras, J. M., G. R. Cañadas, M. M. Gea & Arteaga, P. (Ed.). *Actas de las I Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 55-61). Granada: Universidad de Granada.

- Batanero, C. (2015). Understanding randomness: challenges for research and teaching. In: *Proceedings of 9th Congress of European Research in Mathematics Education – CERME* (pp 34-49). Prague, Czech Republic.
- Batanero, C. & Godino, J. (2002). *Estocástica y su didáctica para maestros: Proyecto Edumat Maestros*. Granada: Universidade de Granada.
- Bauer, M. W. (2002). Análise de conteúdo clássica: uma revisão. In: M. W. Bauer & G. Gaskell (Ed.). *Pesquisa qualitativa com texto, imagem e som: um manual prático*. Tradução de P. A. Guareschi. (pp. 189-217). Petrópolis, RJ: Vozes.
- Brasil. Ministério da Educação. Secretaria de Educação Básica. (2018). *Base Nacional Comum Curricular*. Brasília, DF.
- Bryant, P. & Nunes, T. (2012). *Children's understanding of probability: a literature review*. London: Nuffield Foundation.
- Camargo, B. V. & Justo, A. M. (2013). *Tutorial para uso do software de análise textual IramuTeQ*. <http://www.iramuteq.org/documentation/fichiers/tutoriel-en-portugais>
- Camargo, B. V. (2005). ALCESTE: Um programa informático de análise quantitativa de dados textuais. In: A. S. P. Moreira, B. V. Camargo, J. C. Jesuíno & S. M. Nobrega (Eds.). *Perspectivas teórico-metodológicas em representações sociais* (pp. 511-539). João Pessoa, PB: UFPB.
- Franklin, C.; Kader, G.; Carrol, J. M. J.; Peck, R.; Perry, M. & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics (GAISE) Education: A Pre-K-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.
- Gal, I. (2005). Towards probability literacy for all citizens: building blocks and instructional dilemmas. In: G. A. Jones (Ed.). *Exploring probability in school* (pp. 39-63). Springer.
- Konold, C. (1991). Understanding students' beliefs about probability. In: E. Von Glasersfeld (Ed.). *Radical constructivism in mathematics education*. (pp. 139-156). Dordrecht: Kluwer.
- Larson, R. & Farber, B. (2010). *Estatística Aplicada*. (4. ed.). São Paulo, SP: Pearson Prentice Hall.
- Leblanc, J-M. (2015). Proposition de protocole pour l'analyse des données textuelles: Pour une démarche expérimentale en lexicométrie. *Nouvelles Perspectives en Sciences Sociales (NPSS)*, 11(1), 25-63.
- Magalhães, M. N. & Lima, A. C. P. (2005). *Noções de Probabilidade e Estatística*. (6. ed.). São Paulo, SP: EDUSP.
- Marchand, P. & Ratinaud, P. (2012). L'analyse de similitude appliqué aux corpus textuels: les primaires socialistes pour l'élection présidentielle française. In: *Actes de 11 Journées internationales d'Analyse statistique des Données Textuelles, JADT* (pp. 687-699). Liège, Belgique.
- Marocci, L. M. & Nacarato, A. M. (2013). Um ambiente de aprendizagem baseado na resolução de problemas: a possibilidade de circulação de significados sobre a Probabilidade por meio da linguagem. *Educação Matemática Pesquisa*, 15(1), 101-123.
- Morgado, A. C.; Pitombeira, J. C.; Carvalho, P. C. P. & Fernandez, P. (2004). *Análise combinatória e probabilidade*. Rio de Janeiro, RJ: SBM.

- Mutombo, E. (2013). A bird's-eye view on the EC environmental policy framing: ten years of Impact assessment at the commission. In: *Proceedings of 1st International Conference on Public Policy - ICCPP*. Grenoble: França.
- Nascimento-Schulze, C. M. & Camargo, B. V. (2000). Psicologia social, representações sociais e métodos. *Temas de psicologia*, 8(3), 287-299.
- Oliveira, J. F. R. de. (2015). Tutorial (básico) de utilização do Iramuteq. Goiânia, GO: UFG. https://files.cercomp.ufg.br/weby/up/771/o/Tutorial_-_Revis%C3%A3o.pdf
- Oliveira Júnior, A. P.; Datori Barbosa, N.; Souza, N. G. S. & Cardoso, K. M. (2019). A apreensão do conceito de experimento aleatório: Resolução de Problemas e jogo pedagógico. *Caminhos da Educação Matemática em Revista*, 9(2), 238-257.
- Oliveira Júnior, A. P. de & Kian, F. A. (2022). A linguagem probabilística de alunos do quinto ano do ensino fundamental: identificando fenômenos ou experimentos aleatórios. *Conjecturas*, 22(1), 1572-1586.
- Oliveira Júnior, A. P. de; Kian, F. A. & Santos, L. R. S. (2023). Ambiguidade lexical em probabilidade: conhecimento de alunos do ensino fundamental sobre acaso, aleatório e incerteza. *Areté*, 9(17), 99-126.
- Oliveira Júnior, A. P. de; Silva Santos, L. R. & Kian, F. A. (2023). Evaluating Lexical Ambiguities About Brazilian Students in The Fifth Year of Elementary School Towards Statistical Graphs and Tables. *Paradigma*, 44(3), 151-174.
- Pinheiro, J. I. D.; Cunha, S. B. da; Carvajal, S. S. R. & Gomes, G. C. (2015). *Estatística Básica: a arte de trabalhar com dados*. (2 ed.). Rio de Janeiro, RJ: Elsevier.
- Reinert, M. (1990). ALCESTE, une méthodologie d'analyse des données textuelles et une application: Aurélia de G. de Nerval. *Bulletin de méthodologie sociologique*, 26(1), 24-54.
- Salviati, M. E. (2017). *Manual do Aplicativo Iramuteq* (Apostila de Curso). Planaltina, DF: Embrapa Cerrados. <http://www.iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>
- Tavares, I. A. (2019). *Iramuteq: um software para análises estatísticas qualitativas em corpus textuais*. 2019. 40f. Trabalho de Conclusão de Curso (Graduação em Estatística). Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal, RN.
- Vásquez, C. O. (2014). *Evaluación de los conocimientos didáctico-matemáticos para la enseñanza de la probabilidad de los profesores de educación primaria en activo*. 2014. 56f. Tesis (Doctoral em Educación). Programa de Doctorado em Educación, Universitat de Girona, Espanha.
- Vásquez, C. O. (2018). Surgimiento del lenguaje probabilístico en el aula de educación primaria. *REnCiMa*, 9(2), 374-389.
- Vásquez, C. A. O. & Alsina, A. (2017). Lenguaje probabilístico: un camino para el desarrollo de la alfabetización probabilística. Un estudio de caso en el aula de Educación Primaria. *Bolema*, 31(57), 454-478.