

FLORES, V. M.; REBECHI, R. R. Linguística de Corpus aplicada à Libras: ferramentas de transcrição para criação de corpus. *ReVEL*, edição especial, v. 21, n. 20, 2023. [www.revel.inf.br]

LINGUÍSTICA DE CORPUS APLICADA À LIBRAS: FERRAMENTAS DE TRANSCRIÇÃO PARA CRIAÇÃO DE CORPUS

Vanize Martins Flores¹

Rozane Rebechi²

vanizeflores@gmail.com

rozanereb@gmail.com

Resumo: A Língua Brasileira de Sinais (Libras), assim como as outras línguas de sinais, enfrenta o desafio de ser transcrita em elementos característicos da linguagem oral para permitir análises por meio de métodos da Linguística de Corpus. Diversos sistemas de transcrição, como o HamNoSys e o Stokoe Notation, além de sistemas de escrita que também são utilizados para transcrever, como o SignWriting e o Elis, bem como o uso de glosas, são empregados para identificar e registrar os sinais. Este estudo tem como objetivo analisar o desempenho de duas dessas ferramentas, ELAN³ e HamNoSys⁴, para transcrever dois vídeos sobre a Covid-19 com interpretação para a Libras produzidos por duas Instituições de Ensino Superior de Goiás, e discutir as vantagens e limitações de cada uma. Para tanto, apoia-se nos métodos de transcrição descritos por dois grupos de pesquisa, um brasileiro (QUADROS et al. 2020) e um alemão (EBLING 2016), que utilizaram os dois sistemas, respectivamente, para transcrever os sinais da sua pesquisa. Concluímos que para os propósitos desta pesquisa a ferramenta ELAN se mostrou mais adequada para transcrever os dois vídeos sobre a COVID-19, pois o resultado da transcrição pode ser processado por softwares de análise textual, como o AntConc. A inserção de etiquetas para padronizar a classificação dos sinais permitiu a análise dos contextos em que os sinais ocorreram utilizando-se as ferramentas disponíveis no software AntConc.

Palavras-chave: Linguística de Corpus; Libras; Sistemas de transcrição.

ABSTRACT: Brazilian Sign Language (Libras), like other sign languages, faces the challenge of being transcribed into elements characteristic of oral language to enable analyses through Corpus Linguistics methods. Various transcription systems, such as HamNoSys and Stokoe Notation, besides writing systems also used to transcribe, such as SignWriting and Elis, HamNoSys, in addition to the use of glosses, are employed to identify and record signs. This study aims to evaluate the performance of ELAN and HamNoSys transcription tools, highlighting the advantages and limitations of each approach in transcribing videos in Libras. To do so, it relies on the methods described by two research groups, one Brazilian (QUADROS et al. 2020) and one German (EBLING 2016), which used the two systems, respectively, to convert signs from your search. We concluded that, for the purposes of this research, the ELAN method proved more suitable for transcribing two videos on COVID-19, as it enables the transcription to be processed by text analysis tools, such as AntConc. The insertion of labels to standardize the classification of signs allowed for the analysis of the contexts in which the signs occur, through the tools available on the AntConc software.

KEYWORDS: Corpus linguistics; Brazilian Sign Language; Transcription systems.

¹ Doutoranda no Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul – UFRGS.

² Professora no Departamento de Línguas Modernas do Instituto de Letras da Universidade Federal do Rio Grande do Sul – UFRGS.

³ Disponível em: <https://archive.mpi.nl/tla/elan>. Acesso em: 10 out. 2023.

⁴ Disponível em: <https://www.sign-lang.uni-hamburg.de/dgs-korpus/hamnosys-97.html>. Acesso em: 30 out. 2023.

Introdução

A Linguística de Corpus (LC) é uma abordagem para compilação e análise de conjuntos de textos autênticos reunidos criteriosamente, de acordo com os objetivos da pesquisa. A LC se concentra na análise quantitativa e qualitativa dos padrões linguísticos encontrados nos corpora e envolve a aplicação de métodos estatísticos e computacionais para extrair informações significativas sobre a estrutura e o uso da linguagem (Berber Sardinha 2004; Tagnin 2013). Essa abordagem tem sido especialmente útil para explorar questões como variação linguística geográfica, diacrônica, de gênero, entre outras, além de aquisição da linguagem, estudos de tradução e outras temáticas.

Os pesquisadores em LC frequentemente utilizam ferramentas de processamento de linguagem natural e estatísticas para analisar grandes conjuntos de dados linguísticos. Essa abordagem possibilita uma visão mais abrangente e objetiva dos padrões linguísticos do que os métodos tradicionais baseados em intuição ou que consideram apenas amostras de textos. A compilação e análise de dados linguísticos textuais criteriosamente coletados e processados por ferramentas computacionais têm sido amplamente empregadas nas últimas décadas em diversas áreas de pesquisa, como Estudos de Tradução (Baker 1993), Terminologia (Bevilacqua 2021), Estilística (Mcintyre 2019) e Discurso (Rebechi 2019).

Entre as inúmeras possibilidades de aplicação dessa metodologia, a investigação de termos e fraseologias recorrentes, a fim de se observar, por exemplo, a consistência terminológica, é especialmente relevante para a análise da linguagem de especialidade (Cf. Carvalho e Rebechi 2021). Até mesmo o discurso oral pode ser investigado por meio de ferramentas de análise textual (ver, entre outros, Prado 2019), visto que sua transcrição já conta com softwares de reconhecimento de voz, tais como o reconhecimento de fala do Windows 10, a digitação por voz do Google Docs, o Gboard, o Dragon by Nuance⁵, entre outros, que transformam as falas em texto escrito, formato exigido pelas ferramentas de análise textual, como AntConc (Laurence 2022), Sketch Engine (Kilgarriff et al. 2014) e outras.

⁵ Disponível em: <https://zapier.com/blog/best-text-dictation-software/#windowsspeech>. Acesso em: 30 set. 2023.

No que tange às línguas de sinais (LS), a aplicação de ferramentas de LC para sua investigação tem como primeiro obstáculo a necessidade de transformar a modalidade visual em texto escrito, que então poderá ser processado por softwares. Possivelmente impactadas por essa dificuldade, as pesquisas que aliam LC e LS ainda são escassas.

Para auxiliar na tarefa de transcrição dos sinais em texto escrito, existem algumas ferramentas disponíveis. Este estudo tem como objetivo analisar o desempenho de duas dessas ferramentas de transcrição, o ELAN⁶ e o HamNoSys⁷, a fim de verificar qual delas pode ser mais adequada para os objetivos desta pesquisa. Para isso, foram utilizados dois vídeos sobre a Covid-19 com interpretação para a Libras produzidos por duas Instituições de Ensino Superior de Goiás.

A realização desta análise surge da necessidade de selecionar uma ferramenta de transcrição da Língua Brasileira de Sinais (Libras) adequada para auxiliar em uma pesquisa em nível de doutorado iniciada em 2022 junto ao Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul - UFRGS. A proposta da pesquisa de doutorado é coletar um corpus em Libras sobre temas relacionados à saúde disponíveis no site TelessaúdeRS⁸ e analisar a frequência dos sinais para determinado conceito clínico, bem como verificar o contexto em que o sinal foi inserido. Para isso, serão realizadas entrevistas em formato de vídeo com tradutores e intérpretes de Libras a fim de coletar os sinais utilizados por eles e, posteriormente, o material será transcrito para o português brasileiro. A transcrição do material coletado é um passo importante, pois a partir disso será possível ter textos compatíveis para a leitura em softwares de LC.

Fundamentam este estudo pesquisas realizadas no Brasil e no exterior. No Brasil, um dos grupos que têm aliado LC e Libras é o Corpus de Libras⁹ (Quadros 2020), coordenado por Ronice Quadros da Universidade Federal de Santa Catarina (UFSC). Desde 1995, esse grupo vem construindo diversos corpora utilizando o Sistema de Anotação Eudico Annotator (ELAN) para a transcrição dos vídeos. Já Sarah Ebling (2016), pesquisadora da Universidade de Zurique, construiu junto ao seu grupo de pesquisa um corpus paralelo de anúncios de trens em língua alemã e

⁶ Disponível em: <https://archive.mpi.nl/tla/elan>. Acesso em: 10 out. 2023.

⁷ Disponível em: <https://www.sign-lang.uni-hamburg.de/dgs-korpus/hamnosys-97.html>. Acesso em: 30 out. 2023.

⁸ Disponível em: <https://www.ufrgs.br/telessauders/materiais-pacientes/>. Acesso em 10 out. 2023.

⁹ Disponível em: <http://corpuslibras.ufsc.br/> 2020. Acesso em: 15 ago. 2023.

língua de sinais suíça-alemã (DSGS) utilizando o Sistema de Notação de Hamburgo para Língua de Sinais (HamNoSys), que possibilita a transcrição considerando as características visuais da LS.

Este trabalho está organizado da seguinte forma: na seção 1, descrevemos brevemente a história das LS e da Libras. A seção 2 discute estudos que aliam Linguística de Corpus e Libras, e apresenta os sistemas de transcrição avaliados. Na seção 3, relatamos o processo da utilização do ELAN para esta pesquisa e justificamos a escolha desse sistema para a nossa pesquisa de doutorado.

1. Contextualizando a LS

Para contextualizar a LS e a comunidade surda, apresentamos um breve recorte histórico a fim de facilitar a compreensão do cenário atual dessa área de estudo. Antes de tudo, é importante salientar que a educação de pessoas surdas foi construída por pessoas “leigas”, bem como pelos chamados “especialistas” da área, influenciando não somente a construção da educação para pessoas surdas, mas também a aquisição e o desenvolvimento da LS. Ladd (2013: 13) define leigo como “qualquer um que não esteja diretamente empregado em domínios relacionados com Surdos, nem em domínios profissionais adjacentes”. Já “especialistas” se referem àqueles que se ocupavam em estudar a surdez como deficiência e suas implicações.

Segundo Karin Strobel (2008), a história das pessoas surdas surge a partir de duas linhas de pensamentos diferentes. A primeira trata da surdez da perspectiva clínica, em que as pessoas surdas são representadas como deficientes, e mesmo como anormais, por conta da falta de audição. A segunda linha sai em defesa da cultura surda, abordando a surdez por um viés antropológico, considerando-se a forma como a pessoa surda é narrada a partir de diferentes contextos sociais, bem como sua construção de identidade.

Muitos pensadores gregos e romanos acreditavam que a pessoa surda não desenvolvia uma linguagem e, portanto, não poderia desenvolver pensamento nem aprendizagem. Conforme Sacks (2010), o filósofo grego Aristóteles (384-322 a.C.), por exemplo, afirmou que a audição era fundamental para a linguagem. Outro filósofo grego a se destacar no assunto é Sócrates (470-399 a.C.). Ele foi o primeiro a trazer uma reflexão sobre o que, posteriormente, seria classificado como língua de sinais, conforme segue:

Se não tivéssemos voz nem língua, mas apesar disso desejássemos manifestar coisas uns para os outros, não deveríamos, como as pessoas que são mudas, nos empenhar em indicar o significado pelas mãos, cabeças e outras partes do corpo? (Sacks 2010: 31).

Naturalmente, devemos ler a citação acima compreendendo que as escolhas das palavras retratam um pensamento da época, quando, por exemplo, aqueles que não conseguiam se comunicar por meio da fala eram classificados como “mudos”. Atualmente, compreende-se que a língua de sinais não é um apanhado de gestos, mas sim uma modalidade visual-espacial, com estrutura gramatical própria e independente da modalidade de língua oral.

Já no século XVI, os pioneiros na educação de pessoas surdas foram Pedro Ponce de León (1520-1584) e Juan Pablo Bonet (1579-1629). Entretanto, há indícios de que o médico italiano Girolano Cardano (1501-1576) defendesse a instrução das pessoas surdas e sua linguagem durante o mesmo período que León e Bonet, sendo que o seu interesse nesse assunto foi motivado pela surdez do primeiro filho.

Conforme Strobel (2008), o uso da leitura labial e a introdução do aparelho auditivo em pessoas surdas iniciaram a partir dos estudos do médico Johann Conrad Amman (1669-1724). Ele passou a se dedicar à educação das pessoas surdas e a fazer com que o sujeito surdo associasse o som das palavras à escrita. Na França, o abade Charles Michel de L’Epée (1712-1789) deu início aos “Sinais metódicos”, baseados na língua de sinais e na gramática francesa. Ele aplicou essa metodologia na educação de pessoas surdas após conhecer duas irmãs gêmeas surdas que utilizavam a língua de sinais para se comunicar.

A *Institution Nationale des Sourds-Muets à Paris*, fundada em 1755 por L’Epée, foi a primeira escola a receber ajuda dos órgãos públicos. Infelizmente, L’Epée recebeu muitas críticas de educadores oralistas que defendiam a proibição da língua de sinais e praticavam a oralização das pessoas surdas. Roch-Ambroise Sicard (1742-1822) se tornou diretor do instituto após o falecimento de L’Epée, dando continuidade à metodologia do abade. Segundo Sacks (2010: 27), L’Epée “treinou numerosos professores para os surdos, e estes, na época da morte do abade, em 1789, já haviam criado 21 escolas para surdos na França e na Europa”.

O norte-americano Thomas Hopkins Gallaudet (1787-1851) visitou o instituto na França, surpreendeu-se com os “sinais metódicos” e passou a utilizar as teorias de L’Epée em seu ensino. Ele buscava, pela Europa, metodologias de estudo para

peças com surdez e professores qualificados a fim de fundar uma escola de surdos nos EUA. Nessa visita, Gallaudet conheceu Laurent Clerc, que atuava no ensino das peças surdas no instituto de Paris. Laurent chegou aos Estados Unidos em 1816, a convite de Gallaudet e, juntos, fundaram, em 1817, a primeira escola de surdos nos EUA com o nome de *American Asylum for the Deaf*, que mais tarde passou a se chamar *American School for the Deaf*, em Hartford.

Praticamente todos os professores de surdos que estudaram em Hartford tornaram-se fluentes em língua de sinais. Sacks (2010: 32) acrescenta que “o aumento da alfabetização e educação entre os surdos foi tão espetacular nos Estados Unidos quanto fora na França, e logo difundiu-se por outras partes do mundo”.

Enquanto o movimento em prol da educação das peças surdas se desenvolvia em outros países, o Brasil, através de Dom Pedro II, convidou o francês surdo Eduard Huet (1822-188?) para trazer sua metodologia de ensino aos surdos brasileiros. Huet veio ao Brasil em 1855 e, em 26 de setembro de 1857, foi fundada a primeira escola para surdos no Rio de Janeiro, o Imperial Instituto dos Surdos-Mudos. Depois de algumas alterações, a escola recebeu o nome de Instituto Nacional de Educação de Surdos (INES) (Strobel 2008: 89). Em 1861, Huet deixou a direção do Instituto devido a questões particulares e, a partir de então, diferentes diretores assumiram o instituto, incluindo peças ouvintes.

Em 1869, havia 550 professores de surdos espalhados pelo mundo. Desses, 41% eram professores surdos que frequentavam as escolas nos EUA. Pouco antes, em 1864, o Congresso dos Estados Unidos aprovou uma lei autorizando a fundação da *Columbia Institution for the Instruction of the Deaf and Dumb and Blind*, nomeada Gallaudet College em 1894 e que, posteriormente, se tornou Gallaudet University. A Gallaudet segue sendo a única universidade no mundo totalmente dedicada ao ensino das peças surdas na área das artes liberais.

Apesar de a LS ter sido uma conquista na educação de surdos, ainda havia, na Europa e nos EUA, aqueles que discordavam dessa metodologia de ensino e defendiam o oralismo. Para fortalecer os estudos oralistas, em 1880 foi realizado o Congresso Internacional de Professores de Surdos em Milão. Esse foi um marco importante para a comunidade surda, pois influenciou a educação dos surdos ao longo dos séculos. Com a participação de 163 peças ouvintes e 1 pessoa surda, todos professores de surdos e de diferentes países, ficou decidido que a educação de surdos deveria ser realizada a partir do método oralista. Como afirma Skliar (2001), a

oralização não começou após as deliberações do congresso, pois ela já era prática comum de ensino, mas o Congresso de Milão legitimou a sua prática. Após o Congresso de Milão, como ficou conhecido, a educação de pessoas surdas desestabilizou-se.

Durante aproximadamente cem anos (1880-1980), as pessoas surdas foram obrigadas a esconder suas línguas de sinais, sua cultura e sua identidade para se submeterem a treinamentos oralistas, no qual predominava o etnocentrismo ouvintista. Segundo Skliar (2001), o que caracteriza o ouvintismo é a obrigatoriedade dos surdos se encaixarem no conjunto de representações dos ouvintes. Segundo Brito (2010), caso os estudos das línguas de sinais já estivessem ocorrendo e os pesquisadores linguistas estivessem presentes no congresso, muito provavelmente o rumo da educação de surdos teria sido outro em vários países.

Décadas mais tarde, em 1960, o linguista William Stokoe publicou o artigo *Language Structure: an Outline of the visual Communication System of the American Deaf*, trazendo reflexões acerca da cultura surda e da língua de sinais. Ele defendia que havia uma estrutura linguística na língua de sinais, assim como há na língua oral. Stokoe, que investigou na Língua de Sinais Americana (ASL) os seus níveis fonológicos e morfológicos, conforme nos apresenta Gesser (2009: 14), apontou a existência de três parâmetros que, posteriormente, foram classificados como parâmetros primários, constituintes dos sinais. São eles: configuração de mão (CM), ponto de articulação (PA) ou locação (L) e movimento (M). A Libras segue esse padrão de estruturação do sinal apresentado por Stokoe.

Alguns anos depois, as leis, o movimento de reconhecimento da pessoa surda e os estudos sobre a comunidade surda, a língua de sinais e a cultura surda ganharam forma com o crescente número de publicações acadêmicas. A sanção da Lei nº 10.436 de 24 de abril de 2002 é um dos principais fatores que contribuíram para a propagação da língua de sinais. Ela reconhece a Libras como meio legal de comunicação e expressão para a comunidade surda do Brasil. Ainda, a legislação reafirma que a Libras possui um sistema linguístico de natureza visual-motora e uma estrutura gramatical própria. Para Brito (2010), o que distingue a língua de sinais da língua oral não é a utilização das mãos e do corpo ou o uso do aparelho fonador, mas a forma como cada língua se organiza.

Além do reconhecimento legal da Libras como língua, outras leis buscam assegurar o acesso à informação para pessoas surdas. O Decreto Lei nº 5.626/05, que

regulamenta a Lei nº 10.436/02, orienta sobre a educação de surdos e o atendimento a pessoas surdas, determinando a obrigatoriedade de haver profissionais tradutores e intérpretes de Libras em espaços públicos, a fim de incluir a pessoa surda na sociedade. Além disso, há a Lei nº 14.191/21, que trata sobre a modalidade de educação bilíngue (Libras como primeira língua e português escrito como segunda língua) para surdos, a Lei Brasileira de Inclusão (LBI) nº 13.146/15, e outras leis que não serão mencionadas nesta pesquisa, mas que possuem valor inestimável para a comunidade surda.

Os estudos surdos compreendem um campo fértil de produção de saberes e conhecimentos pelo qual a comunidade acadêmica tem se interessado cada vez mais. Entre vários estudiosos, destacam-se, no Brasil, Lucinda Ferreira Brito (1993), Ronice Müller de Quadros (1997, 2013, 2017 entre outros) e Lodenir Karnopp (2013), que tratam sobre aquisição da linguagem, educação de surdos e a gramática da língua de sinais. Gladis Perlin (2001) e Karin Strobel (2008), também autoras nacionais, pesquisam sobre a construção da identidade surda e da cultura surda. Pesquisadores internacionais também possuem um papel importante no debate sobre os estudos surdos, entre eles destacamos o britânico Paddy Ladd (2013), que discutiu o significado da existência surda, criando o conceito de deafhood, ou surdidade, e o argentino Carlos Skliar (2001 e 2013) que, além de tratar da identidade, como Ladd, ainda discute sobre a educação bilíngue para pessoas surdas.

Conforme Skliar (2001), o discurso clínico-terapêutico ligado à surdez fez com que os surdos buscassem se encaixar em um modelo corretivo, impactando a construção da educação de surdos. O autor propõe que a surdez seja pensada além da dicotomia entre modelo clínico e modelo antropológico, mas representada como uma diferença, uma experiência visual, um lugar de identidade múltipla e multifacetada e dentro do discurso sobre a deficiência.

Portanto, conhecendo a história da língua de sinais, compreendendo que os estudos relacionados a Libras estão em crescente desenvolvimento, mas observando a escassez de trabalhos que tratem da aplicação da LC à Libras, passamos a apresentar a utilização das ferramentas de transcrição da LS a fim de tornar possível a sua leitura em softwares de análise de corpus.

2. A Linguística de Corpus na Libras

A aplicação da metodologia subjacente à LC na Libras ainda é incipiente. Uma das razões que pode ajudar a explicar essa escassez é a dificuldade de transcrição dos sinais em caracteres que a maior parte das ferramentas computacionais processa. Um dos desafios substanciais ao analisar uma forma de expressão utilizando ferramentas originalmente concebidas para outra modalidade reside na necessidade de transcrição. Por exemplo, para a investigação do discurso oral com o auxílio das ferramentas da LC é imperativo transformar os diálogos falados em palavras escritas. Embora, na atualidade, possamos contar com recursos computacionais que tornam esse processo mais eficiente, sua complexidade persiste. Mesmo com avanços tecnológicos, é possível que ocorram erros, como a grafia equivocada de palavras homófonas em contextos específicos. Por exemplo, a ferramenta pode grafar “seção” em vez de “sessão”. É crucial notar, no entanto, que tais imperfeições tendem a ser “neutralizadas” diante dos acertos, mitigando eventuais impactos negativos nos resultados obtidos.

Contudo, a transcrição de sinais para a modalidade escrita é ainda mais desafiadora, já que compreende não só a interpretação do sinal pelo transcritor, mas também a transformação desse sinal em uma sequência alfanumérica. Quadros (2016) enfatiza que realizar a transcrição da LS para a língua portuguesa com o objetivo de coletar dados por meio das ferramentas de LC é de suma importância, pois esses dados poderão auxiliar em novos estudos linguísticos e na construção de ferramentas de apoio aos profissionais da área da LS.

Um dos textos que buscamos como referência metodológica para o desenvolvimento desta pesquisa intitula-se “A transcrição de textos do Corpus de Libras” (Quadros 2016). Nele são exemplificados os processos de registro, transcrição e compilação do Corpus utilizado no projeto, além de ser debatida a importância de se olhar mais para a Libras por meio da LC, bem como de voltar a atenção para a LC pela perspectiva da Libras.

O primeiro corpus criado a partir do projeto Corpus de Libras começou a ser compilado em 1995, abrangendo o processo de aquisição da linguagem de um grupo de crianças surdas cujos pais também são surdos. Com o avanço das pesquisas do grupo, foram inseridos dados de crianças surdas filhas de pais ouvintes, crianças ouvintes filhas de pais surdos e crianças surdas com implante coclear filhas de pais surdos. Conforme o projeto se solidificou e seus resultados se tornaram públicos, outros pesquisadores puderam ter acesso a essa metodologia e desenvolver seus

trabalhos utilizando os padrões de registro e transcrição desenvolvidos para a construção do Corpus de Libras.

Esse modelo de registro está organizado em um projeto chamado “Proposta de Manual de Transcrição do Corpus LIBRAS Inventário Nacional” (Quadros et al. 2020)¹⁰. Ele possibilitou que um glossário mais robusto da Libras fosse alimentado por pesquisadores de Santa Catarina e outros grupos de pesquisa de diferentes estados, não somente os que trabalhavam com a temática da aquisição da linguagem. Além desse manual, é possível encontrar tutoriais para realizar a transcrição do material em vídeo utilizando a ferramenta ELAN, processo que será explicado abaixo.

Encontramos no site do projeto vídeos postados por pesquisadores de Santa Catarina, Rio Grande do Sul, São Paulo, entre outros estados brasileiros¹¹. É possível ver nos vídeos pessoas sinalizando, individualmente, algum tema relacionado a uma das três propostas do projeto, conforme será explicado abaixo. Para buscar os registros em vídeos, basta clicar em “Acervo” (Figura 1) e escolher uma região do mapa do Brasil para visualizar os vídeos disponíveis.

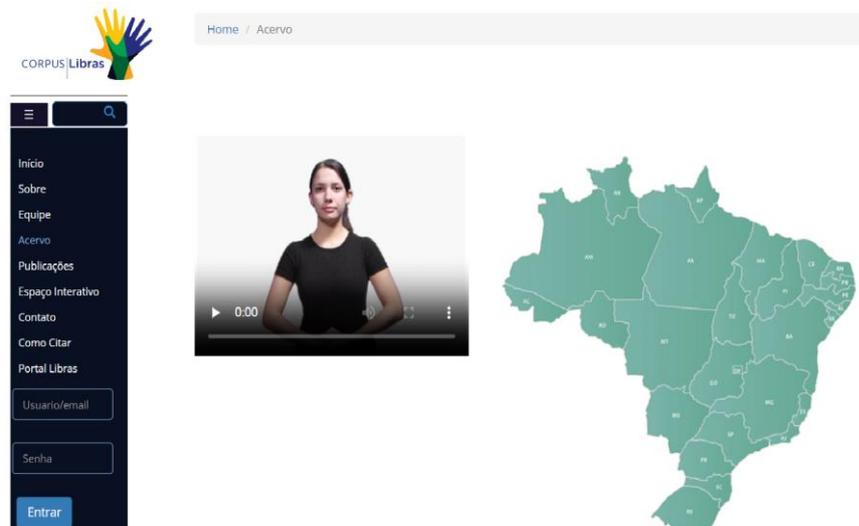


Figura 1: Aba Acervo localizada na página do Corpus de Libras¹²

Caso o estado escolhido tenha projetos registrados no site, uma nova janela se abrirá com o material disponibilizado, como é possível observar na Figura 2.

¹⁰ Disponível em: <https://corpuslibras.ufsc.br/publicacoes/index?page=2>. Acesso em: 10 out. 2023.

¹¹ Disponível em: <https://corpuslibras.ufsc.br/dados>. Acesso em: 15 ago. 2023.

¹² Disponível em: <https://corpuslibras.ufsc.br/dados>. Acesso em: 10 out. 2023.



Figura 2: Aba de São Paulo localizado no Acervo do Corpus de Libras¹³

Atualmente, o Corpus de Libras conta com 3 projetos: “Inventário Nacional de Libras”, com coleta de sinais utilizados por pessoas surdas em três faixas etárias (dos 19 aos 29 anos, dos 30 aos 49 anos e acima de 50 anos); “Materiais de produção acadêmica” construídos com o apoio de diferentes universidades em diferentes modalidades de ensino (graduação, pós-graduação, palestras e afins); e “Antologias Literárias em Libras”, em que diferentes pessoas são chamadas para recitar poesias e contar histórias. Todo o material coletado pode ser encontrado no site do projeto Corpus de Libras, de forma gratuita, a partir de um cadastro que gera login e senha.

Uma proposta para a realização das transcrições é o uso de glosas na língua escrita para a qual os sinais são traduzidos. Porém, recai sobre o linguista uma grande responsabilidade, pois transcrever é um ato de análise por si só (Johnston 1991 apud Quadros 2016). Nesse caso, a transcrição está intimamente ligada à análise e isso pode influenciar o resultado da pesquisa.

Outro projeto que utiliza método de transcrição semelhante ao do Corpus de Libras é o “Auslan Corpus”, desenvolvedor do glossário identificado como ID-glosa. O grupo começou a coletar dados em 2005 e tem como proposta reunir os vídeos, as anotações e os metadados obtidos para disponibilizar ao público no *Auslan Lexical Database*. O projeto reuniu cerca de 150 horas de material em vídeo sinalizado por diferentes usuários nativos ou quase nativos da língua australiana de sinais que, posteriormente, foi transformado em 1.700 textos. Para isso, foram utilizados os materiais disponibilizados pelo *Sociolinguistic Variation in Auslan Project (SVIAP)* e

¹³ Disponível em: <https://corpuslibras.ufsc.br/dados>. Acesso em: 10 out. 2023.

pelo *Endangered Languages Documentation Project* (ELDP), conforme nos apresenta Johnston (2008: 83).

O projeto SVIAP contou com a participação de 211 pessoas sinalizantes moradores das cidades de Sydney, Melbourne, Brisbane, Adelaide e Perth, enquanto o ELDP realizou a entrevista com 100 participantes, também moradores das cinco cidades citadas acima. A coleta foi realizada a partir de uma entrevista, de conversação livre e de respostas relacionadas a materiais que lhes eram apresentados, como, por exemplo, uma história ilustrada, ou desenho animado ou um conto traduzido para a Auslan. Esse banco de dados pode ser encontrado tanto no formato de dicionário (Auslan Signbank), disponibilizado *on-line* e com acesso aberto, quanto no formato de um banco de dados de referência com acesso limitado a pesquisadores.

Conforme o autor, o ID-gloss funciona da seguinte forma:

Por exemplo, se uma pessoa sinaliza CASA (um sinal iconicamente relacionado com a forma de um telhado), mas, na verdade, quer dizer LAR, ou executa uma forma particularmente grande e exagerada do sinal CASA, referindo-se à mansão, (sem que essa forma modificada em si seja um lexema reconhecido e característico da língua), então o ID-glossa CASA ainda seria usado em ambas as instâncias para identificar o sinal na anotação. (Johnston 2008: 84. Tradução nossa)¹⁴

Na busca por padronizar a forma de transcrição do material no Auslan Corpus, foram utilizadas todas as variantes fonológicas e morfológicas de um lexema. Para a transcrição do material em vídeo foi utilizada a ferramenta ELAN. Com a utilização dessa ferramenta de transcrição, foi possível gerar padrões de transcrições, bem como modelos de *tags* (etiquetas) fixas. São exemplos de glosa a utilização de letras maiúsculas para os sinais (ex.: COMER) e com hífen quando for mais de uma palavra (ex.: COMER-MAÇÃ), fs- para identificar palavras que foram soletradas (ex.: (fs-) NUNCA), o uso de IND quando houver apontação (ex.: IND ELA), entre outras glosas descritas tanto por Johnston (2008) quanto por Quadros (2016).

Johnston (2008) e Quadros (2016) compreendem a importância de se utilizarem as glosas para facilitar o processo de transcrição, porém, assumem que esse processo também apresenta limitações. Exemplo disso é a impossibilidade de expressar em glosa algumas características específicas de um sinal, como, por

¹⁴ *For example, if a person signs HOUSE (a sign iconically related to the shape of a roof) but actually means home, or performs a particularly large and exaggerated form of the sign HOUSE, implying mansion, (without that modified form itself being a recognized and distinctive lexeme of the language) then the ID-gloss house would still be used in both instances to identify the sign in the annotation.*

exemplo, a configuração de mão, o movimento, a expressão não manual (expressões faciais e corporais) e outros parâmetros responsáveis pela construção do sinal.

Para transcrever essas características morfológicas, de forma que fossem reconhecíveis por ferramentas de análise textual, por exemplo, seria necessário criar um método para descrevê-las na glosa, o que poderia influenciar a interpretação do sinal. Por exemplo, se a sinalização do termo GRIPE for acompanhada de uma expressão facial que denota tratar-se de uma “gripe mais suave”, pode ser traduzido como RESFRIADO. Mas se a informação sobre a expressão facial não estiver disponível em alguma parte da transcrição, outro pesquisador ao acessar o material escrito interpretará a palavra somente como GRIPE.

Com o objetivo de manter as nuances dos sinais, Prillwitz e seus colaboradores criaram em 1987 o sistema HamNoSys na Universidade de Hamburgo, Alemanha, com a proposta de ser utilizado internacionalmente. O sistema tem em torno de 200 símbolos e traz símbolos icônicos para facilitar seu uso, além de ser um sistema econômico por conta da quantidade de símbolos. A seguir será explicado um pouco mais sobre as características desse sistema.

2.1 Sistema de Notação de Hamburgo - HamNoSys

O HamNoSys utiliza símbolos para transcrever os sinais em um nível fonético-fonológico. Integrado de forma fluida às ferramentas padrões de computação, o sistema apresenta uma sintaxe formal, seguindo um princípio de composicionalidade. Esse sistema é dinâmico e está constantemente aberto a melhorias, mantendo-se atualizado em consonância com a evolução das LS. Ele se baseia no sistema de anotação que Stokoe (2005) desenvolveu para descrever os parâmetros da língua de sinais. Fazem parte do sistema de anotação criado por Stokoe a configuração de mão (no caso, apenas o formato da mão) e o movimento. Esse sistema é aprimorado no HamNoSys, onde são inseridas a orientação da mão, a locação, a expressão facial e a possibilidade de registro da sinalização com as duas mãos (Hanke 2004).

Observe a seguir um exemplo apresentado por Hanke (2004) de uma transcrição da Língua de Sinais Alemã (DGS) para o sistema HamNoSys, em que o

sinal de Hamburgo está transcrito a partir da sua configuração de mão, orientação, locação e o movimento.

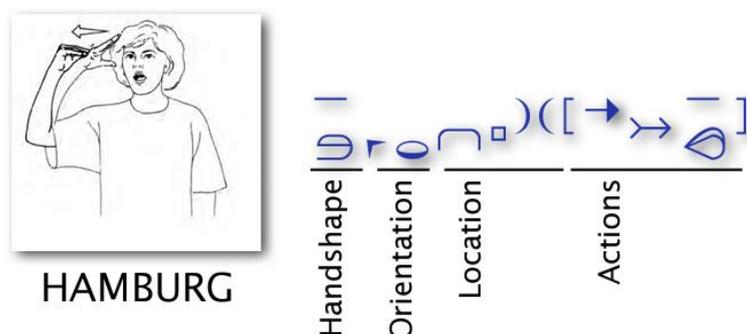


Figura 3: Exemplo da transcrição de um sinal em DGS utilizada pelo HamNoSys¹⁵

Segundo Hanke (2004), os objetivos desse sistema são: a) a utilização internacional, ou seja, praticamente todas as línguas de sinais podem ser transcritas de forma padronizada; b) a representação da iconicidade do sinal, criando símbolos que representem, na forma mais próxima possível, os parâmetros constantes no sinal; c) a economia, levando-se em consideração que a transcrição do sinal evitaria longas explicações no texto escrito sobre o movimento, a expressão facial ou outro parâmetro; d) a integração com ferramentas de informática padrão, em que a transcrição pode ser realizada em um computador e compilada em uma base de dados; e) a descrição da sintaxe formal, que deve ter uma sintaxe bem definida e seguir o princípio dos parâmetros da língua de sinais; f) e a manutenção da extensibilidade, ou seja, a busca contínua por aperfeiçoamento da ferramenta HamNoSys.

O projeto de Ebling (2016) utilizou o HamNoSys para a transcrição de anúncios de trens em alemão para a língua de sinais suíço-alemã, justificando o seu uso devido à dificuldade na utilização das glosas para transcrever os sinais para a língua escrita. Apesar de afirmar que as glosas “fornecem uma representação facilmente legível e pesquisável das línguas de sinais” (Ebling 2016: 119), a pesquisadora também comenta sobre a dificuldade de expressar um vocabulário utilizado em uma modalidade visual por meio de uma língua oral, e optou pelo

¹⁵ Disponível em: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HamNoSys_2018.pdf. Acesso em: 10, out. 2023.

sistema HamNoSys devido a sua proximidade com a natureza visual da língua de sinais.

Entretanto, a dificuldade encontrada na utilização desse sistema para a compilação de um corpus a ser lido em *softwares* de LC é a incompatibilidade de leitura, visto que é gerado um arquivo escrito com símbolos. Atualmente, os programas de processamento de corpus possuem suporte apenas para a leitura de textos escritos com caracteres alfanuméricos e em formato *txt* (arquivo de texto). Ainda, apesar de ser uma ferramenta disponibilizada de forma gratuita, o *layout* do teclado do HamNoSys funciona apenas no sistema operativo macOS que é desenvolvido e distribuído pela empresa Apple Inc. desde 2001 e destinado exclusivamente aos computadores Mac. Sendo assim, é necessário explorar mais esse sistema para conhecer as possibilidades de processamento computacional dos símbolos, bem como a sua disponibilidade para os usuários da ferramenta.

2.2 Sistema de Anotação Eudico Annotator - ELAN

O ELAN (EUDICO Linguistic Annotator) é uma ferramenta de anotação que oferece funcionalidades para criação, edição, visualização e pesquisa de anotações em dados de vídeo e áudio¹⁶. Desenvolvido no Instituto Max Planck de Psicolinguística, na Holanda, o ELAN foi projetado para fornecer uma base tecnológica sólida para a anotação e exploração de gravações multimídia, com foco na análise de idiomas, línguas de sinais e gestos. No entanto, ele pode ser utilizado por qualquer pessoa que trabalhe com corpora de mídia, ou seja, dados de vídeo e/ou áudio, para fins de anotação, análise e documentação. Veja a seguir o *layout* da janela principal do ELAN:

¹⁶ Disponível em: https://www.mpi.nl/tools/elan/docs/ELAN_manual.pdf. Acesso em: 28, nov. 2023.

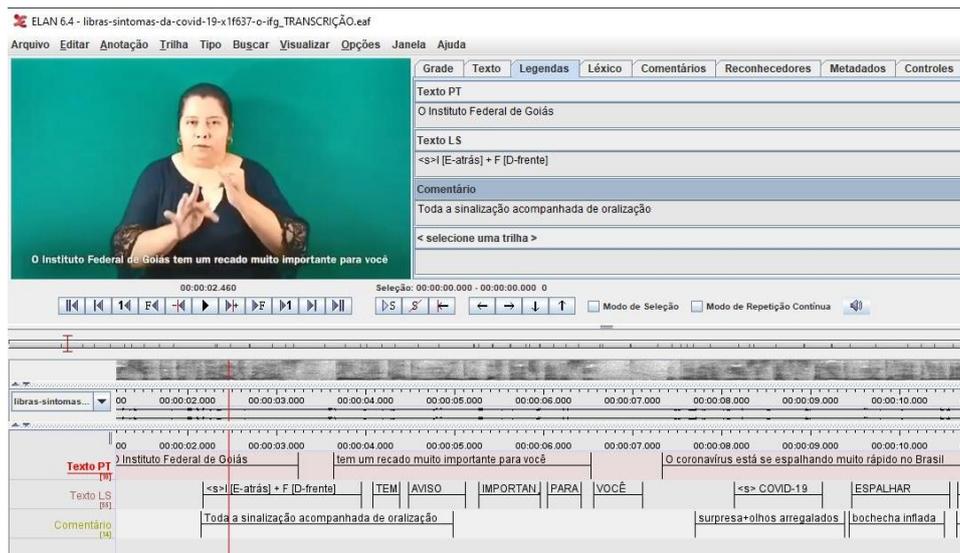


Figura 4: Tela do ELAN

As principais características e funcionalidades do ELAN incluem:

1. **Exibição de Sinais:** Permite a exibição de sinais de fala e/ou vídeo, juntamente com suas anotações.
2. **Vinculação Temporal:** Facilita a vinculação temporal de anotações a fluxos de mídia, possibilitando a sincronização precisa.
3. **Vinculação de Anotações:** Permite a vinculação de anotações entre si, facilitando a organização e análise de informações relacionadas.
4. **Camadas de Anotação Personalizáveis:** Suporta um número ilimitado de camadas de anotação, proporcionando flexibilidade para os usuários definirem categorias e hierarquias.
5. **Suporte a Diferentes Conjuntos de Caracteres:** Lida com diferentes conjuntos de caracteres para garantir a representação adequada de diversos idiomas e símbolos.
6. **Exportação e Importação de Dados:** Oferece a capacidade de exportar anotações em diversos formatos, como texto delimitado por tabulação, e permite a importação de vários formatos de arquivo.
7. **Opções de Pesquisa:** Facilita a pesquisa eficiente por meio de seus recursos de busca.

O ELAN é uma ferramenta versátil, adequada para pesquisadores, linguistas e profissionais que lidam com análise de linguagem, línguas de sinais ou gestos em contextos multimídia. Ele fornece uma estrutura eficaz para a organização e

interpretação de dados complexos de áudio e vídeo. Cada projeto no ELAN consiste em, no mínimo, dois tipos de arquivos: um ou mais arquivos de mídia e um arquivo de anotação. O ELAN possui várias opções de exportação do arquivo, como podemos ver na Figura 5 a seguir:

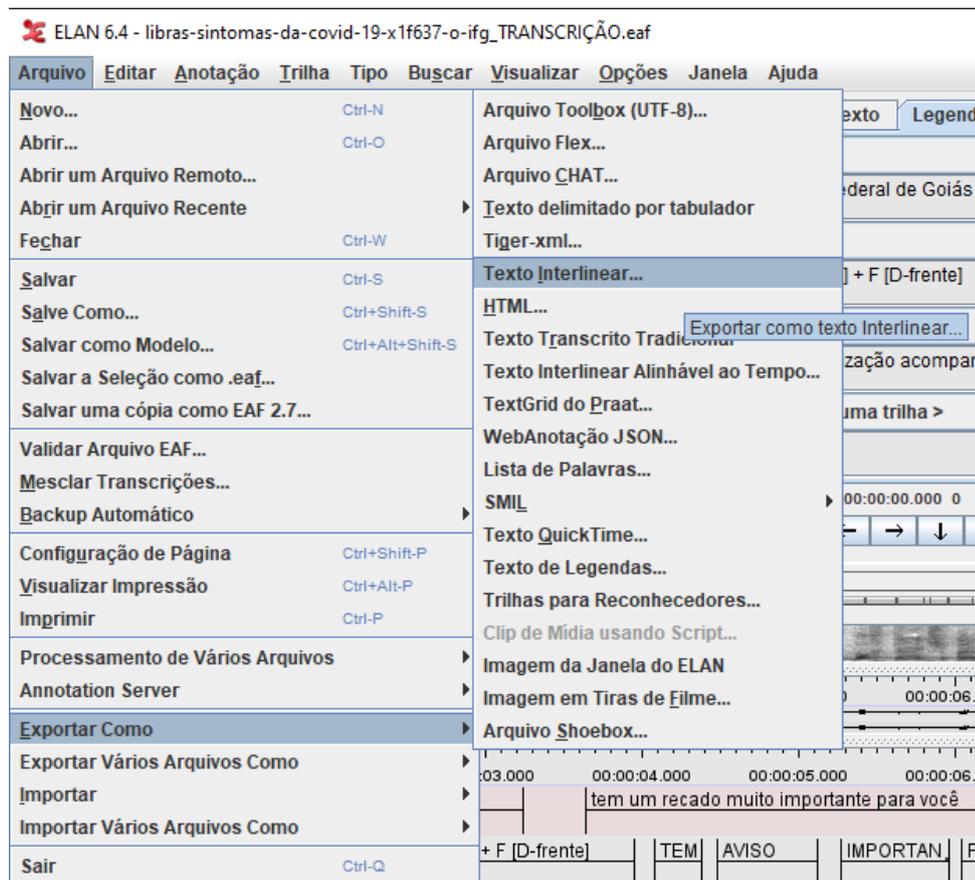


Figura 5: Tela de exportação de arquivo do ELAN

O que é observado, tanto no HamNoSys quanto no ELAN, é que ambas as ferramentas de transcrição possuem a necessidade de seguir critérios bem estabelecidos no momento de realizar as transcrições, sempre respeitando os objetivos do trabalho. Afinal, é a partir do corpus que poderemos confirmar ou não as hipóteses e desenvolver a pesquisa (Tagnin 2015). Caso as fontes de extração do corpus não sejam confiáveis e caso as glosas não sejam escolhidas criteriosamente, os resultados da pesquisa podem ficar aquém do esperado.

Partindo da variedade de opções de exportação do ELAN, que permite aos usuários escolher o formato mais apropriado para atender às necessidades específicas de seu projeto ou para facilitar análises adicionais em diferentes plataformas ou software, a facilidade de manuseio da ferramenta e as possibilidades oferecidas pelas

funcionalidades do ELAN, escolhemos essa ferramenta como *software* de transcrição dos vídeos em Libras para o português do Brasil. Portanto, descreveremos abaixo os detalhes sobre a utilização do ELAN para a construção do Corpus de 2 vídeos curtos, disponíveis *on-line*, que tratam sobre a pandemia da Covid-19 e justificamos a sua escolha.

3. Coleta e investigação do corpus da Covid-19 em Libras

Devido à facilidade de manuseio e possibilidade de realizar a transcrição por camadas, as chamadas trilhas, optamos por utilizar o ELAN como software auxiliar para a criação do corpus de estudo. Para isso, foram selecionados 2 vídeos curtos que estavam disponíveis *on-line* e que tratavam da pandemia da Covid-19. O primeiro vídeo consta na página oficial do Instituto Federal de Goiás no Facebook¹⁷ e o segundo, no canal oficial da TV UFG, no Youtube¹⁸. Ambos os vídeos contam com um narrador, um tradutor e intérprete de Libras e a legenda em português.

A escolha do material levou em consideração a similaridade dos assuntos, pois ambos tratam dos sintomas que o vírus causa ao entrar em contato com o hospedeiro. Além disso, os dois vídeos foram produzidos no estado de Goiás em instituições de ensino federal. Um dos vídeos tem a duração de 50 segundos e o outro, 1 minuto e 30 segundos. Porém o conteúdo do segundo vídeo ocupa menos tempo do que o total, pois é precedido e encerrado por um *card* informativo.

Ao inserir os vídeos no ELAN é possível criar trilhas que serão utilizadas pelo transcritor para inserir as informações desejadas. Esse processo de criação de trilhas com conteúdos diferentes facilita a compartimentação das informações que serão analisadas. Na trilha destinada à transcrição da LS para texto, utilizamos *tags* que irão servir como etiquetas para padronização da busca. São elas: <s> para sinais de termos da área da saúde, <in> para incorporação de numeral e <cl> para classificador.

A glosa e as *tags* utilizadas na transcrição dos materiais apresentados neste trabalho foram elaborados por nós, com o objetivo de testar a utilização do ELAN e o processamento do corpus no AntConc. Porém, como citado na seção 2, já existem

¹⁷ Disponível em: <https://www.facebook.com/IFG.oficial/videos/libras-sintomas-da-covid-19/633465714170476/>. Acesso em: 15 ago. 23.

¹⁸ Disponível em: <https://www.youtube.com/watch?v=XKMTPfuduoM>. Acesso em: 15 ago. 23.

sugestões de glosas que podem ser aplicadas no decorrer de uma transcrição. Abaixo, é possível observar o uso da *tag* <s>, significando que este sinal representa um termo clínico, e a palavra em maiúscula para descrever o sinal, conforme Figura 6 abaixo:

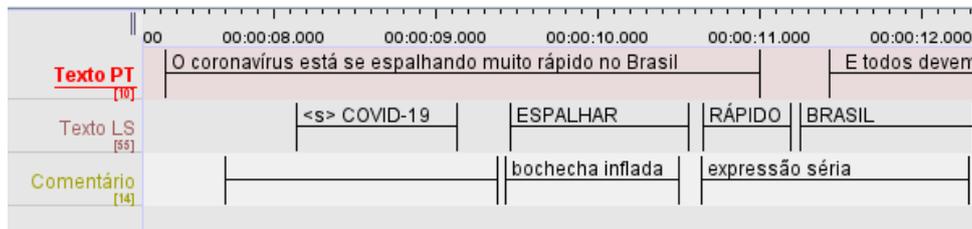


Figura 6: Trilhas do vídeo IFG no software ELAN.

Observe que, para transcrever o sinal de coronavírus, foi utilizada a palavra COVID-19, pois o sinal para ambos os termos é o mesmo. Logo, quando são utilizadas as palavras coronavírus, SARS-CoV-2 ou COVID-19, a identificação da glosa segue COVID-19 para auxiliar na anotação. A escolha acima segue a sugestão apresentada por Johnston (2008) na sua estruturação da ID-glosa para transcrição.

Sobre a *tag* <in>, referente à incorporação de numeral, ela é utilizada quando a configuração de mão do sinal é substituída por uma configuração de mão em numeral para, assim, expressar quantas vezes o evento ocorreu. Podemos observar um exemplo de numeral na Figura 7, quando o intérprete de Libras sinaliza “1 dia” e “4 dias”.



Figura 7: Incorporação de numeral nos sinais 1 DIA e 4 DIAS, respectivamente, retirado do vídeo da TV UFG

Já a *tag* <cl>, que se refere aos classificadores, foi utilizada quando o intérprete fazia uso desse recurso para descrever algo em sua forma, densidade, grandeza ou outra característica pertinente para a tradução, como pode ser visto na Figura 8.



Figura 8: Classificador para a palavra ESPIRRAR retirado do vídeo da TV UFG.

Após a compilação do corpus, escolhemos o AntConc (Anthony 2022)¹⁹ para realizar a análise, por ser de disponibilidade gratuita. No AntConc foram inseridos o corpus de estudo, ou seja, os textos em Libras transcritos para o português brasileiro obtidos no ELAN em formato *txt*, e uma nota retirada do Portal do Ministério da Educação sobre o Dia Nacional do Surdo²⁰ para ser usado como corpus de referência. A escolha do corpus de referência seguiu o critério da relação do tema saúde e surdez e o fato de ser significativamente maior do que o corpus de estudo²¹.

Utilizando-se a função WordList (lista de palavras individuais) do software, foi levantada a lista de todas as palavras dos corpora, bem como sua frequência. O corpus composto das transcrições dos dois vídeos em Libras - doravante, corpus de estudo - totalizou 216 palavras (*tokens*), enquanto o corpus de referência contabilizou 2363 palavras (*tokens*). Por meio do contraste com o corpus de referência, o utilitário Keyword (lista de palavras-chave) identificou as palavras-chave do corpus de estudo, ou seja, aquelas que ocorrem com frequência estatisticamente maior neste do que no corpus de referência. A partir da *tag* <cl>, por exemplo, é possível observar na Figura 9 que foram utilizados classificadores para sinalizar sintomas específicos:

File	Left Context	Hit	Right Context
1 [[Libras] Covid.txt	encontrar primeiro<in> gripe simples objetivo respirar tosse gripe pingar<	cl> *	aponta pescoço* dor inchar<cl> 1 dia espaço 4 dias sumir
2 [[Libras] Covid.txt	objetivo respirar tosse gripe pingar<cl> *aponta pescoço* dor inchar<	cl> 1	dia espaço 4 dias sumir segundo<in> gripe igual primeira<
3 [[Libras] Covid.txt	T-E rosto doença alergia coçar-rosto-peito<int> espirro<	cl>	olhos coçar coçar-garganta<int> próximo/mais vermelho olhos<
4 [[Libras] Covid.txt	cl> olhos coçar coçar-garganta<int> próximo/mais vermelho olhos<	cl	P> quarto<in> Covid-19<s> febre tosse cansaço S-

¹⁹ Disponível em: <https://www.laurenceanthony.net/software>. Acesso em: 30 ago. 23

²⁰ Marques, Luciano. Primeiro aluno surdo de medicina sonha proporcionar atendimento mais autônomo aos deficientes. Portal MEC. c2018. Disponível em: <http://portal.mec.gov.br/component/tags/tag/dia-nacional-do-surdo>. Acesso em: 15 ago. 23.

²¹ Berber Sardinha (2004) recomenda que o corpus de referência seja, no mínimo, três vezes maior do que o corpus de estudo.

Figura 9: Software AntConc com os Key Word In Context de classificadores

Ao olharmos o contexto em que os classificadores ocorrem, na sua maioria, observamos que são utilizados para a sinalização de sintomas como, por exemplo, coriza, dor de garganta, vermelhidão nos olhos e espirro. Isso pode indicar a necessidade de criação de um glossário em que o foco seja discutir a semântica da frase e o impacto da escolha do sinal para a interpretação daquele termo.

Em suma, as vantagens encontradas no ELAN são a possibilidade de compartimentar as informações tornando mais fácil a análise segmentada, bem como a opção de salvar as transcrições em formato *txt*, o que permite a leitura do material no AntConc e em outros softwares de processamento de corpus. Entretanto, a desvantagem desse sistema de transcrição é a impossibilidade de representação dos parâmetros morfológicos presentes no sinal. Ainda assim, até o presente momento o ELAN tem se mostrado eficaz para o desenvolvimento das transcrições.

4. Conclusão

A despeito das tentativas históricas de apagar a comunidade surda, sua cultura e as Línguas de Sinais (LS), estas persistem e, contrariamente, ganharam reconhecimento não apenas por meio de legislações, mas também em esferas acadêmicas e sociais. Nesse contexto, os pesquisadores de LS podem agora direcionar sua atenção para as características e padrões dessas línguas, fortalecendo-as por meio do estudo de seu léxico e estruturas gramaticais, inclusive em diferentes gêneros textuais em que essas línguas se manifestam.

O desenvolvimento, embora restrito, de alternativas para integrar a língua de sinais na LC tem impulsionado esses estudos, proporcionando maior acesso à informação para os surdos e ampliando as pesquisas conduzidas por pessoas surdas, oferecendo uma nova perspectiva nos estudos surdos e na tradução. Este trabalho apresentou uma visão sobre a utilização da LC na compilação de um Corpus em Libras, destacando o uso do sistema de escrita HamNoSys e da ferramenta de transcrição ELAN.

O objetivo foi analisar o desempenho do sistema HamNoSys e do ELAN para selecionar o método mais adequado ao subsequente processamento do texto pelo software AntConc. A transcrição realizada no ELAN revelou-se a melhor opção para compilar um corpus de Libras na área da saúde, visto que o documento final pode ser

salvo em arquivo de texto, formato necessário para a compilação de corpus no AntConc. Em resumo, a oportunidade de estudar a língua de sinais com as ferramentas da LC promove o desenvolvimento educacional e social da comunidade surda, enquanto também beneficia a comunidade ouvinte, podendo gerar novas metodologias aplicáveis às pesquisas das línguas orais.

Referências Bibliográficas

BAKER, Mona. Corpus Linguistics and Translation Studies. Implications and Applications. In Baker, M., Francis, G.; Tognini-Bonelli, E. (Ed.). *Text and technology*. Amsterdam and Philadelphia: Benjamins, 1993. p. 233–250.

BERBER SARDINHA, T. *Linguística de Corpus*. São Paulo, Barueri: Editora Manole, 2004.

BEVILACQUA, C. et. al. *A terminologia do Patrimônio Cultural Imaterial*. (Projeto de Pesquisa). Porto Alegre, UFRGS, 2021.

BRASIL. Lei nº 10.436, de 24 de abril de 2002. Regulamenta a Língua Brasileira de Sinais – LIBRAS. Brasília, 2002. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/l10436.htm. Acesso em: 12 set. 2023.

_____. Lei nº 13.146, de 6 de julho de 2015. Estatuto da Pessoa com Deficiência. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência, Brasília, 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em: 12 set. 2023.

_____. Lei nº 14.191, de 3 de agosto de 2021. Lei de Diretrizes e Bases da Educação Nacional. Altera a Lei nº 9.394, de 20 de dezembro de 1996, Brasília, 2021. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14191.htm. Acesso em: 12 set. 2023.

BRITO, Lucinda Ferreira. *Integração social & educação de surdos*. Rio de Janeiro: Babel, 1993.

_____. *Por uma gramática de Língua de Sinais*. Rio de Janeiro: Tempo Brasileiro, 2010.

CARVALHO, Y. S.; REBECHI, R. R. Inteligibilidade e convencionalidade em textos de divulgação da área médica em português brasileiro. *Relin*, v. 29, n. 2, p. 959-998,

2021. Disponível em: <http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/17462>>. Acesso em: 10 out. 2023.
- EBLING, Sarah. Building a parallel corpus of German: Swiss German Sign Language train announcements. *International Journal of Corpus Linguistics*, [s. l.], v. 21, n. 1, p. 115-129, 2016.
- GESSER, Audrei. *LIBRAS? Que língua é essa? Crenças e preconceitos em torno da língua de sinais e da realidade surda*. São Paulo: Parábola Editorial, 2009.
- Instituto Federal de Goiás (IFG) – oficial. *Libras - Sintomas da Covid-19*. Facebook, 18 de abril de 2020. Disponível em: <https://www.facebook.com/IFG.oficial/videos/libras-sintomas-da-covid-19/633465714170476/>. Acesso em: 15 ago. 2023.
- HANKE, T. HamNoSys - representing sign language data in language resources and language processing contexts. In: STREITER, O.; VETTORI, C. (ed): *LREC Workshop proceedings: Representation and processing of sign languages*. Paris: ELRA, 2004, p. 1-6.
- JOHNSTON, T. Corpus Linguistics and Signed Languages: No Lemmata, No Corpus. *Paper Presented at the 3rd Workshop on the Representation and Processing of Sign Languages (LREC)*, Marrakech, Morocco, May 2008. p. 82-87. Disponível em: <http://www.lrec-conf.org/proceedings/lrec2008/>. Acesso em: 30 set. 2023
- KARNOPP, Lodenir B. *Produções culturais em língua brasileira de sinais (Libras)*. Letras de Hoje, Porto Alegre, v. 48, n. 1, p. 407-413, 2013.
- KILGARRIFF, A. et al. The Sketch Engine: ten years on. *Lexicography*. Equinox Publishing, 1(1), jul. 2014. p. 7-3. Disponível em: <http://dx.doi.org/10.1007/s40607-014-0009-9>. Acesso em: 10 out. 2023
- LADD, Paddy. *Em busca da surdidade: Colonização dos Surdos*. Cascais: Sud'Universo, 2013. v. 1.
- LAURENCE, A. AntConc (Version 4.2.0) [Computer Software]. Waseda University, 2022. Disponível em: <https://www.laurenceanthony.net/software>. Acesso em: 30 ago. 2023
- MCINTYRE, D.; WALKER, B. *Corpus Stylistics: Theory and Practice*. Edinburgh: Edinburgh University Press, 2019. 376 p.
- PERLIN, Gladis T. T. Identidades Surdas. In: SKLIAR, Carlos. *A surdez: um olhar sobre as diferenças*. Porto Alegre: Mediação, 2021. v. 2, p. 51-73.

- PRADO, M. C. de. *A relevância da Pragmática no ensino do inglês aeronáutico: um estudo baseado em corpora*. 2019. Tese (Doutorado em Estudos Lingüísticos e Literários em Inglês) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2019. DOI doi:10.11606/T.8.2019.tde-16122019-181408. Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8147/tde-16122019-181408/pt-br.php>. Acesso em: 10 out. 2023.
- QUADROS, Ronice M de. *Educação de surdos: A aquisição da linguagem*. Porto Alegre: Artes Médicas, 1997.
- _____.; STUMPF, Marianne Rossi; LEITE, Tarcísio de Arantes (org.). *Estudos de Língua de Sinais Leitura: VI*. Florianópolis: Insular, 2013.
- _____. A transcrição de textos do Corpus de Libras. *Revista Leitura: Línguas de Sinais: abordagens teóricas e aplicadas*, Porto Alegre, v. 1, n. 57, p. 8-34, 2016.
- _____. *Língua de herança: língua brasileira de sinais*. Porto Alegre: ArtMed, 2017.
- _____.; SCHMITT, D.; LOHN, J. T; LEITE, T. de A. *Corpus de Libras*. [S. l.], 2020. Disponível em: <http://corpuslibras.ufsc.br/> 2020. Acesso em: 30 set. 2023.
- REBECHI, Rozane R. God, nation and family in the impeachment votes of Brazil's former president Dilma Rousseff: A corpus-based approach to discourse. *Journal of Corpora and Discourse Studies*, [s. l.], n. 2, p. 144–174, 2019.
- SACKS, Oliver. *Vendo vozes: uma viagem ao mundo dos surdos*. Tradução: Laura Teixeira Motta. São Paulo: Companhia das Letras, 2010.
- SKLIAR, Carlos. *A surdez:: um olhar sobre as diferenças*. Porto Alegre: Mediação, 2001. v. 2.
- _____. *Atualidade da educação bilíngue para surdos: processos e projetos pedagógicos*. Porto Alegre: Mediação, 2013. v. 4.
- STOKOE, William C. Sign Language Structure:: An Outline of the Visual Communication Systems of the American Deaf. *Journal of Deaf Studies and Deaf Education*, [s. l.], v. 10, ed. 1, p. 3–37, 1 jan. 2005.
- STROBEL, Karin. *Surdos: Vestígios Culturais não Registrados na História*. Tese (Doutorado em Educação) - Universidade Federal de Santa Catarina, Florianópolis, 2008.
- TAGNIN, Stella E. O. A Linguística de Corpus na e para a tradução. In: VIANA, Vander; TAGNIN, Stella E. O. (org.). *Corpora na Tradução*. São Paulo: Hub Editorial, 2015. p. 19-56.

TV UFG. *[LIBRAS] COVID-19 | Quais são os principais sintomas?* YouTube, 06 de abril de 2020. Disponível em <https://www.youtube.com/watch?v=XKMTPfuduoM>. Acesso em: 15 de agosto de 2023.