

**CRIAÇÃO DO CODIC - *CORPUS* ORAL DE DIVULGAÇÃO CIENTÍFICA:
CONSIDERAÇÕES LINGUÍSTICAS E METODOLÓGICAS**

**CREATION OF CODIC - ORAL *CORPUS* OF SCIENTIFIC DISSEMINATION:
LINGUISTIC AND METHODOLOGICAL CONSIDERATIONS**

Jackson Wilke da Cruz Souza¹

Universidade Federal da Bahia

Resumo: Diante da escassez de trabalhos que disponibilizam recursos para investigação linguística sobre a Divulgação Científica (DC) na modalidade oral, objetiva-se neste trabalho apresentar o *Corpus* Oral de Divulgação Científica (CODiC). O CODiC é composto por transcrições de respostas de especialistas (professores, técnicos e alunos) da Universidade Federal de Alfenas às perguntas enviadas pela comunidade externa ao projeto de extensão “A Voz da Ciência”, que foram veiculadas durante a programação de uma rádio local. Baseando-se nos princípios teóricos da Linguística de *corpus*, o CODiC constitui-se de transcrições das cinco primeiras temporadas temáticas do projeto, relativo ao período do segundo semestre de 2019 até o segundo semestre de 2021, somando um pouco mais de 4 horas de gravação, 32.094 tokens e 7.685 types. Apresentam-se caracterizações linguísticas e de metadados sobre a maneira como o conhecimento técnico-científico circula em espaços midiáticos abertos à DC. Ademais, são feitas reflexões sobre as mobilizações discursivas realizadas pelos interlocutores em determinado momento histórico, ideológico e cultural, o que auxilia a compreender o CODiC como um instrumento linguístico.

Palavras-chave: *Corpus* oral; Linguística de *corpus*; Divulgação Científica.

Abstract: Given the scarcity of works that provide resources for linguistic research on Scientific Dissemination (SD) in the oral modality, the objective of this work is to present the *Corpus* Oral de Divulgação Científica (CODiC). CODiC is made up of transcripts of responses from experts (teachers, technicians, and students) from the Federal University of Alfenas to questions sent by the external community to the “A Voz da Ciência” extension project, which were broadcast during local radio programming. Based on the theoretical principles of *Corpus* Linguistics, CODiC consists of transcriptions of the first five thematic seasons of the project, covering the period from the second half of 2019 to the second half of 2021, totaling a little more than 4 hours of recording, 32,094 tokens and 7,685 types. Linguistic and metadata characterizations are presented on the way in which technical-scientific knowledge circulates in media spaces open to SD. Furthermore, reflections are made on the discursive mobilizations carried out by the interlocutors in each historical, ideological, and cultural moment, which help to understand CODiC as a linguistic instrument.

Keywords: Oral *corpus*; *Corpus* linguistics; Scientific dissemination.

Submetido em 11 de abril de 2023.

Aprovado em 04 de setembro de 2023.

¹ Doutor em Linguística pela Universidade Federal de São Carlos Atua como docente na Universidade Federal da Bahia no Instituto de Ciência, Tecnologia e Inovação e no Programa de Pós-Graduação em Língua e Cultura. E-mail: jackcruzsouza@gmail.com.

Introdução

Devido a acontecimentos sociais, políticos e econômicos dos últimos anos, acompanhamos uma profusão de atividades de Divulgação Científica (DC) que contribuem para a construção de imaginários sobre *o fazer científico e quem o faz*. Nesse esteio, percebe-se que a DC pode ser considerada como importante e poderosa ferramenta de combate e resistência à *Fake News* (DANTAS; DECCACHE-MAIA, 2020), à desinformação (FREIRE, 2021) e ao negacionismo (CORREIA; MARTINS, 2022) que se espalham notoriamente na sociedade. Isso se deve ao fato de a DC acessar, em certa medida, esferas subjetivas da sociedade, permitindo com que haja deslocamentos de significados e concepções sobre ciência, cientistas/pesquisadores e das instituições de ensino e pesquisa envolvidas nesse processo (SOUZA, 2021).

Desde que centros de pesquisas e, em especial, universidades perceberam a importância de dialogar com a comunidade não especializada e externa acerca de suas atividades de pesquisa, tecnologia e inovação, bem como explicar conceitos e aspectos científicos a fim de torná-los inteligíveis por esse público, a DC tem ganhado notoriedade. A universidade, compreendida como um dos agentes de promoção da DC como prática (LIMA; GIORDAN, 2021), tem construído pontes de interlocução com esse público, mobilizando meios e estratégias de comunicação que tendem ir ao encontro do público-alvo em diferentes práticas de divulgação.

Essas estratégias comunicativas podem se dar em registros linguísticos escrito, oral e multimodal. Em cada um deles adotam-se estratégias específicas para promover explicações, exemplificações, demonstrações e definições de metodologias e conceitos científicos, por exemplo, que sejam funcionais do ponto de vista linguístico-cognitivo, em determinados veículos de comunicação, como revistas, *podcasts* e vídeos curtos.

Apesar da vasta utilização das redes sociais, que implicam a utilização de recursos e estratégias que demandam aspectos multimodais, uma das formas ainda democráticas e das mais antigas de promover a DC é o rádio. Esse aspecto mais democrático ocorre pelo alcance que se pode obter quando comparado aos meios e veículos de comunicação que são promovidos pela escrita, especialmente em programações que atinjam comunidades bem específicas.

Com relação aos estudos e pesquisas sobre produção e circulação de textos de DC, há uma vasta literatura que investiga as operações linguísticas utilizadas por divulgadores em textos escritos para promover a interlocução com a comunidade não especializada.

Cataldi (2008), ao analisar a mídia impressa, considerou três grandes grupos de operações linguísticas que ocorrem em textos escritos de DC, a saber: (i) *Expansão*, que se trata de inclusão de informações, com o propósito de disponibilizar significados conceituais para efetuar participação cognitiva e comunicativa do leitor; (ii) *Redução*, que se trata de estratégias de condensação ou de eliminação de informações científicas, mas mantendo-se conceitos relevantes; e (iii) *Variação*, que são procedimentos que se baseiam nos níveis lexical, semântico para promover a reformulação discursiva do texto.

Salienta-se que o registro escrito, ao ser utilizado em determinados veículos de comunicação, possibilita a utilização de certos recursos estruturais para promover maior compreensão do conteúdo, como notas de rodapé e/ou explicativas, imagens e infográficos, por exemplo. Essas possibilidades se enquadram na proposta de Cataldi (2008), ao conceber que uma nota de rodapé pode promover *Expansão* sobre o que está sendo dito, ou mesmo indicar um material suplementar de maneira a não comprometer ou interromper a narrativa científica.

Entretanto, quanto ao registro oral, percebe-se que há poucos estudos que investigam como o texto e o discurso de DC se estruturam e se realizam, e como eles mobilizam seus interlocutores no processo de letramento científico e inserção de um público não especializado na cultura científica. Esses estudos podem partir da caracterização de usos, operações e construções linguísticas específicas, que passam a ter funções determinadas na produção de sentido em um conjunto de dados linguísticos concebido especialmente para essa finalidade, como propõe a Linguística de *Corpus* (LC).

Nesse sentido, tem-se um instrumento linguístico que pode ficar limitado à descrição ou representação da atividade linguística de determinados grupos de falantes que o *corpus* representa. Entretanto, sabendo-se que (i) a construção de um *corpus* linguístico não é a simples organização de textos (orais, escritos e/ou multimodais), mas que evidencia certo objetivo de pesquisa, e (ii) é possível acessar dimensões políticas e históricas, demonstrando que não há neutralidade nos enunciados que compõem os textos, pode-se afirmar o *Corpus* Oral de Divulgação Científica (CODiC) pode ser considerado um instrumento linguístico.

Assim, objetiva-se neste trabalho destacar os processos metodológicos de criação do CODiC, entendendo-o como um instrumento linguístico que possibilita investigação e caracterização de usos técnicos e estratégias linguísticas para a popularização de conceitos e metodologias científicas. Para tanto, são descritas características linguísticas

e de metadados que elucidam como os enunciados linguísticos mobilizam sentidos a partir de um recorte histórico, político e ideológico.

Seguindo os princípios metodológicos da LC, em que se prevê representatividade, naturalidade e autenticidade aos dados linguísticos (SARDINHA, 2000), espera-se contribuir com a própria área e a Linguística descritiva, publicizando um instrumento e recurso linguístico-computacional que permitirá investigar diferentes fenômenos sobre a linguagem no âmbito da Enunciação, Sintaxe, Semântica e Sociolinguística em determinadas situações comunicativas, e perceber como interlocutores produzem, circulam e recebem a DC como discurso e materialidade.

Este artigo está organizado em outras quatro seções, além desta Introdução. Na seção 2, apresentam-se os trabalhos relacionados a esta pesquisa sob a forma de Revisão da literatura. Na Seção 3, apresenta-se a metodologia deste trabalho, equacionada em tarefas executadas ao longo do projeto de pesquisa. Na Seção 4, são apresentados e discutidos os resultados a partir do *corpus* de maneira quantitativa e qualitativa. Por fim, na Seção 5, tecem-se considerações finais, além de apontar para trabalhos futuros.

1. Recuperando Alguns Conceitos e Procedimentos em Linguística de *Corpus*

Para o desenvolvimento do CODiC, esta pesquisa se baseou nos princípios teóricos da LC. Por conta disso, destaca-se a importância de recuperar conceitos importantes à área de LC, seguindo a proposta de Biber (1993) e Biber, Conrad e Reppen (1998), que incluem o registro escrito e oral na noção de “textos” que compõem o *corpus* linguístico, e Berber Sardinha (2004).

Biber (1993), propõe que algumas etapas são necessárias para a construção de um *corpus* linguístico, a saber: (i) planejamento, (ii) coleta e (iii) armazenamento dos textos. Em (i), definem-se os parâmetros situacionais que variam entre os textos de diferentes comunidades discursivas, além de ser o momento de definir quais os fenômenos linguísticos que serão investigados. Como resultado dessas decisões, parte-se para a coleta dos textos e decide-se o formato de armazenamento deles.

Biber (1993, p.12) chama à atenção sobre a necessidade de *representatividade*, a despeito de *corpora* eletrônicos, que oferecem “uma base empírica sólida para ferramentas e descrições linguísticas de uso geral, permitindo a análise de uma dimensão que, de outra forma, não seria possível”. Assim, quando se pensa em “representatividade” é necessário levar em conta que o conjunto de textos deve ser adequado para promover generalizações acerca dos fenômenos linguísticos investigados.

Apesar de Berber Sardinha (2004) salientar que não existam critérios objetivos para determinar a representatividade do *corpus*, ele destaca que a amostra deve ter uma dada extensão. Esse procedimento tenta garantir que o corpus represente nele fenômenos linguísticos ou aspectos outros que estejam no horizonte da pesquisa. Para tanto, o autor destaca que determinar a população-alvo e os fenômenos linguísticos a serem estudados são estratégias relevantes para contornar a difícil tarefa de definir a extensão do *corpus*, sempre norteando-se pelas perguntas “*representativo do quê*” e “*representativo para quem*”.

Para que o *corpus* represente a variedade linguística de determinada construção fonética-fonológica, por exemplo, é importante que ele contenha diferentes variantes e que tenham sido produzidas naturalmente (em contraposição a uma criação artificial dos sons neste caso). Assim, é possível dizer que quanto mais raro é um fenômeno, maior deverá ser o *corpus* para poder representá-lo; ao passo que, quanto mais comum o fenômeno em observação, menor será o *corpus*. Nesse sentido, há uma relação direta entre a representatividade e o tamanho do *corpus*, considerando os objetivos de pesquisa que conduziram a construção daquela coletânea de textos.

Ainda, Biber (1993), na mesma perspectiva, contrasta a extensão do *corpus* (em quantidade de textos e de palavras, por exemplo) à definição de uma população-alvo, com relação às decisões sobre os métodos de amostragem. Levar em conta apenas características estatísticas pode conduzir à seleção de textos que conduzem análises equivocadas, como deixar de catalogar determinados fenômenos porque ocorrem em baixa frequência ou não ocorrem naquele conjunto de textos compilados.

Diante disso, Berber Sardinha (2004) propõe três pontos que o *corpus* precisa atender para responder essas duas últimas questões: (i) *Tipologia*, em que se detalham aspectos do *design* do *corpus* e as formas como ele foi coletado; (ii) *Representatividade* e (iii) *Extensão*, que juntos apontam para o tamanho do conjunto de dados, partindo do princípio que a língua é um sistema probabilístico (HALLIDAY, 1991) e, portanto, quanto maior o *corpus*, maior é a chance de descrever e encontrar fenômenos linguísticos que estão sob foco dos estudos.

Biber, Conrad e Reppen (1998, p.4) caracterizam os estudos realizados em LC como (i) *empíricos*, que permitem análise de padrões reais de uso em textos naturais; (ii) *extensos*, que é uma coleção grande de textos coletados criteriosamente; (iii) apresentar análises metodológicas, com utilização de abordagens automáticas, semiautomáticas e/ou manuais, além de análises qualitativas e quantitativas.

Por fim, cabe destacar que um *corpus* linguístico não representa apenas fenômenos linguísticos isolados de seus falantes, já que a naturalidade é um aspecto imprescindível em LC. Os *corpora* linguísticos, antes de mais nada, representam decisões baseadas em objetivos de pesquisa, o que justifica o tamanho esforço para que haja representatividade sobre a língua ou alguma de suas modalidades. Pontua-se, então, que quaisquer que sejam os métodos de compilação, os textos não são escolhidos ao acaso, pois há, ainda que pouco, enviesamento sobre essas escolhas. Mesmo em sua dimensão técnica, é possível que os *corpora* linguísticos reflitam uma dimensão político-histórica, no sentido de não haver neutralidade e não haver razão de ser feito, conforme Aquino (2020) aponta sobre a caracterização dos instrumentos linguísticos.

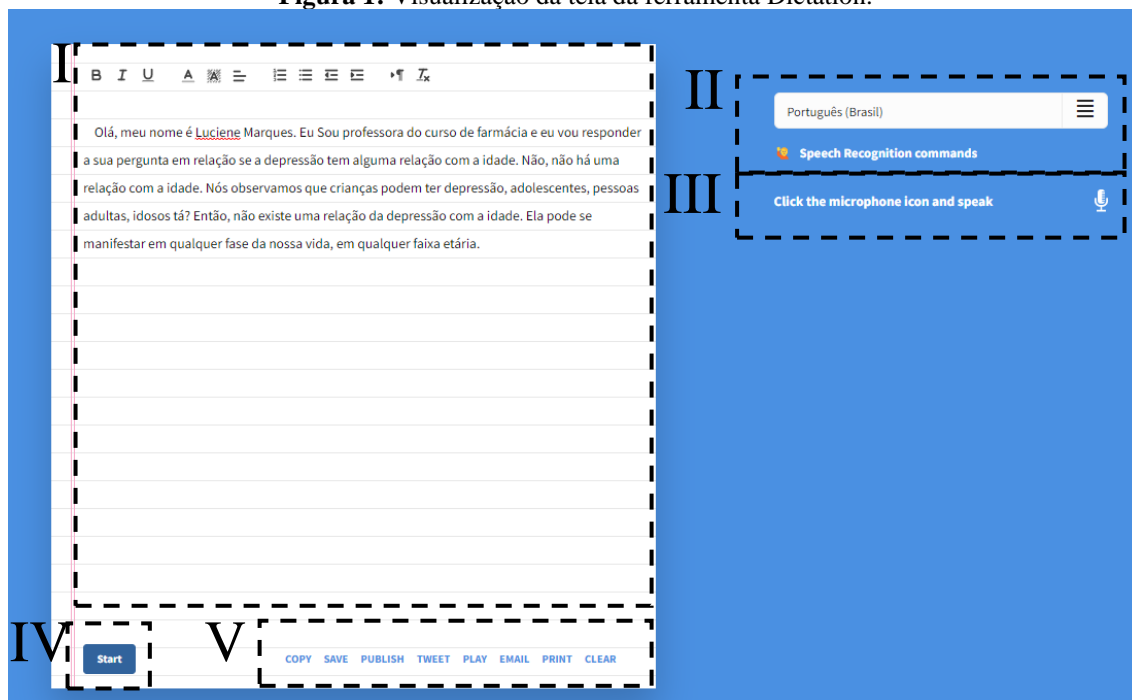
Além disso, os textos que são compilados que fazem parte do *corpus* são recuperados a partir de determinadas fontes, em dado recorte temporal e produzidos naturalmente por certos falantes. Assim, os *corpora* auxiliam não apenas sobre o conhecimento da língua e dos usos linguísticos dos falantes, mas também permite acessar ideologias e pensamentos dominantes e marginais, percepções sobre o mundo em dado momento histórico e cultural. Por conta disso, é possível considerar que o *corpus* contribui aos estudos linguísticos sendo também um objeto cultural e técnico-histórico, com pontua Medeiros (2020) ao analisar glossários.

2. Metodologia

A priori, a criação do CODiC está alinhada à investigação de fenômenos inerentes à Sintaxe, à Semântica e à Enunciação. Nesse sentido, as decisões metodológicas de transcrição dos áudios, anotação de certos fenômenos linguísticos e formatos de armazenamento dos arquivos estão subjacentes a esses níveis de análise linguística.

Quanto à *transcrição*, foi utilizada a ferramenta on-line e gratuita Dictation². Essa ferramenta possui extensão para diversos idiomas, além de permitir que qualquer formato de áudio possa ser processado e que a transcrição seja exportada para diferentes plataformas (como redes sociais e E-mail), como demonstrado na Figura 1.

² Disponível em: <https://dictation.io/>

Figura 1: Visualização da tela da ferramenta Dictation.

Fonte: Elaboração própria.

Como demonstrado na Figura 1, tem-se a tela de apresentação da ferramenta Dictation, organizada em cinco seções. Em (I), tem-se a área de transcrição do áudio; nela, a partir do reconhecimento do áudio, a ferramenta imprime na tela o texto compreendido, permitindo que o usuário também faça correções e/ou inserções na transcrição, bem como edições (alinhamento, destaques, paragrafação etc.). Em (II), o usuário deve selecionar qual o idioma do áudio que será transcrito. Em (III), deve ser selecionado o tipo de microfone que será utilizado para reconhecer o áudio: o usuário pode optar entre um microfone interno do computador, que pode reconhecer qualquer formato de áudio, ou o microfone externo, para transcrever o áudio dito pelo usuário. Em (IV), tem-se o botão “Start” que permite iniciar, pausar e interromper a transcrição. Por fim, em (V), têm-se as opções de exportar a transcrição, admitindo diferentes formatos e extensões.

Para esta pesquisa, foi utilizado o microfone interno do computador, sendo que a transcrição foi feita com base no áudio que está disponível nas plataformas de *streaming* do projeto de extensão “A voz da ciência”. As transcrições foram feitas a partir de respostas elaboradas por pesquisadores ou especialistas de determinados assuntos a perguntas enviadas pelo público externo à universidade. As respostas foram veiculadas inicialmente na Rádio Federal de Alfenas durante sua grade de programação e,

posteriormente, armazenadas em uma plataforma de *streaming*, de onde se extraíram os áudios. Por não ter sido utilizado o áudio original das respostas, a ferramenta Dictation precisou lidar com ruídos do áudio original e com falas concomitantes a músicas de fundo. Como resultado, foi necessário incluir uma fase de revisão para corrigir equívocos do processo automático de transcrição, como demonstrado em (1).

(1)

- a) *Olá meu nome é Luciene Marques Eu Sou professora do curso de farmácia e eu vou responder a sua pergunta em relação se a depressão tem alguma relação com a idade são com a idade observamos que crianças podem ter depressão adolescentes pessoas adultas idosos então não existe uma relação da depressão com a idade se manifestar fase da nossa vida em qualquer faixa etária.*
- b) *Olá, meu nome é Luciene Marques. Eu Sou professora do curso de farmácia e eu vou responder a sua pergunta em relação se a depressão tem alguma relação com a idade. Não, não há uma relação com a idade. Nós observamos que crianças podem ter depressão, adolescentes, pessoas adultas, idosos tá? Então, não existe uma relação da depressão com a idade. Ela pode se manifestar em qualquer fase da nossa vida, em qualquer faixa etária.*

Em (1), tem-se um exemplo de uma transcrição do *corpus* cuja temporada discutida foi Saúde mental. Em (1a), tem-se a transcrição da resposta como o *software* realizou a tarefa. Percebe-se diversos desvios sintáticos, caracterizados pela ausência de determinadas palavras na construção frasal. Assim, após a revisão manual, a transcrição que consta no *corpus* é a versão exemplificada em (1b).

Nesse processo de revisão dos textos, mantiveram-se repetições de palavras e sons (como “*remédio remédio*” e “*eh eh eh*”), repetição de sons iniciais de palavras (como “*saúde me mental*”) e desvios da norma padrão da língua (como “*foi feito muitos procedimentos*”), a fim de garantir que a transcrição, apesar da revisão, pudesse representar textos naturais, como proposto pela LC.

Além disso, foram tomadas cinco outras decisões metodológicas sobre o processo de transcrição:

- Diante de sons vocálicos, optou-se por acrescentar a consoante ‘h’, resultando em “ah”, “eh”, “ih”, “oh” e “uh”, para não provocar ambiguidade com conectivos ou repetições silábicas nos áudios;

- Os números foram transcritos por extenso e nunca em algarismos, como “um” e “dois mil e dezenove”;
- As siglas foram sempre transcritas em caixa-alta, como “UNIFAL” e “AIDS”;
- Ao que se refere à pontuação, os pontos-finais eram acrescentados automaticamente quando a ferramenta Dictation identificava a finalização de enunciados; quando isso não ocorria, pontos de finalização (ponto-final e interrogação) e vírgulas foram acrescentadas durante a fase de revisão;
- As transcrições não apresentam nenhuma segmentação em unidades de sentido ou turnos, já que se tratam de um tipo de monólogos (provocados por perguntas acessadas pelos especialistas e/ou divulgadores científicos posteriormente ao ato de fala), e pela transcrição não apresentar informações prosódicas que poderiam auxiliar nesse processo.

Vale destacar que essas duas últimas decisões descritas, para além das justificativas linguísticas e de projeto já apresentadas, baseiam-se no uso posterior de ferramentas de processamento automático de texto. Assim, essas decisões poderiam auxiliar o processo de busca de automáticas por padrões linguísticos, como a estrutura de apresentação de conceitos científicos diante de determinados assuntos, por exemplo.

Quanto ao *formato de disponibilização do corpus*, elegeu-se o formato .xml (do inglês, “*Extensible Markup Language*”). Esse formato permite a inserção de dados (como marcadores linguísticos específicos) e metadados (como informação de estruturação e duração das respostas) ao longo do *corpus*, como demonstrado na Figura 2. Além disso, esse formato facilita tanto a consulta quanto o processamento manual e/ou automático das informações.

Figura 2: Exemplo de anotação xml do CODiC.

```

<CODIGO_T1-A11>
  <PERGUNTA_EDITADA>Por que o vírus não é capaz de entrar pela barreira da pele?</PERGUNTA_EDITADA>
  <RESPOSTA>
    <TEXTO>
      <APRESENTACAO>Olá, meu nome é Bárbara Ávila, eu sou médica dermatologista, professora da Faculdade de Medicina da UNIFAL.</APRESENTACAO>
      <CONTEUDO>O que hoje nós sabemos é que o coronavirus ele apresenta diversas proteínas. Uma delas é responsável pela sua ligação no receptor de enzima conversora de angiotensina, o que faz com que o vírus seja capaz de penetrar nas células. Esse receptor não está presente nas camadas da pele, por isso o vírus não é capaz de penetrar a barreira da pele.</CONTEUDO>
    </TEXTO>
    <AUDIO>
      <INICIO_DA_RESPOSTA>0:00:37</INICIO_DA_RESPOSTA>
      <FINAL_DA_RESPOSTA>0:01:08</FINAL_DA_RESPOSTA>
      <DURACAO_DA_RESPOSTA>0:00:31</DURACAO_DA_RESPOSTA>
      <DURACAO_DO_AUDIO_COMPLETO>0:01:41</DURACAO_DO_AUDIO_COMPLETO>
      <LINK>https://open.spotify.com/episode/2t6RPNlBI2g5YDplv2u1lw?si=7149ff8a41ef4598</LINK>
    </AUDIO>
  </RESPOSTA>
  <RESPONDENTE>
    <NOME>Bárbara Ávila</NOME>
    <SEXO>Feminino</SEXO>
    <CURSO_DE_GRADUACAO>Medicina</CURSO_DE_GRADUACAO>
    <AREA_DO_CONHECIMENTO>Ciências da saúde</AREA_DO_CONHECIMENTO>
    <TITULACAO>Especialização</TITULACAO>
    <ATUACAO_NA_UNIFAL>Docente</ATUACAO_NA_UNIFAL>
    <UNIDADE_ACADEMICA>Faculdade de Medicina [FAMED]</UNIDADE_ACADEMICA>
  </RESPONDENTE>
</CODIGO_T1-A11>

```

Fonte: Elaboração própria.

O .xml é um formato de linguagem de anotação que permite realizar marcações especiais no texto por meio de *tags* entre colchetes angulares que delimitam o início e o final da informação marcada. Na Figura 2, por exemplo, tem-se as *tags* <RESPOSTA> e </RESPOSTA> que abrigam hierarquicamente outras *tags*, a saber: texto e áudio. Em cada uma dessas outras duas *tags* há delimitações de outras informações, como trechos de apresentação dos respondentes, a resposta elaborada por eles, a duração e o link dos áudios completos, por exemplo. Salienta-se que esse formato de linguagem de anotação é compatível com editores de texto (como o Notepad ++) e interfaces gráficas para anotação de *corpus* que processem essa extensão de arquivo. O CODiC está disponível em um repositório eletrônico ³sob a licença *Creative Commons*.

3. Resultados e Discussão

Nesta seção estão detalhados resultados desta pesquisa, organizados em Caracterização do *corpus*, Organização das informações, Considerações sobre o perfil dos respondentes e Considerações sobre a oralidade no *corpus*.

3.1. Caracterização do *corpus*

A caracterização do *corpus*, em certa medida, reflete decisões do projeto de extensão do qual os áudios foram transcritos. A equipe executora do projeto de extensão elege um tema que julgue ser atual e relevante à comunidade externa, a qual pode enviar

³ Disponível em: <https://github.com/jackcruzsouza/CODiC>.

suas dúvidas por meio de redes sociais⁴. Em seguida, as perguntas são encaminhadas a possíveis especialistas e/ou divulgadores científicos da universidade (professores, técnicos e/ou alunos) que atuarão como respondentes. Eles são instruídos a produzirem respostas curtas com linguagem acessível pelo público não especializado (aspecto que será retomado na subseção 3.3).

Para elaboração do CODiC, foram consideradas as cinco primeiras temporadas do projeto com as seguintes temáticas “Covid-19”, “Atemático”, “Saúde mental”, “Uso Racional de medicamentos” e “LGBTQIAP+”. De acordo com a Tabela 1, a quantidade de áudios transcritos referentes a cada uma das temporadas decaiu ao longo das temporadas. As transcrições foram realizadas a partir dos áudios disponíveis na plataforma *streaming* do projeto, e não constam todos os áudios que foram veiculados na programação da Rádio Federal de Alfenas, especialmente para as temporadas 4 e 5.

Tabela 1: Caracterização do CODiC.

Temporada/ Assunto	Qnt. de áudios	Qnt. de <i>tokens</i>	Qnt. de <i>types</i>	Duração média das respostas	Duração total das respostas
T1: Covid-19	44	8.779	1.876	00:01:30	01:05:50
T2: Atemática	40	8.407	1.908	00:01:33	01:02:05
T3: Saúde mental	35	7.053	1.599	00:01:37	00:56:28
T4: Uso racional de medicamentos	23	3.711	1.080	00:01:09	00:26:31
T5: LGBTQIAP+	24	4.144	1.222	00:01:16	00:30:18
TOTAL	166	32.094	7.685	00:01:25	04:01:12

Fonte: Elaboração própria.

Com relação à Temporada 1, destaca-se que essa temporada foi marcada por estar sendo produzida concomitante à pandemia de Covid-19. Nesse sentido, as respostas elaboradas refletem certas preocupações técnicas, informativas e científicas que se tinham à época da temporada, como, por exemplo, a caracterização de sintomas, material de máscaras e material de assepsia, como demonstrado em (2). Ainda que a informação em si não seja mais adequada, dado que a instrução de utilizar máscaras de tecido não sejam mais indicadas por órgãos fiscalizadores de saúde, a resposta foi mantida no *corpus* por possibilitar o mapeamento da preocupação tanto da comunidade externa (no sentido de levantar pontos de questionamento) quanto da universidade (no sentido de tornar acessível a informação fidedigna à realidade) criarem interlocução sobre esse ponto.

⁴ Disponível em: <https://www.instagram.com/avozdaciencia/>.

Outra justificativa para a manutenção de respostas como (2) no *corpus* é o fato de a resposta possibilitar análises de estratégias linguísticas utilizadas no discurso oral para a popularização científica, para além da questão temporal levantada anteriormente.

(2)

Pergunta: Como devo higienizar a minha máscara de tecido?

Resposta: Olá, meu nome é Leonardo Almeida, sou professor da UNIFAL. As máscaras de pano, elas podem ser lavadas com água e sabão, assim como elas também podem ser deixadas de molho em água sanitária por aproximadamente meia hora. Depois disso, essas máscaras devem ser lavadas com água e sabão e também podem ser penduradas para secarem naturalmente. Quando elas estiverem secas, as máscaras devem ser passadas a ferro quente antes de serem utilizadas novamente. É importante lembrar que as máscaras são de uso pessoal, não podendo ser compartilhada com outras pessoas, e somente as máscaras de pano são reutilizáveis, são laváveis. As máscaras descartáveis só são utilizadas uma única vez.

A Temporada 2 foi tida como atemática pois foi a única em que o projeto se abriu à comunidade externa para responder quaisquer perguntas enviadas, e não a partir da decisão de um tema específico. Aqui estão contempladas respostas com relação à atuação e disponibilidade de serviços da universidade no município mineiro de Alfenas, importância de certos animais aos seres humanos (como abelhas e peixes), uso de medicamentos e como lidar com casos de suicídio.

Na Temporada 3 contemplou-se a discussão sobre saúde mental, tendo como respondentes profissionais da área da saúde. Nessa temporada, percebe-se que há duas naturezas de respostas, a depender da formação acadêmica dos respondentes: se eram de áreas que propiciavam maior discussão com a temática e seus subtemas (como as áreas de Psicologia, Medicina e Enfermagem), as respostas apresentam maior detalhamento informativo; se os respondentes eram de áreas afins, (como Odontologia), as respostas foram precisas, apesar de concisas, destacando-se o aspecto de interlocução com o público ouvinte (como “*Não desista da psicoterapia*”).

A Temporada 4 também foi executada durante a pandemia de Covid-19, concomitante à polêmica em torno do uso de medicamentos que não tinham eficácia comprovada cientificamente para combater a doença. Na tentativa de produzir mais adesão do público ouvinte, decidiu-se não nomear a temporada com os medicamentos questionados à época, mas tornar a discussão mais abrangente. Como resultado, as perguntas enviadas contemplaram tais medicamentos, mas também outros mais comuns

(como dipirona e anti-inflamatório, por exemplo), bem como o uso devido de cada um deles (quanto à dosagem e ao vencimento, por exemplo).

Por fim, na Temporada 5 discutiram-se temas relacionados à comunidade LGBTQIAP+. Ainda que o foco deste estudo esteja nas respostas elaboradas, é interessante perceber que a maioria das perguntas enviadas ao projeto foram anônimas, diferentemente das temporadas 1, 2 e 4, em que, possivelmente, explica-se devido à própria temática da temporada. Essa é a única temporada que escapa à área da saúde no CODiC, possibilitando verificar construções frasais e estratégias linguísticas que advêm de formações e experiências acadêmico-profissionais distintas. Ademais, é possível detectar menor interlocução nas respostas, ao passo que há maior indicação de materiais complementares e fontes de mais informações (como sites institucionais e organizações civis municipais).

Quanto à *duração das respostas*, os respondentes eram instruídos a produzirem textos com duração máxima de 2 minutos, já que a grade de programação da rádio disponibilizou 3 minutos de inserção midiática. Assim, as respostas elaboradas tiveram duração média total de 00:01:25, indicando capacidade de síntese dos respondentes em relação ao conteúdo elaborado em formato de DC.

Quanto à *análise estatística* do CODiC, utilizou-se a ferramenta AntConc (ANTHONY, 2005), a qual possibilitou a contabilização de sequência de caracteres separados por espaços em branco (*tokens*) e a quantidade dessas sequências sem repetição (*types*). A diferença entre *tokens* e *types* permite inferir certa novidade linguística que os respondentes empregam em seus textos. Observou-se que a variabilidade linguística das respostas é bastante pequena devido à relação numérica entre *tokens* e *types* (no caso, 24.409), sendo possivelmente explicado por usos linguísticos mais simplificados em função da intenção discursiva e argumentativa, do público-alvo e do suporte de comunicação.

3.2. Organização das informações

Como visto, o *corpus* está disponibilizado em formato .xml, o qual permite segmentar trechos de informação para recuperação e análise manual e automática mais facilmente por meio de *tags*. As *tags* utilizadas no *corpus* dizem respeito à Pergunta editada, Resposta e Respondente.

a) *Tag “Pergunta editada”*

A *tag* “Pergunta editada” compreende a pergunta enviada ao projeto de extensão pela comunidade externa à universidade. Para ir ao ar, a pergunta pode ter passado por algum processo de edição técnica (para remover silêncios e repetições de palavras, por exemplo) ou textual (para substituir palavras e/ou expressões preconceituosas, ou retirar apreciações do público sobre o projeto). Visando a atender às exigências da Lei Geral de Proteção de Dados (LGPD), todas as perguntas do *corpus* disponibilizado foram anonimizadas, ou seja, as informações relativas a nome, idade e/ou localização foram removidas da transcrição.

b) *Tag “Resposta”*

Na *tag* “Resposta”, armazenam-se duas macro-categorias, a saber: “resposta” e “áudio”. Em “resposta”, dividiu-se em “apresentação” e “conteúdo”, que foram elaborados pelos respondentes. Essa divisão se deu por observar a roteirização no trecho compreendido em “apresentação”, já que os respondentes eram instruídos a dizerem seus nomes, formação acadêmica e unidade acadêmica que atuavam na UNIFAL-MG. Em “conteúdo” consta a resposta elaborada, em que a *tag* foi inserida não seguindo padrões sintáticos, mas onde, de fato, ocorreram, como demonstrado na Figura 3.

Figura 3: Exemplo de uso de *tags* no CODiC.

```
<CODIGO_T1-A6>
<PERGUNTA_EDITADA>Se eu tiver roupas velhas em casa, como moletom e camisas, eu posso utilizá-las
pra fazer as máscaras? Qual o perigo disso? E qual a eficiência dessa máscara?</PERGUNTA_EDITADA>
<RESPOSTA>
<TEXTO>
<APRESENTACAO>Bom dia, eu sou o professor Lucas Lopardi Franco. Eu sou Professor na
faculdade de ciências farmacêuticas da Universidade Federal de Alfenas</APRESENTACAO>
<CONTEUDO>e eu vou responder agora essa pergunta sobre a eficiência das máscaras
artesanais, se elas são confiáveis. E sim elas são confiáveis. A gente hoje tem um bom
respaldo de estudos que mostram a confiabilidade desses tecidos, mas é importante olhar o
tipo de material que você tá usando. Existe um tipo de tecido, por exemplo, que não é de
grande capacidade de retenção que à base de TNT. Então, se você tá usando como máscara
artesanal uma camada de TNT a segurança não é muito grande, mas se você usa, por exemplo,
o algodão 2 ou 3 camadas você já tem uma grande eficiência sobre isso. Então, você pode
fazer essas máscaras artesanais, você tem que avaliar se elas têm mais de uma camada de
tecido e o tipo de tecido que ele é feito. É mais comum hoje para facilitar a respiração
você, por exemplo, fazer dois tecidos de algodão e um de TNT neh 2 camadas de algodão e 1
de TNT interna você pode inclusive nessas máscaras que têm duas camadas de algodão externa
e interna ela tem um compartimento mesmo artesanal neh ela tem um compartimento pra você
trocar o do meio e aí você pode, por exemplo, até usar o filtro de papel que já foi
mostrado uma alta eficiência de retenção. Então uma saída também é você usar essas
máscaras de tecido artesanais que tenham um um espaço interno pra você trocar pra você
pode usar um filtro de papel para aumentar a segurança, um de TNT pro dia a dia pra
facilitar a respi a respiração. Então sim: pode usar máscara artesanal, mas tem que tomar
cuidado com o tipo de material. Se você quiser acompanhar mais o projeto nas mídias
sociais, a gente mostrou lá os materiais mais seguros, as camadas, tem tudo mais detalhado
lá. Então sim: pode usar a máscara artesanal.</CONTEUDO>
```

Fonte: Elaboração própria.

Como visto, um dos pressupostos da LC é observar os fenômenos linguísticos em seus contextos naturais de ocorrência. Nesse sentido, julgou-se ser necessária essa divisão no interior da resposta para que pudessem ser recuperados mais facilmente os trechos em

que houvesse certa espontaneidade na resposta elaborada, para que fenômenos ligados à oralidade pudessem ser observados, além de ser possível manipular e/ou recuperar trechos específicos: do “conteúdo”, da “apresentação” do respondente, ou do “texto” da pergunta.

Ainda sobre a resposta, na *tag* “Áudio” compreende-se a caracterização numérica dos dados sobre a resposta e o áudio completo (com introdução, pergunta, resposta e agradecimentos), além do endereço eletrônico do áudio completo na plataforma de *streaming* do projeto.

c) *Tag “Respondente”*

Nesta *tag* estão compreendidas as informações sobre os respondentes, a saber: Nome, Sexo, Curso de graduação, Área de conhecimento, Titulação, Atuação na UNIFAL-MG e Unidade acadêmica. Todas essas informações foram extraídas a partir do Portal de Dados Abertos da UNIFAL-MG⁵ (que armazena informações acerca de seus servidores) e da Plataforma Lattes (a respeito das informações acadêmicas).

3.3. Considerações sobre o perfil dos respondentes

É importante destacar que a produção textual em DC depende da mobilização não apenas de conceitos científicos, mas também da adoção de estratégias discursivas para promover interlocução e compreensão do conteúdo com o público-alvo. Sabe-se ainda, o quão difícil é transpor os muros acadêmicos, já que isso exige mais que o domínio de uma linguagem tida como acessível, mas também fazer com que as pesquisas científicas acolham agendas sociais e culturais, para que a comunidade externa compreenda e apoie a necessidade de pesquisas científicas.

O divulgador científico pode, por vezes, ser um cientista ou pesquisador acadêmico, que, no imaginário brasileiro, são (a) pessoas inteligentes que fazem coisas úteis à humanidade, (b) pessoas comuns com treinamento especial, (c) pessoas que servem a interesses econômicos e produzem conhecimento em áreas nem sempre desejáveis, (d) pessoas que se interessam por temas distantes das realidades das pessoas ou ainda (e) pessoas excêntricas de fala complicada (CGE, 2019).

Compreender o perfil dos divulgadores científicos permite-nos repensar nas estratégias linguísticas e multimodais de reformulação discursiva (FINATTO; EVERS; STEPHANI, 2016) em determinado registro da língua, bem como avaliar a pertinência

⁵ Disponível em: <https://sistemas.unifal-mg.edu.br/app/si3/home.php>.

da fonte de informação utilizada (CUNHA; GIORDAN, 2015). Como resultado, ter-se-ão maneiras populares de tornar mais acessível e compreensível o conhecimento que se é produzido na Universidade, a partir de dimensões objetivas e subjetivas da DC (SOUZA, 2021).

Tabela 2: Caracterização de perfil dos respondentes no CODiC.

SEXO	TEMPORADA					ATUAÇÃO			TITULAÇÃO						ÁREA DO CONHECIMENTO						
	T1	T2	T3	T4	T5	Aluno	Técnico	Docente	Graduação	Graduando	Especialização	Mestrando	Mestrando	Doutorando	Doutorando	Artes	Ciências da Saúde	Ciências da Saúde	Ciências da Saúde	Humanas	Sociais
Fem.	8	16	6	12	14	4	6	44	1	1	5	0	1	3	40	0	8	35	5	2	3
Masc.	14	12	4	9	7	3	5	37	0	1	6	4	2	0	32	1	10	26	4	0	4

Fonte: Pocobello, Silva e Souza (2022).

De acordo com a Tabela 2, no projeto, há maior presença de especialistas do sexo feminino (56) quando se compara a pessoas do sexo masculino (46). Quanto à atuação na UNIFAL-MG, há participação maior de docentes (81), seguidos de técnicos (11) e alunos (7), sendo que a maioria possui o título de doutor/a (72). Com relação à área do conhecimento, percebe-se que há massiva presença de respondentes da área de Ciências da saúde (61), ao passo que as outras áreas do conhecimento estão sub-representadas, bastante explicado por conta dos temas abordados no CODiC e à natureza das perguntas enviadas ao projeto.

A partir dessas informações iniciais, Pocobello, Silva e Souza (2022) apontaram que o perfil de divulgadores do projeto de extensão “A voz da ciência” se concentra em indivíduos do sexo feminino, com doutorado e que atuam como docentes na UNIFAL-MG, predominantemente da área de Ciências da saúde. Tal perfil advém da correlação entre a temática das temporadas e a própria configuração do corpo de servidores e alunos

da referida universidade. Ademais, é possível conjecturar que a não articulação do projeto com outras áreas do conhecimento, por vezes, corrobora um estereótipo social acerca da concepção de Ciência e da figura do cientista: aquela deve sempre se conectar a áreas de Ciências exatas ou da Saúde, enquanto esta deve sempre estar enquadrada na imagem de uma pessoa em um laboratório.

Cabe sinalizar uma distinção terminológica e conceitual que está presente no CODiC. As pessoas que responderam às perguntas, até o momento, estão sendo tratadas como “respondentes” ou como “divulgadores”, mas não exclusivamente como “especialistas”. Para ser considerado esta última categoria, seria necessário que os respondentes possuíssem formação e atuação acadêmicas compatíveis às perguntas. Um jornalista, por exemplo, pode ser considerado como um divulgador ao abordar uma temática sobre ciência, por exemplo; porém, isso não o torna um especialista no assunto.

É possível perceber que, na Temporada 5, cuja temática foi LGBTQIAP+, por exemplo, a maioria dos respondentes tinham alguma ligação com a temática (constitutiva, participativa ou ideológica, por exemplo), sem que necessariamente pesquisassem questões voltadas ao tema. Uma das atualizações que o CODiC receberá em uma de suas próximas versões será o acréscimo da informação “especialista” (aquele que tem engajamento científico com a questão) ou “divulgador” (aquele que responde à pergunta a partir de fontes científicas sem que ele tenha engajamento científico com a questão) a cada um dos respondentes, para que análises mais específicas sobre os usos linguísticos e estratégias de DC possam ser estudadas.

3.4. Considerações sobre a oralidade no *corpus*

Por ser um *corpus* oral, o CODiC tem representatividade de traços e fenômenos linguísticos desse registro da língua, considerando que este molda-se às necessidades comunicativas dos interlocutores (no caso, respondentes e ouvintes), além das características formais do suporte midiático que publiciza os áudios.

Antes, é necessário refletir que o discurso de DC está engendrado em um complexo sistema de atividades de produção da DC, o que inclui mecanismos linguísticos, como apontam Lima e Giordan (2018).

Figura 4: Sistema de atividades de produção da DC.

Fonte: Lima e Giordan (2018).

Lima e Giordan (2018) concebem a DC dentro de um sistema de atividades complexas e interligadas, ora diretamente, ora indiretamente, conforme ilustrado na Figura 4. A Linguagem verbo-visual, em especial, é entendida como um sistema simbólico capaz de promover ilustração do tema e representação do cientista. Segundo os autores, o objeto da produção da DC é a cultura científica, tida como um “empreendimento humano produzido por meio de processos comunicacionais tendo em vista a comunicação e inserção de novos membros nessa cultura” (LIMA; GIORDAN, 2018, p. 501). Por conta disso, na representação da Figura 4, a Linguagem verbo-visual, conecta-se diretamente ao Divulgador, à Cultura Científica e a Comunidade de divulgadores para promover comunicação e inserção.

Porém, como demonstrado inicialmente, muitos recursos linguísticos conhecidos atualmente que são empregados nesse tipo de linguagem estão restritos ao registro escrito. Se por um lado ilustrações visuais e notas de rodapé explicativas não são possíveis de serem utilizadas, por outro, utilizam-se outras estratégias de explicação nos textos orais, como demonstra-se em (3)

(3)

Olá, meu nome é Ana Carolina Padovan, sou professora de microbiologia da UNIFAL. Para você desinfetar suas roupas e sapatos pense o seguinte: sabão e água são os nossos maiores aliados. Então a sua roupa você pode lavar da forma como você tá acostumada: pode colocar na máquina, pode colocar no tanquinho, pode lavar à mão. O importante é esfregar bem com sabão e água. Se você estiver usando uma máscara de pano, é importante que você deixa essa máscara de molho pelo menos por 30 minutos com sabão e água antes de lavar normalmente. Já o seu sapato dê preferência pra desinstalá-lo fora da sua casa. Você pode manter uma vasilha com uma solução de água 9 partes de água, uma parte de água sanitária pra você passar na solas do seus sapatos. Se o seu sapato é

delicado, você pode usar também água e sabão, sempre nas solas e de preferência manter o seu sapatos, mesmo que higienizados, fora da sua casa.

A resposta transcrita em (3) advém da primeira temporada, cuja temática foi Covid-19. Observa-se que a respondente faz uso de ao menos três estratégias para construir seu discurso: (i) *construção imagética*, em que a respondente utiliza a construção “pense o seguinte”, no intuito de conduzir o ouvinte a visualização da cena instrutiva que será construída no desenrolar da narrativa; (ii) *interlocução direta*, em que, de maneira explícita (como em “*se você estiver...*” e “*você pode lavar...*”), tenta promover proximidade com o ouvinte; e (iii) *sequenciação*, em que por meio de procedimentos seriados, promove a instrução de como poderia ser promovida a higienização de máscaras, roupas e sapatos, utilizando o tipo textual injuntivo, tido como comum e circulante na sociedade em outros gêneros textuais.

A *repetição* também pode ser tida como marca da oralidade na Linguagem verbo-visual. Dentro da sequência de instruções, a respondente repete a estrutura de condição (“*se o seu sapato...*”, por exemplo) e de modalização (“*pode lavar à mão...*”, bem como um conjunto de palavras (como “*água*”, “*sabão*”, “*água sanitária*”, “*máquina*”, “*tanquinho*”, “*lavar*” e “*esfregar*”, por exemplo), que se circunscrevem em um determinado campo semântico para tentar promover ênfase às instruções fornecidas.

Outro aspecto presente nos textos que constituem o CODiC é o planejamento e não-planejamento das respostas, como ilustrado em (4)

(4)

a) *Olá, eu sou Larissa Bueno. Médica de Família e professora do curso de Medicina da Universidade Federal de Alfenas. Muitas pessoas podem achar que as relações sexuais com o mesmo gênero são mais tranquilas em relação às doenças, mas a relação sexual homoafetiva transmite sim as DSTS, que nada mais são do que as doenças sexualmente transmissíveis, e as doenças transmitidas elas podem ser desde o HPV, até mesmo o HIV que é uma doença crônica, grave. Portanto o cuidado e a prevenção eles devem sempre existir.*

b) *Meu nome é Robson, sou professor aqui da Unifal, minha especialidade é nefrologia e vim responder a pergunta do José Alfredo aqui sobre a respeito do uso das medicações para diabetes. Então tomar dois comprimidos porque comeu mais doce, isso não vai funcionar muito bem porque os medicamentos, principalmente via orais, eles demoram a fazer o efeito, geralmente eles têm um tempo maior neh pra fazer efeito, diferente da insulina que age mais rapidamente. Porém, o remédio não vai substituir a insulina. Então, o medicamento, o comprimido. Então o ideal é que você não aumente, pelo contrário, que você tente reduzir as próximas refeições com eh... alimentos neh,*

reduza um pouco os alimentos pra tentar reduzir essa glicemia, e retornar no médico pra reavaliar se tem necessidade de aumentar as doses ou não dos medicamentos.

As respostas apresentadas em (4) advêm das Temporadas 5 e 4, respectivamente. Ao analisar (4b), percebe-se a utilização demasiada de marcadores discursivos de conclusão para introduzir ideias (como “*Então tomar dois comprimidos...*”), pausas (marcado por “...”), quebra de continuidade discursiva (como em “*Então, o medicamento, o comprimido. Então o ideal é que você não aumente...*”) e hesitações (evidenciado por “eh”), o que leva a concluir que o referido texto não apresenta planejamento. Já (4a) demonstra certo planejamento textual por conta do uso do conectivo de conclusão para finalizar uma ideia (em “*Portanto o cuidado e a prevenção...*”), e apresenta marcas da oralidade com a retomada do tema, em relação ao rema, pelo pronome “eles” em “[*o cuidado e a prevenção*] *eles* devem sempre existir”.

Outra consideração sobre a oralidade no CODiC é a construção de um interlocutor receptor da resposta. Como visto, a resposta elaborada é provocada por uma pergunta que é encaminhada ao projeto de extensão que, posteriormente, é veiculada na programação da Rádio. É interessante notar que quem envia as perguntas ao projeto, na interlocução assume um papel apenas de provocador da resposta, já que quem envia as perguntas pode não ser do município mineiro de Alfenas, como observado por Vilaça *et al.* (2022). Como resultado, observa-se que mesmo de maneira explícita, como em (3), a interlocução, na verdade, está sendo construída em função de uma representação do possível ouvinte da rádio.

Com relação a isso, há respostas no *corpus* em que há explicitude de quem envia a pergunta (como “*O NOME-DO-PERGUNTADOR, de LOCALIDADE-DO-PERGUNTADOR, gostaria de saber se os fumantes correm o maior risco de infecção pelo coronavírus...*”), do próprio ouvinte (como “*Então, o problema eh, querido ouvinte, não é a prática do sexo sem a camisinha e sim o contato muito próximo...*”), ou ainda dos dois simultaneamente (como em “*Caros NOME-DO-PERGUNTADOR e ouvintes...*”).

Ainda, percebe-se que em algumas respostas há movimentos enunciativos, que fazem com que os respondentes ora se comuniquem evocando quem envia a pergunta, ora com o ouvinte, como demonstrado em (5).

- (5)
a) Pergunta: *Sou gay, posso doar sangue?*

b) *Resposta:* *Olá, pessoal. Eu sou Iara, médica hematologista e professora da UNIFAL e a resposta é sim. Até que enfim rege o bom senso, paciência e a justiça. A partir de maio de 2020 o STF derrubou a cláusula que proibia homossexuais de doarem sangue. Até 2020 homossexuais não podiam doar sangue, isso é ridículo. Agora você pode procurar o banco de sangue mais próximo de você, independente de sua orientação sexual, e doar sangue desde que você tenha tido apenas um parceiro nos últimos seis meses. Diante disto, procure o banco de sangue, ajude o próximo. Espero que eu tenha respondido, deixando tudo isso muito claro e até uma próxima oportunidade. Tchau!*

Observa-se que a resposta em (5b) é elaborada a partir da pergunta enviada em (5a), da quinta temporada do projeto. A resposta inicia-se diretamente ao ouvinte que enviou a pergunta, seguida de informações evocadas pela respondente. Porém, mais ao final da resposta, a construção “*Agora você pode procurar o banco de sangue mais próximo de você, independente de sua orientação sexual(...)*” parece dirigir-se tanto ao perguntador quanto ao ouvinte da programação, tendo em vista que aquele explicitou sua orientação sexual, ao passo que a resposta evoca “independentemente de sua orientação sexual”. A resposta é finalizada sendo redirecionada ao perguntador, com o desejo de que a pergunta tenha sido respondida, de fato.

Por fim, cabe pontuar que, apesar de as transcrições do CODiC serem provenientes de uma plataforma de *streaming* e estarem armazenados em formato de *podcast*, os áudios poderão ser analisados a partir de concepções de suporte midiático sobre o rádio. Nas análises que se derivam, então, implica-se considerar tempo de duração dos áudios, construção da imagem do cientista, construção imagética de um público-alvo mais amplo e mecanismos linguísticos a partir da relação dialógica e interlocucional entre respondentes e (possíveis) ouvintes.

Considerações Finais

O objetivo deste trabalho foi pautar as etapas de criação do CODiC, como recurso linguístico que possibilita investigação e caracterização linguísticas de especialistas na popularização de conhecimento científico. Como resultado, mais que apresentar o *corpus*, foi destacada uma metodologia de coleta e análises iniciais de aspectos relevantes para a LC.

Ao longo deste artigo, apresentaram-se fragmentos do que os consulentes ao *corpus* poderão encontrar ao acessá-lo e baixá-lo. Antes, será necessário partir do princípio que os respondentes representados ao longo do *corpus* mantêm interlocução contínua e rotineira com seus pares especialistas, que valorizam e entendem o processo

de construção e constituição do conhecimento científico. Esse não é um processo fácil, como destaca Souza *et al.* (2022), pois

requer que o texto fuja aos jargões científicos e que, ao mesmo tempo, não caia no engano de ser um texto simples; requer que o autor ao explicar e pontuar fatos, teorias e pesquisas científicas não as simplifique a ponto de torná-las vulgar, simples ou inexistente; requer entendimento de como os sentidos circulam, se produzem e se (re)significam em uma sociedade hiper conectada; por fim, requer que o autor fuja à proposição de um leitor médio e parta do princípio que o óbvio nunca é tão óbvio (SOUZA *et al.*, 2022, p.6).

Nesse sentido, o consulente não terá à sua disposição apenas um recurso para analisar textualmente as construções frasais e mobilização de conceitos científicos em disputa, mas também a possibilidade de observar como interlocutores acadêmicos se deslocam discursivamente para se inserem em atividades comunicativas novas de mobilização de sentidos a fim de tornar mais acessível o conhecimento.

Vale ressaltar que, ao analisar as respostas produzidas pelos respondentes, o CODiC reflete um recorte histórico-ideológico peculiar e, infelizmente, expressivo na sociedade atual: desinformação generalizada, aumento de notícias falsas, pensamentos anti e pseudocientíficos. É necessário partir do princípio que a DC produz intersecções entre esferas de criação ideológicas, em que as atividades pleiteiam motivos, objetivos, diretrizes, agentes e ferramentas culturais (LIMA; GIORDAN, 2021), importantes no processo de resistência da Ciência. As respostas, nesse sentido, não foram produzidas ao acaso: elas foram provocadas a partir de perguntas que se ancoram em temáticas escolhidas pelo projeto de extensão “A voz da Ciência”. É possível dizer que as respostas fazem parte de um projeto discursivo do grupo extensionista, na tentativa de mobilizar sentidos e fomentar o conhecimento. Tais pontos também poderão ser explorados no *corpus*, pois ele foi tido na concepção de que a DC “se apresenta como uma ferramenta fundamental para repensar os caminhos que a ciência faz atualmente [...] que podem se afastar de uma classe elitizada e se aproximar de um leitor não especialista” (SOUZA *et al.* 2022 p.6).

Como alguns dos trabalhos futuros que serão desenvolvidos a partir do CODiC destacam-se: (i) acréscimo de meta informação sobre (não) espontaneidade nos textos; (ii) inserção de campos semânticos em cada uma das temporadas; (iii) anotação enunciativa, com vistas a identificar quando os respondentes alteram ou alternam seus alvos interlocucionais. Ademais, há outros trabalhos que servirão de ampliação da

representatividade do *corpus*, como acrescentar outras mídias que suportam o registro oral (como vídeos e podcasts transcritos, por exemplo), e aplicar notação de oralidade para que sejam identificados e analisados aspectos prosódicos, resultando em outras versões do CODiC.

Agradecimentos

Agradeço a todos alunos, professores e comunidade externa que estiveram envolvidos no projeto de extensão “A Voz da Ciência”, especialmente ao idealizador do projeto, Professor Dr. Lucas Lopardi Franco.

Referências

AQUINO, José Edicarlos de. Gramática: instrumento técnico/ferramenta político-histórica. In: MEDEIROS, Vanise; Esteves, Phellipe Marcel da. S. et al. (Org.). **Almanaque de Fragmentos: ecos do século XIX**. Campinas: Pontes, 2020, p. 113-118.

ANTHONY, L. AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom. In: IPCC 2005. **Proceedings**. International Professional Communication Conference, 2005. IEEE, 2005. p. 729-737.

BERBER SARDINHA, T. **Linguística de corpus**. Editora Manole Ltda: São Paulo, 2004.

BERBER SARDINHA, T. Linguística de *corpus*: histórico e problemática. **Delta**, v. 16, p. 323-367, 2000.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus linguistics: Investigating language structure and use**. Cambridge University Press, 1998.

BIBER, Douglas. Representativeness in *corpus* design. **Literary and linguistic computing**, [s.l.], v.8, n.4, p. 243-257, 1993.

BRASIL. Ministério da Ciência, Tecnologia, Inovações (MCTI). **Percepção pública da C&T no Brasil – 2019 – Resumo executivo**. Brasília: Centro de Gestão e Estudos Estratégicos, 2019.

CATALDI, C. A definição utilizada como estratégia divulgativa sobre transgênico na mídia impressa. **Vertentes**, São João Del-Rei/Brasil, v.1, n.32, 2008.

CORREIA, M.; MARTINS, M.C. O papel das atividades de divulgação científica na promoção da literacia científica. **Saber & Educar**, [s.l.], v.31, nº1, 2022.

CUNHA, M.B.; GIORDAN, M. Divulgação científica na sala de aula: implicações de um gênero. In CUNHA, M.B.; GIORDAN, M. **Divulgação científica na sala de aula: perspectivas e possibilidades**. Ijuí: Editora Unijuí, 2015.

DANTAS, L.F.S.; DECCACHE-MAIA, E. Divulgação Científica no combate às Fake News em tempos de Covid-19. **Research, Society and Development**, [s.l.], v. 9, n. 7.F., 2020.

FINATTO, M.J.B.; EVERS, A.; STEFANI, M. Letramento científico e simplificação textual: o papel do tradutor no acesso ao conhecimento científico. **Letras (Santa Maria)**. Santa Maria/Brasil. V.26, n.52, p.135-158, 2016.

FREIRE, N.P. Divulgação científica imuniza contra desinformação. **Ciência & Saúde Coletiva**, [s.l.], v. 26, p. 4810-4810, 2021.

HALLIDAY, M.A.K. Towards probabilistic interpretations. In **Functional and systemic linguistics: Approaches and uses**, v. 19991, p. 39-61, 1991.

LIMA, G.S.; GIORDAN, M. Da reformulação discursiva a uma práxis da cultura científica: reflexões sobre a divulgação científica. **História, Ciências, Saúde-Manguinhos**, [s.l.], v. 28, p. 375-392, 2021.

LIMA, G.S.; GIORDAN, M. O movimento docente para o uso da divulgação científica em sala de aula: um modelo a partir da teoria da atividade. *Revista Brasileira de Pesquisa em Educação em Ciências*, v.18, n.2, p.493-520, 2018.

MEDEIROS, Vanise. Glossários. In MEDEIROS, Vanise; Esteves, Phellipe Marcel S. et al. (Org.). **Almanaque de Fragmentos**: ecos do século XIX. Campinas: Pontes, 2020, p. 09-20.

POCOBELLO, F.S.; SILVA, M.E.F.A.; SOUZA, J.W.C. Análise de perfil dos divulgadores científicos do projeto de extensão A voz da ciência. **Anais** do Congresso de Extensão Universitária da UFABC. [s.l.], v.6, n.1, p. 10-12, 2022.

SOUZA, J.W.C. A divulgação científica cindida entre a objetividade e subjetividade: algumas reflexões e contribuições. **Trem de Letras**, [s.l.], v. 8, n. 3, p. e021004-e021004, 2021.

SOUZA, J.W.C.; ARAUJO, M.P.; PEREIRA, L.N.; PEREIRA, G.A. Editorial: Sobre revitalizações e recomeços: Um processo em movimento. **Caderno de Estudos Interdisciplinares**, [s.l.], v. 4, n. 1, p. 5-10, 21 nov. 2022.

VILAÇA, A.M.S.; CARVALHO, B.A.S.; SOUZA, J.W.C.; SOUZA, J.P.R.; DOMINGUETE, L.H.; SILVA, Y.A. Análise de perfil do público-alvo do projeto de divulgação científica A Voz da Ciência. **Anais** do Congresso de Extensão Universitária da UFABC. [s.l.], v.6, n.1, p.7-9, 2022.