

PROPOSTAS METODOLÓGICAS PARA A ORGANIZAÇÃO DE OBRAS TERMINOGRÁFICAS EM LÍNGUA DE SINAIS BASEADAS EM CORPUS

METHODOLOGICAL PROPOSALS FOR THE ORGANIZATION OF TERMINOGRAPHIC WORKS IN SIGN LANGUAGE BASED ON CORPUS

Joyce Cristina Souza*
Marianne Rossi Stumpf**
Patricia Tuxi***

Resumo: No campo dos Estudos do Léxico, a Terminologia tem sido destaque nos últimos anos, tornando-se objeto de estudo para muitos pesquisadores que trabalham com a Língua Brasileira de Sinais (Libras). Da mesma forma ocorre o crescimento de obras terminográficas. Sob a justificativa da inexistência de sinais que representem os termos e/ou os conceitos desejados, é comum a prática metodológica guiada pelo processo de criação seguido da validação de sinais-termo. Fundamentados nos pressupostos da Linguística de Corpus, apresentamos neste trabalho cinco propostas metodológicas para a realização de pesquisas e organização de instrumentos linguísticos (glossários e dicionários) dedicados às terminologias em Libras, tendo como princípio a valorização do uso que o falante faz da língua em contextos reais de comunicação para a compilação de corpus linguístico autêntico.

* Doutoranda em Linguística pela Universidade Federal de Santa Catarina, mestre em Linguística pela Universidade Federal de São Carlos e licenciada em Letras pela Universidade de Franca. E-mail: joyce.souza@ufscar.br

** Professora do Programa de Pós-Graduação em Linguística (PPGL) da Universidade Federal de Santa Catarina (UFSC). E-mail: marianne@ead.ufsc.br

*** Professora do Programa de Pós-Graduação dos Estudos da Tradução (POSTRAD) e do Programa de Pós-Graduação em Linguística (PPGL) da Universidade de Brasília (UnB). E-mail: ptuxi@unb.br

Palavras-chave: Terminologia; Terminografia; metodologia de pesquisa; Linguística de Corpus; língua de sinais.

Abstract: In the field of Lexical Studies, Terminology has stood out in recent years, becoming the subject of study for many researchers working with Brazilian Sign Language (Libras). Similarly, there has been a growth in terminological works. Under the justification of the absence of signs representing the desired terms and/or concepts, it is common to follow a methodological practice guided by the process of creating followed by the validation of term-signs. Grounded in the assumptions of Corpus Linguistics, this paper presents five methodological proposals for conducting research and organizing linguistic tools (glossaries and dictionaries) dedicated to terminologies in Libras. The guiding principle is the valorization of the speaker's use of language in real communication contexts as an authentic linguistic corpus.

Keywords: Terminology; Terminography; Research Methodology; Corpus Linguistics; sign language.

Introdução

Nas últimas duas décadas, a língua de sinais tem se tornado objeto de estudo em diferentes áreas de investigação no cenário brasileiro. O reconhecimento legal da Língua Brasileira de Sinais, doravante Libras, como meio de comunicação e expressão das pessoas surdas no ano de 2002, ocorreu por meio da Lei Federal nº 10.436 (BRASIL 2002), bem como pela sua regulamentação por meio do Decreto nº 5.626 (BRASIL 2005), o que justifica esse movimento no âmbito acadêmico.

O investimento científico nos estudos linguísticos da Libras tem reverberado no campo das Ciências do Léxico (Lexicografia, Lexicologia, Terminografia e Terminologia), em que se nota uma exponencial produtividade de pesquisas e produtos. Neste trabalho, daremos atenção somente ao tratamento das unidades terminológicas, ou seja, dos termos propriamente ditos, no âmbito da Terminologia e da Terminografia.

Sob a justificativa da inexistência de sinais que representem os termos e/ou os conceitos desejados, ou da falta de materiais que os registrem, é comum a prática metodológica guiada pelo processo de criação de sinais seguido de sua validação. De acordo com CALVET (2007), as situações linguísticas podem ser geridas partindo de uma abordagem *in vivo* ou *in vitro*. Esta última refere-se aos problemas relacionados ao uso das línguas e, para o autor, isso ocorre quando,

em seus laboratórios, linguistas analisam as situações e as línguas, as descrevem, levantam hipóteses sobre o futuro das situações linguísticas propostas para solucionar os problemas e, em seguida, os políticos estudam essas hipóteses e propostas, fazem escolhas, as aplicam (CALVET 2007:70).

A gestão de uma determinada língua, para além do Estado, é de responsabilidade também daqueles que a utilizam em contextos diversos, institucionais ou não. Assim, “as escolhas que acontecem na esfera *in vivo* não podem ser desconsideradas, uma vez que muitas delas ocupam um espaço relevante no pensar sobre a língua” (CORREA E GÜTHS 2015: 144).

Nesse sentido, apartando-nos da tradição metodológica (criação de sinais e validação), apresentamos neste trabalho outros caminhos para a realização de pesquisas e organização de instrumentos linguísticos (glossários e dicionários) dedicados ao léxico especializado da Libras, de modo a pensar sobre a língua partindo da própria língua e de seus falantes, e tendo como princípio a valorização do uso que o falante faz da língua em contextos reais de comunicação.

Para realizar tal tarefa, ancoramo-nos nos pressupostos da Linguística de *Corpus* (LC) que

[...] ocupa-se da coleta e da exploração de corpora, ou de conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador. (SARDINHA 2004:3)

Vale ressaltar que o uso de *corpus* sempre foi um recurso utilizado nas pesquisas linguísticas. A concepção de *corpus* empregada pela Linguística define-o como “um conjunto, tão variado quanto possível, de enunciados efetivamente emitidos por usuários de uma referida língua em determinada época” (DUCROT; TODOROV 2001 *apud* ALUÍSIO; ALMEIDA 2006:157), ou como “um conjunto de textos escritos ou falados numa língua, disponível para análise” (TRASK, 2004 *apud* ALUÍSIO; ALMEIDA 2006:157). Todavia, a concepção de *corpus* passou a ter um novo emprego com o advento da LC.

Para SINCLAIR (2005), pioneiro no trabalho com dicionário compilado a partir de um *corpus* computadorizado,

a corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as

possible, a language or language variety as a source of data for linguistic research (SINCLAIR 2005:4).

Historicamente, as pesquisas envolvendo o campo da Terminologia e da Libras utilizam os estudos da Linguística de Corpus e extraem inicialmente os termos de textos em português, para depois identificá-los ou criá-los em língua de sinais.

Essa identificação, em alguns casos, ocorre por meio da busca terminológica em dicionários ou glossários, e em outros casos, por meio da própria comunidade surda, via entrevistas, formulários, questionários, grupo focal, entre outros.

Considerando as atividades laborais e acadêmicas das autoras proponentes deste trabalho, a saber, duas professoras universitárias, sendo uma surda e uma ouvinte, e uma tradutora intérprete de língua de sinais (TILS) de uma universidade federal; considerando o quadro de servidores (docentes e técnicos) surdos e TILS lotados nas instituições públicas federais de Ensino Superior; e considerando ainda as terminologias utilizadas cotidianamente nas universidades, sobretudo nas atividades administrativas, reuniões e colegiados, decidimos iniciar uma jornada rumo à identificação do vocabulário especializado empregado pelos falantes de Libras nesse contexto.

A fim de explorar novos percursos metodológicos para o fazer terminológico envolvendo a língua de sinais, apoiamo-nos nos preceitos da LC, de modo a observar o comportamento linguístico da comunidade surda acadêmica¹ por meio de evidências empíricas. Mas, antes de iniciar a descrição das etapas percorridas e dos desafios encontrados em cada uma delas, dedicamos a seção subsequente para fazer uma contextualização da pesquisa que nos motivou a investir nesse propósito científico.

1. O contexto da pesquisa

Podemos identificar como ponto inicial do contexto da pesquisa a criação dos cursos de Letras com habilitação em Língua Brasileira de Sinais (Libras).

¹ Entende-se por comunidade surda acadêmica a comunidade composta por falantes da Libras, que compreende professores surdos, professores ouvintes (fluentes em língua de sinais), tradutores intérpretes, técnicos-administrativos e discentes surdos.

Tais formações surgiram como resposta ao Decreto n. 5.626/2005, que regulamenta a Lei n. 10.436/2002 (Lei de Libras), e como garantia para a acessibilidade aos surdos por meio da Libras, conforme previsto na Lei n. 5.296/2004 (Lei de Acessibilidade).

Os cursos oferecem formação em licenciatura ou bacharelado, sendo o primeiro para a formação de professores da área da língua de sinais e o segundo para tradutores e intérpretes de Português - Libras. Segundo Quadros e Stumpf (2014), a sua criação ocorreu com o objetivo de atender às demandas da área de ensino e de formação e inclusão dos surdos.

Os cursos inicialmente foram oferecidos pela Universidade Federal de Santa Catarina (UFSC), na modalidade à distância. Iniciaram no ano de 2006, quando foram abertos polos em Brasília (UnB), Florianópolis (UFSC), Fortaleza (UFCE), Goiânia (CEFET-GO), Manaus (UFAM), Rio de Janeiro (INES), Salvador (UFBA), Santa Maria (UFSM) e São Paulo (USP).

No ano de 2008 novas universidades receberam o curso Letras - Libras, possibilitando, dessa vez, a formação em bacharelado além da formação em licenciatura, e foram abertos polos em Belém (UEPA), Belo Horizonte (CEFET-MG), Brasília (UnB), Campinas/SP (UNICAMP), Curitiba (UFPR), Dourados/MS (UFGD), Florianópolis (UFSC), Fortaleza (UFC), Goiânia (CEFET- GO), Natal (CEFET- RN), Porto Alegre (UFRGS), Recife (UFPE), Rio de Janeiro (INES), Salvador (UFBA) e Vitória (UFES).

Em 2010 ocorreu a titulação das primeiras turmas de formandos na licenciatura e, em 2012, as demais turmas de licenciatura e bacharelado. Entre os anos de 2006 e 2012, formaram-se 312 bacharéis e 378 licenciados.

Atualmente no Brasil há um total de 40 cursos de licenciatura e 8 cursos de bacharelado no formato presencial, e também polos de ensino no formato de Educação a Distância (EaD).

A expansão da oferta de cursos resultou no aumento significativo de pesquisas voltadas para a Língua Brasileira de Sinais - Libras no meio acadêmico. Esse fato teve como fator principal o aumento de surdos como discentes, docentes e técnicos, que passaram a realizar pesquisas da própria língua.

Com a chegada desses novos agentes educacionais e do alunado, desafios singulares emergiram, sobretudo provenientes de situações comunicacionais

administrativas. Tais contextos (ultra)passam questões linguísticas, uma vez que vão desde compreender e se apropriar da linguagem e do léxico especializado, utilizado nas reuniões de colegiado de cursos, departamentais e de outras instâncias, até sistematizar o funcionamento, a logística e a estrutura organizacional da universidade.

Por ser incipiente a presença da Libras e de seus falantes nesses espaços, seria natural que houvesse inicialmente alguma dificuldade relacionada tanto ao âmbito conceitual quanto ao campo das terminologias que abarcam esse novo universo, o que de fato aconteceu.

Além do alunado, há também a questão dos Tradutores e Intérpretes que atuam nas Instituições de Ensino Superior - IES. Tuxi (2017) destaca em sua pesquisa de doutorado momentos de vivências em que ocorriam interpretações (português - Libras) durante reuniões administrativas, quando foi possível constatar o uso contínuo da datilologia² no processo de tradução de termos como **colegiado**, **decano**, **departamento**, **extensão**. O uso da datilologia tem o seu valor já reconhecido em pesquisas, como afirma Castro Júnior (2011); no entanto, o seu uso excessivo e repetitivo revela uma lacuna de sinais em língua de sinais nessa área.

Outro ponto que ficou evidente durante as interpretações de palestras, leituras de atas e reuniões de departamento foi a ausência de sinais para **histórico escolar**, **créditos**, **disciplina optativa** e **disciplina obrigatória**, entre outros. A ausência desses sinais evidenciou a necessidade de a Língua de Sinais Brasileira ocupar um lugar do conhecimento científico e tecnológico apresentado pelo saber da língua majoritária, no caso o português, que permeia os espaços sociais e acadêmicos.

De acordo com Krieger (2001), o termo é uma unidade lexical especializada de determinada área técnica e científica, é objeto de estudo da Terminologia e é descrito e registrado pela Terminografia. Em consonância, para Tuxi (2017), o termo possui uma configuração que o distingue dos itens lexicais comuns da língua geral pelo fato de ser utilizado em discursos

² Datilologia é o uso do alfabeto manual para expressar nome de pessoas, de localidades e outras palavras que não possuem um sinal. Em uma representação de transcrição do português para a LSB é representada pela palavra separada, letra por letra por hífen (FERREIRA-BRITO, 2010: 23).

especializados. Em línguas de sinais há também essa diferença, ou seja, sinais comuns que ocorrem na língua geral e sinais especializados que ocorrem em discursos técnicos e científicos. Nesse sentido, os primeiros são chamados de sinais e os segundos, de sinais-termo, conforme Tuxi (2017:50):

Sinal

1. Sistema de relações que constitui de modo organizado as línguas de sinais.
2. Propriedades linguísticas das línguas dos surdos. Nota: a forma plural - sinais - é a que aparece na composição língua de sinais.

Sinal-termo

1. Termo da Língua de Sinais Brasileira que representa conceitos com características de linguagem especializada, próprias de classe de objetos, de relações ou de entidades. 2. Termo criado para, na Língua de Sinais Brasileira, denotar conceitos contidos nas palavras simples, compostas, símbolos ou fórmulas, usados nas áreas especializadas do conhecimento e do saber. 3. Termo adaptado do português para representar conceitos por meio de palavras simples, compostas, símbolos ou fórmulas, usados nas áreas especializadas do conhecimento da Língua de Sinais Brasileira.

Assim, a diferença entre *sinal* e *sinal-termo* está no uso funcional da língua frente a conceitos abstratos e definições de determinados objetos no campo da especialidade. Tal diferença é percebida no processo de organização e registro dos sinais-termo em linhas de pesquisa voltadas para a Terminologia e Terminografia das línguas de sinais.

Para dar andamento à pesquisa, identificamos trabalhos científicos sobre o processo de criação e validação de sinais. Nesses estudos, os sinais são assim designados: unidades lexicais sinalizadas (ULS), unidades terminológicas sinalizadas (UTS) e unidades terminológicas complexas sinalizadas (UTCS). Para tanto, demos início a um levantamento de teses e dissertações da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelas palavras de busca “terminologia AND libras”. Foram encontrados 41 resultados, sendo 29 pesquisas de mestrado e 12 de doutorado. Elaboramos um quadro para exibir o ano e o tipo de obra, com se vê a seguir.

Quadro 1: Resultado de busca de dissertações e teses no âmbito da Terminologia

| ANO | DISSERTAÇÃO | TESE | ANO | DISSERTAÇÃO | TESE |
|------|-------------|-------|------|-------------|-------|
| 2014 | ----- | 2 | 2021 | 4 | 5 |
| 2013 | 2 | ----- | 2020 | 4 | 1 |
| 2012 | 3 | ----- | 2019 | 7 | 1 |
| 2011 | 1 | ----- | 2018 | 2 | 1 |
| 2010 | 1 | ----- | 2017 | 1 | ----- |
| 2009 | ----- | 1 | 2016 | 2 | ----- |
| 2008 | ----- | ----- | 2015 | 1 | 1 |
| 2007 | 1 | ----- | | | |

Fonte: Elaborada pelas autoras (2023).

A partir da análise dessas obras e da percepção do trabalho que estava sendo realizado nas teses e dissertações levantadas, as autoras sentiram uma necessidade maior de delinear mais propriamente o conceito de *sinai-termo*, incluindo o modo de coleta de dados, sua organização e registro.

Tal pesquisa tomou força no Congresso Nacional de Pesquisas em Linguística de Línguas de Sinais realizado em 2018 na UFSC, quando se percebeu, nas falas de pesquisadores e congressistas, uma variação significativa entre alguns sinais correspondentes a termos utilizados na academia, como por exemplo: *créditos, disciplinas eletivas, disciplinas optativas, conselho, departamento, colegiado*, entre outros, como referido anteriormente. Em interpretação, recorria-se ao conceito ou fazia-se uso da datilologia.

Daí surgiu a presente pesquisa sobre as terminologias da Libras na esfera acadêmica, realizada a partir de uma parceria interinstitucional entre UFSC, Universidade Federal de São Carlos (UFSCar) e UnB, por meio de uma atividade de extensão submetida à Pró-Reitoria de Extensão (ProEx) de uma das instituições envolvidas.

Desse modo, o foco da pesquisa centrou-se no modo de coleta de dados de sinais-termo da esfera acadêmica, sua identificação, sua organização e registro, o que remodelou o trabalho de pesquisa inicial.

Diante disso, definimos como objetivo deste trabalho demonstrar os caminhos que percorremos, do ponto de vista metodológico, para a compilação de um *corpus* composto de sinais-termo utilizados pela comunidade acadêmica surda sinalizante em contextos de uso nas rotinas administrativas geradas pelas universidades.

2. Considerações metodológicas para a organização de um *corpus* em Libras

Ao iniciar a compilação de um *corpus*, é importante atentar para uma série de elementos que pode implicar na qualidade e resultados da pesquisa. De acordo com Biber, Conrad e Reppen (1998:12), “a corpus is a large and principled collection of natural texts”³, todavia tais textos precisam estar alinhados aos propósitos preliminares do estudo.

Tais textos podem provir de um *corpus* já existente e disponível para uso, seja ele de acesso livre ou mediante pagamento, ou de compilação, em que a parte interessada reúne os textos apropriados para compor o seu próprio *corpus*. Neste último caso, deve-se ter em mente que a compilação de um corpus inclui uma série de etapas que envolve o processo de seleção, manipulação, nomeação dos arquivos, permissão de uso e, por fim, a anotação dos textos.

Dada a motivação e objetivo do presente estudo, focaremos na etapa da seleção, que nos serviu como mola propulsora para explorar novos percursos metodológicos em busca de um fazer terminológico envolvendo a língua de sinais, os quais serão brevemente descritos.

É importante definir o tipo de *corpus* que se deseja compilar, bem como o seu tamanho e composição. Vários autores dedicaram-se a propor uma tipologia de *corpus*, dentre eles Sardinha (2004), que apresenta sete critérios: modo (falado, escrito, ambos); tempo (síncrono, diacrônico, contemporâneo, histórico); seleção (de amostragem, monitor, dinâmico, estático, equilibrado); conteúdo (especializado, regional ou dialetal, multilíngue); autoria (de aprendiz ou de falante nativo da língua); disposição interna (paralelo ou alinhado); e finalidade (de estudo, de referência, de treinamento ou teste).

Uma das principais características de um *corpus* diz respeito à sua composição, uma vez que ele deve ser composto por “dados coletados em situações reais de uso da língua para comunicação” (VASILÉVSKI, 2007:50).

³ Um *corpus* tem como princípio uma grande coleção baseada em textos naturais (Tradução nossa).

Outro aspecto igualmente importante é que o referido *corpus* seja eletrônico e esteja salvo em um formato que viabilize a sua leitura por programas de linguística de *corpus*, para a identificação, armazenamento, busca, seleção e recuperação de informações.

Cabe salientar que, até aqui, as referências e autores mencionados se basearam em línguas faladas de modalidade oral-auditiva, com um sistema de escrita devidamente convencionado e compartilhado pela comunidade de falantes dessas línguas. Do ponto de vista operacional, o registro de uma língua dessa natureza pode se dar tanto de maneira oral (via gravação de áudio ou vídeo) quanto de maneira escrita.

No caso deste estudo, por se tratar de uma pesquisa cuja língua objeto é a Língua Brasileira de Sinais, portanto uma língua sinalizada, de modalidade espaço-visual, a primeira observação a ser feita diz respeito à lógica e ao conceito de texto comumente empregado, e também ao registro dessa língua. De acordo com Silva (2019), o registro de uma língua sinalizada implica no emprego de tecnologias como, por exemplo, as mídias em vídeo.

Nesse sentido, considerando o escopo da pesquisa, que abarca as terminologias utilizadas pela comunidade surda acadêmica cotidianamente nas universidades, sobretudo nas atividades administrativas, em reuniões de colegiados, nosso primeiro desafio foi pensar no tipo de texto que serviria aos propósitos da pesquisa. Haveria textos que contemplassem os diferentes gêneros empregados no contexto desejado em Libras para análise? Se sim, onde e como acessá-los?

Como já mencionado na seção introdutória deste trabalho, no intuito de apartar-nos da tradição metodológica utilizada até o presente momento, apresentamos outros caminhos para a realização de pesquisas e organização de instrumentos linguísticos (glossários e dicionários) dedicados ao léxico especializado da Libras, de modo a pensar sobre a língua partindo da própria língua e de seus falantes, tendo como princípio a valorização do uso que o falante faz da língua em contextos reais de comunicação.

A seguir, discutiremos sobre o nosso percurso de investigação, que nos levou a identificar, até o momento, cinco possibilidades metodológicas para a

compilação de um *corpus* de estudo em Libras, as quais serão apresentadas a seguir.

2.1 Metodologia 1 - Baseada na Linguística de Corpus

Sob a justificativa da inexistência de sinais que representem os termos e/ou os conceitos desejados, ou da falta de materiais que os registrem, a prática metodológica tradicional, como já dissemos, vale-se do processo de criação seguido da validação de sinais-termo. Neste contexto, grupos de trabalho buscam criar sinais a partir do entendimento do conceito para, em seguida, validá-los na comunidade de uso ou de pesquisadores terminológicos em geral. A fim de explorar novos percursos metodológicos para o fazer terminológico e posteriormente terminográfico envolvendo a língua de sinais, buscamos nos apoiar nos preceitos da Linguística de Corpus (LC).

Contudo, a carência de pesquisas e, conseqüentemente, de referenciais teóricos bem fundamentados sobre a aplicação e as implicações da Linguística de Corpus nas/das línguas de sinais, bem como a falta de materiais voltados para essa temática, é substancial. Diante desse cenário, decidimos utilizar os princípios da LC empregada para as línguas orais e replicá-los em um estudo envolvendo uma língua de sinais, no caso, a Libras.

Iniciamos o trabalho pela compilação dos textos que iriam compor o nosso *corpus* de estudo, seguido do tratamento relacionado e da extração terminológica.

2.1.1 Seleção e compilação dos *corpora*

Para a obtenção dos termos desejados para este estudo, foram utilizadas Atas de Reuniões de Colegiados das três instituições: Departamento de Libras da Universidade Federal de Santa Catarina - UFSC; Departamento de Psicologia da Universidade Federal de São Carlos - UFSCar e Departamento de Linguística e Línguas Clássicas da Universidade de Brasília - UnB. Tais documentações, abaixo descritas, compuseram, então, um *corpus* inicial de pesquisa. Faz-se

necessário destacar que tais textos foram escritos em português, e não em Libras, ou seja, a identificação dos termos se deu apenas em português.

- UFSCar - 4 (quatro) Atas de reuniões ordinárias do colegiado do Departamento de Psicologia, 1 (uma) Ata de reunião extraordinária do colegiado do Departamento de Psicologia e 4 (quatro) Pautas de reuniões ordinárias do colegiado do Departamento de Psicologia;
- UFSC - 4 (quatro) Convocações para reunião ordinária do colegiado do Departamento de Libras e suas respectivas Atas;
- UnB - 2 (duas) Atas de reunião do colegiado do Departamento de Linguística, Português e Línguas Clássicas do Instituto de Letras, 1 (uma) Ata de reunião do colegiado do Programa de Pós-Graduação em Estudos da Tradução da UnB e 2 (duas) Pautas de reuniões do colegiado do Departamento de Linguística, Português e Línguas Clássicas.

Embora de gêneros textuais diferentes (ata, pauta, convocação), todos os textos, previamente analisados, traziam a terminologia que buscávamos, como uma amostra. Em termos gerais, um *corpus* deve ser essencialmente representativo da linguagem, do idioma ou de uma variedade dele (SARDINHA, 2004). Junto à representatividade, tem-se a extensão do *corpus*; portanto, quanto mais extenso for o *corpus*, mais representativo ele será. Apesar disso, é importante lembrar que

o corpus é uma amostra de uma população cuja dimensão não se conhece (a linguagem como um todo). Desse modo, não se pode estabelecer qual seria o tamanho ideal da amostra para que represente essa população. Uma salvaguarda é tornar a amostra a maior possível, a fim de que ela se aproxime ao máximo da população da qual deriva, sendo, portanto, mais representativa (SARDINHA, 2004:23)

No caso, a população aqui representada pelo *corpus* é a população acadêmica, que faz uso de uma linguagem específica em determinados contextos específicos, como é o caso das reuniões de colegiado. Os textos todos escritos em português trazem termos técnico-administrativos, já acima citados.

Os textos encontravam-se em formato PDF. Para que ficassem legíveis pelo computador e obedecessem aos comandos das ferramentas

computacionais utilizadas em pesquisas no campo da Linguística de Corpus, foi necessário converter os arquivos em formato TXT, que contém apenas caracteres do teclado, isto é, letras, números e símbolos ortográficos. Essa conversão viabiliza a limpeza e a formatação do *corpus*, que compreende a retirada de elementos visuais do texto, tais como gráficos, imagens, tabelas e demais anotações que não fazem parte do corpo do texto, de modo a prepará-lo para ser lido e processado pela máquina. Essa etapa se refere ao tratamento do *corpus*.

Aluísio e Almeida (2006) alertam-nos para um ponto importante ainda na etapa da compilação, que é sobre a proteção da identidade dos participantes e do pedido de direitos de uso dos textos. No nosso caso, os textos foram disponibilizados pelos próprios departamentos envolvidos, mediante aprovação em Conselho. Porém ao converter o arquivo e iniciar a limpeza do *corpus*, decidimos por desconsiderar os nomes próprios das pessoas mencionadas no corpo dos textos das Atas para preservar as identidades e a integridade das partes citadas.

Quando se trata de um texto escrito, tomar esse tipo de cuidado é relativamente fácil e as negociações quanto à preservação da identidade pode ser menos burocrática. O mesmo não se pode dizer sobre um texto em Libras videossinalizada⁴, como é o caso das próximas metodologias.

Feita a limpeza, com o auxílio do *software Ant-Conc*⁵, iniciamos o processo de exploração dos recursos computacionais para a extração terminológica. Esse *software*, conforme é possível observar na imagem a seguir, possui uma série de recursos, dentre eles um campo que quantifica as *Word Types* e as *Word Tokens*, além da ferramenta *WordList*, que serve para gerar listas de frequência de palavras com base estatística, e a *KeyWords*, que é capaz de gerar uma lista de palavras destacadas do *corpus* de análise. Todo o processo foi elaborado a partir dos textos e com eles foi organizada a *WordList*.

⁴ O termo Libras videossinalizada foi cunhado por Silva (2019) em sua tese de doutorado defendida no Programa de Pós-Graduação em Linguística na UFSC, e se refere ao uso da Libras gravada em vídeo, em que o sinalizante estabelece uma relação com a câmera, com ou sem emprego de recursos multimodais.

⁵ *Ant-Conc* é um *software* mais fáceis e úteis para análise textual e linguística de *corpus*. Foi desenvolvido por Laurence Anthony, Professor na Faculdade de Ciência e Engenharia da Waseda University, Japan. Disponível em: <https://tarciziosilva.com.br/blog/analise-de-texto-com-antconc-introducao-arquivos-e-configuracoes/>. Acesso em: 17 nov. 2022.

Figura 2: Área de trabalho do AntConc 3.5.8 2019 utilizado na pesquisa (para Windows)

| A | B | C | D | E |
|-------------|---|----------------|----------------------|----------------------------------|
| Rank | Frequência Absoluta como Nome ou Nome próprio no texto | 1-grama | Classificação | 1-gramas não selecionados |
| 110 | 4 | legislação | termo | |
| 124 | 3 | pesquisa | termo | |
| 130 | 3 | programa | termo | |
| 263 | 1 | evento | termo | |
| 1 | 47 | departamento | termo | |
| 3 | 31 | colegiado | termo | |
| 4 | 31 | reunião | termo | |
| 5 | 30 | comissão | termo | |
| 6 | 28 | docente | termo | |
| 7 | 21 | professor | termo | |
| 8 | 21 | chefia | termo | |
| 9 | 21 | discussão | termo | |
| 10 | 20 | aprovação | termo | |
| 11 | 19 | relator | termo | |
| 12 | 18 | ata | termo | |
| 13 | 18 | parecer | termo | |
| 17 | 14 | pauta | termo | |
| 18 | 14 | | termo | horas |
| 19 | 13 | chefe | termo | |
| 20 | 13 | extensão | termo | |
| 23 | 12 | unanimidade | termo | |
| 26 | 11 | | termo | intérpretes |

Fonte: Acervo das autoras (2023).

Geradas as listas de palavras contidas no *corpus* compilado, ao acionar a função Clusters/N-Grams do referido *software*, identificamos a relação de possíveis candidatos a termo separada de acordo com a sequência de n itens de uma determinada amostra de texto. No caso do *corpus* gerado, obtivemos uma listagem de candidatos a termos uni-grama, bi-grama e tri-grama, composta por 389, 108 e 70 termos, respectivamente.

Figura 3: Listagem de candidatos a termos uni-grama composto por 389 termos

D1 fx TERMOS SELECIONADOS

| | A | B | C |
|----|----------------|------------------|----------------|
| 1 | UNIGRAMAS UFSC | UNIGRAMAS UFSCar | UNIGRAMAS UnB |
| 2 | 106 | 201 | 82 |
| 3 | legislação | abstenção | afastamento |
| 4 | adjunto | adversa | reunião |
| 5 | administração | afastamento | colegiado |
| 6 | afastamento | alocação | unanimidade |
| 7 | apreciação | alunos | apreciação |
| 8 | aprovação | andamento | ata |
| 9 | aprovação | apreciação | curso |
| 10 | aprovações | apreciar | indicação |
| 11 | assiduidade | aprovação | professor |
| 12 | ata | aprovado | aproveitamento |
| 13 | averiguação | aproveitamento | comissão |
| 14 | banca | artigo | estágio |
| 15 | campus | assumir | departamento |
| 16 | capacitação | atestado | chefe |
| 17 | chefe | atividade | licença |
| 18 | chefia | atribuição | plano |
| 19 | colegiado | ausências | chefia |
| 20 | comissão | autoridade | pauta |
| 21 | concelho | autorização | relatório |
| 22 | concurso | avaliação | disciplina |
| 23 | conselho | banca | representante |
| 24 | consolidação | biênio | instituto |
| 25 | contemplada | bolsista | licenciatura |
| 28 | convocação | candidato | informes |

Fonte: Acervo das autoras (2023).

Adotar esse percurso metodológico levou-nos à identificação dos termos conforme desejávamos, no entanto, apenas em português, como já tínhamos previsto ao utilizar esses textos.

Como a língua-objeto deste estudo, reiterando, é a Libras, e ainda que tivéssemos buscado realizar o mesmo trajeto nessa língua, não haveria como compilar textos em Libras videossinalizada que abarcassem o tema ou vocabulário estimado. Não há um banco de atas, reuniões e convocações em Libras. E, por isso, a falta de registro de textos sinalizados em diferentes gêneros, incluindo aqueles que circulam nas universidades em situações administrativas, inviabilizou a continuação da Metodologia 1, pois o intuito aqui era o de tentar reproduzir as etapas de uma pesquisa baseada nos preceitos da

LC, privilegiando a língua de sinais. Para lidar com esse problema, partimos para uma nova tentativa, a Metodologia 2, de que trataremos a seguir.

2.2 Metodologia 2 - *Corpus* “Entrevistas”

Diante dos obstáculos apresentados na metodologia anterior, recorreremos à entrevista como forma de obtenção das informações desejadas, no caso, as terminologias utilizadas pela comunidade surda acadêmica em língua de sinais.

Segundo Lakatos e Marconi (2003:196), “a entrevista é um importante instrumento de trabalho nos vários campos das ciências sociais ou de outros setores de atividades”, já que, por meio dela, é possível realizar investigações, diagnósticos ou tratamentos que tangenciam fenômenos ou problemas sociais.

Embora nosso recorte não tenha sido de cunho social, e sim linguístico, apostamos na entrevista enquanto possibilidade de instrumento para a organização de um *corpus* para análise. Foram estabelecidos alguns critérios de seleção dos participantes, dentre eles, ser surdo, ser professor universitário há mais de três anos e atuar em uma das três instituições proponentes deste estudo (UFSC, UFSCar ou UnB). Seguindo tais critérios, chegamos a 10 (dez) participantes. Do ponto de vista quantitativo, sabemos que esse número se torna ínfimo se considerarmos a quantidade de professores surdos existente no quadro de servidores que compõe o montante das universidades brasileiras.

Por precaução e por experiência de outros estudos, decidimos por realizar um teste piloto. A aplicação do teste foi essencial para alinharmos o instrumento dessa metodologia. Por se tratar de uma pesquisa envolvendo seres humanos, os participantes assinaram o Termo de Consentimento Livre e Esclarecido (TCLE).

Uma das vantagens de utilizar esse instrumento é a possibilidade de dar ao pesquisador a oportunidade de obter dados relevantes e substanciais que não são encontrados ou disponibilizados por meio de outras fontes e documentos. Vale ressaltar que a aplicação desse instrumento pode ser realizada de diferentes maneiras, a depender do tipo de entrevista utilizada e dos objetivos da pesquisa. No nosso caso, optamos pela entrevista não-estruturada, também conhecida por semiestruturada, na qual traçamos uma espécie de roteiro para guiar a entrevista.

O roteiro foi organizado em duas partes: o Eixo I e o Eixo II. O primeiro tratava da caracterização dos participantes, com perguntas relacionadas a informações mais gerais, tais como: o nome da universidade em que atua, em quais cursos ministra aulas, como foi o percurso para se tornar efetivo etc. Já o segundo continha perguntas cujo objetivo era extrair do participante informações sobre a participação e a dinâmica das rotinas administrativas na universidade.

A entrevista-piloto foi realizada em Libras, no entanto, como nosso objetivo principal era extrair terminologias do contexto acadêmico-administrativo de textos “orais” produzidos espontaneamente pelo participante sinalizante, buscamos elaborar as questões do roteiro de modo a não predizer ou influenciar o participante com sinais que sugerissem as terminologias desejadas. As respostas foram dadas de acordo com as experiências e vivências do participante; todavia, ao analisar o *corpus* posteriormente, observamos uma predominância de sinais relacionados ao léxico comum da Libras e poucos sinais candidatos a sinais-termo. Desse modo, tornou-se inviável investir nas outras dez entrevistas com os professores convidados para participar do estudo, haja vista a insuficiência de dados para sustentar uma pesquisa que se propõe a desenvolver um material terminográfico baseado em *corpus*.

Diante desse novo obstáculo, seguimos na tentativa de encontrar uma metodologia que nos possibilitasse organizar ou selecionar um *corpus* em Libras que fosse representativo, de referência e valorizasse o uso que o falante faz da língua em contextos reais de comunicação enquanto *corpus* linguístico autêntico. Avançamos então para a Metodologia 3.

2.3 Metodologia 3 - *Corpus* “Reuniões de Conselho Universitário”

Após duas tentativas consecutivas sem sucesso para o propósito da pesquisa, e considerando o contexto pandêmico da covid-19 em que o trabalho foi realizado, concluímos que ainda não havíamos explorado o bastante os recursos tecnológicos e a Internet.

Vale lembrar que durante a pandemia diversos estabelecimentos e instituições, sobretudo de ensino, migraram suas atividades do presencial para o remoto. No que tange às Instituições Federais de Ensino Superior (IFES), todas as atividades envolvendo ensino, pesquisa, extensão e gestão passaram a ser realizadas online, de maneira síncrona ou assíncrona, por meio de diferentes ambientes virtuais. Seguindo essa mesma lógica, com as reuniões de colegiados não haveria de ser diferente. Para fins de registro, as reuniões de conselho de curso, de departamento, de centro e universitárias eram gravadas, e algumas disponibilizadas em seus respectivos canais na Internet.

O advento tecnológico, bem como o seu desenvolvimento, viabilizou o surgimento de novos tipos de textos, escritos e falados, tais como o e-mail, sala de bate-papo, blogs, revistas e jornais online, mensagens de texto e interações (HUNDT; NESSELHAUF; BIEWER, 2007). Podemos ainda acrescentar essa lista com *podcasts*, *videocasts* e vídeos disponíveis na Internet, em plataformas como *YouTube*, *TikTok*, entre outros.

Corroborando com Silva (2019), o registro de uma língua sinalizada implica no emprego de tecnologias; logo, a compilação de vídeos produzidos em Libras ou com acessibilidade nesta língua, não só podem, como devem ser realizadas por meio de filmagem com foco na língua de sinais. Diante disso, considerando a possibilidade de acessar conteúdos relacionados às reuniões de colegiados online, elaboramos um novo planejamento rumo a uma metodologia que levasse em conta conteúdos disponíveis na Internet para compor um *corpus* com base nos vídeos disponíveis.

Desse modo, compilamos um *corpus* composto por 18 vídeos coletados na plataforma *YouTube*, conforme a Figura 4. Para localizar os vídeos, utilizamos os descritores: i. reunião universitária; ii. reunião de conselho universitário e iii. reunião conselho universidade, seguidos do complemento “em Libras”.

Figura 4: Fontes onde foram coletados os vídeos na plataforma do *Youtube*



| CORPUS | | |
|---|---|--------------|
| TÍTULO | LINK | UNIVERSIDADE |
| Reunião Conselho Universitário (Libras) | https://www.youtube.com/watch?v=AQWjgT3n_c&ab_channel=Laborat%C3%B3riodeEdi%C3%A7%C3%A3odeLibrasUFPI | UFPI |
| Reunião Consuni - UFG - Libras | https://www.youtube.com/watch?v=Ee_Ayly8z0&ab_channel=UFGOficial | UFG |
| Reunião dos conselheiros de ensino, pesquisa e extensão - CEPEX (LIBRAS) | https://www.youtube.com/watch?v=KyxgBeaF0Dk&ab_channel=Laborat%C3%B3riodeEdi%C3%A7%C3%A3odeLibrasUFPI | UFPI |
| Reunião de conselheiros superiores da UFPI (LIBRAS) | https://www.youtube.com/watch?v=p2NfEkbWlV&ab_channel=Laborat%C3%B3riodeEdi%C3%A7%C3%A3odeLibrasUFPI | UFPI |
| Reunião dos conselhos superiores | https://www.youtube.com/watch?v=i-kr3TH0gCo&ab_channel=UfpTV | UFPI |
| Reunião dos conselheiros universitários da UFPI - LIBRAS | https://www.youtube.com/watch?v=ejw2_CBI-cko&ab_channel=Laborat%C3%B3riodeEdi%C3%A7%C3%A3odeLibrasUFPI | UFPI |
| 3ª Reunião Ordinária do Conselho Superior Provisório - CONSUPRO | https://www.youtube.com/watch?v=59D4rTU6Oq&ab_channel=UFNTOficial | UFNT |
| Reunião do Conselho Superior 29/05/2020 | https://www.youtube.com/watch?v=rgBCKvQR&ab_channel=ifes-institutoFederaldoEsp%C3%ADritoSanto | IFES |
| Reunião Conselho Universitário - 08/06/2020 | https://www.youtube.com/watch?v=jPk_nGUMV4&ab_channel=ConselhoUniversit%C3%A1rioUFSC | UFSC |
| Reunião Conselho Universitário - 26/06/2020 | https://www.youtube.com/watch?v=w-ijr1006QA&ab_channel=ConselhoUniversit%C3%A1rioUFSC | UFSC |
| (LIBRAS) Reunião PROGRAD e coordenadores de curso de graduação da UFPR | https://www.youtube.com/watch?v=Fr2OdRkolk&ab_channel=EventosUFPRTV | UFPR |
| CEG - 27/01/2021 - Sessão remota - 1ª parte | https://www.youtube.com/watch?v=M1ARTKotKmg&ab_channel=webTVUFJ | UFJ |
| CEG - 31/03/2021 - Sessão remota | https://www.youtube.com/watch?v=FaGnl2EWlQ&ab_channel=webTVUFJ | UFJ |
| 7ª Reunião Geral Extraordinária do Consepe | https://www.youtube.com/watch?v=NWRif-X667w&ab_channel=UFERNOficial | UFERN |
| CEG - 10/02/2021 - Sessão remota | https://www.youtube.com/watch?v=aG1YfFlTo&ab_channel=webTVUFJ | UFJ |
| Coronavírus - Reunião para debater propostas de retomada do calendário acad | https://www.youtube.com/watch?v=6fonTw5ob8s&ab_channel=UFPRTV | UFPR |
| 5ª Reunião Ordinária do Consepe | https://www.youtube.com/watch?v=MxvVlQQWw&ab_channel=UFERNOficial | UFERN |
| 247ª Reunião Ordinária do Consuni - 2ª Sessão | https://www.youtube.com/watch?v=NudQRmc01A&ab_channel=UFSCarOficial | UFSCar |

Fonte: Elaborada pelas autoras (2022).

De acordo com Sardinha (2004), a extensão do *corpus* comporta três dimensões: o número de palavras, o número de textos e o número de gêneros. Os sites acima apresentados oferecem um número significativo de vídeos, que passamos a chamar de Videotextos, fonte do nosso *corpus*. Dada a especificidade do material e a língua-objeto estimada, identificamos na ocasião aproximadamente duas dezenas de videotextos, aparentemente com bastante insumo para análise.

Embora dezoito não seja um número grande do ponto de vista quantitativo, ainda assim é um número significativo de materiais sobre a área pesquisada. É importante observar também a diversidade da natureza dos vídeos encontrados, haja vista que eles são oriundos de universidades federais diferentes.

Não obstante, alguns desafios emergiram da prática metodológica adotada, desde a seleção dos vídeos para compor o *corpus* até a análise dos vídeos selecionados. O primeiro deles foi o fato de grande parte dos vídeos não apresentarem uma sinalização natural, ou seja, feita por surdos sinalizantes de língua de sinais como primeira língua. Nos vídeos identificados, a comunicação era mediada por intérpretes de Libras. Segundo Sardinha (2004:29), “além de representativo, o *corpus* deve ser adequado aos interesses do pesquisador, que deve ter uma questão a investigar para a qual necessite de um *corpus* específico”. Para a realidade desta pesquisa, a questão levantada pelas

pesquisadoras diz respeito às terminologias utilizadas pela comunidade acadêmica surda sinalizante nas rotinas administrativas geradas pelas universidades, partindo do pressuposto do uso que o falante faz da língua em contexto real de comunicação.

O segundo desafio encontrado foi a qualidade do material. Apesar da quantidade, poucos eram diretamente em língua de sinais, e os que foram identificados em Libras possuíam poucos termos, ou seja, o repertório de análise foi muito restrito e sem grande impacto para a constituição de um *corpus* para a finalidade desejada.

Após a análise do *corpus* coletado, percebemos que por meio dessa metodologia não seria possível atingir o objetivo proposto inicialmente, o que nos fez buscar um novo percurso, que apresentaremos a seguir.

2.4 Metodologia 4 - *Corpus* “Reunião de Departamento”

Observando a trajetória percorrida até aqui, é possível afirmar que existem diferentes formas para se compilar um *corpus*, e que a sua eficácia depende dos objetivos propostos para o trabalho. As estratégias utilizadas nas metodologias anteriores foram úteis parcialmente e nos serviram para identificarmos novas formas de explorar o fazer científico do ponto de vista metodológico. Isso não quer dizer que não possam ser úteis em sua totalidade para outras demandas de pesquisa.

A Metodologia 4, na verdade, foi uma tentativa de continuação do que já havia sido organizado na Metodologia 3. Dentre os vídeos anteriormente encontrados, buscamos identificar reuniões de colegiados em Libras que foram realizadas pelos departamentos das universidades federais envolvidas na pesquisa. Para tanto, foram reunidas seis gravações, cujos participantes (surdos ou ouvintes) eram sinalizantes de Libras, língua oficial de condução da reunião.

De posse do material, nomeamos e armazenamos os vídeos em uma pasta virtual compartilhada entre as autoras, para facilitar o acesso e manuseio dos dados. Em seguida, partimos para o tratamento do *corpus* utilizando o *software*

EUDICO Linguistic Annotator (ELAN)⁶ para analisar o conteúdo dos vídeos e consequentemente para realizar as anotações necessárias para viabilizar o restante do processo. Com as anotações realizadas, demos início à extração, seleção e listagem de sinais-termo.

Infelizmente, nem todas as universidades dispõem de um departamento cujos membros sejam majoritariamente falantes de Libras, nativos ou não, possibilitando que as reuniões possam ser realizadas diretamente nessa língua, sem que haja a mediação de intérpretes.

O fato de não termos tido acesso a registros de reuniões em Libras de outras universidades limitou parcialmente o estudo, no sentido de ampliar as buscas, observar variações, catalogar outros sinais-termo etc.

Ainda assim, até o presente momento, esta foi a metodologia que mais se aproximou do objetivo da pesquisa. Identificamos ao todo cento e setenta e sete (177) sinais-termo, produzidos por falantes da Libras em produções discursivas espontâneas no *corpus* analisado.

Durante as análises, via ELAN, gerou-se a seguinte dúvida entre as pesquisadoras: qual seria a forma mais adequada de anotação do *corpus*? Seria melhor transcrever, traduzir ou glosar o texto sinalizado⁷? Que impacto essa escolha causaria na pesquisa? Considerando estarmos diante de um *corpus* cuja língua objeto é a Libras, de modalidade visual-espacial, as produções corporais envolvem geralmente as mãos, a face e o tronco. Tais produções integram um *corpus* de línguas de sinais e apresentam-se em vídeo (QUADROS 2016).

Para o questionamento levantado acima, optamos por adotar as glosas nas anotações dos sinais-termo encontrados no *corpus* a partir da leitura realizada no ELAN. Essa escolha não impactou negativamente a pesquisa, todavia reconhecemos que o ideal para um trabalho com *corpus* é que se faça uso de uma boa ferramenta computacional que seja compatível com a natureza do texto utilizado (sinalizado, falado ou escrito), e que permita ao pesquisador explorar dados linguísticos para extrair dali ocorrências da língua e realizar

⁶ O ELAN é um *software*, uma ferramenta computacional utilizada para anotar e transcrever manualmente e semiautomaticamente gravações de áudio ou vídeo.

⁷ Por não ser o foco do trabalho não nos ateremos em explicar conceitualmente a diferença e do que trata cada uma dessas operações (transcrever, traduzir e glosar). Aos interessados na temática sugerimos como leitura o texto *A transcrição de textos do Corpus Libras*, de autoria da pesquisadora Ronice Müller de Quadros (2016).

análise linguística em diferentes níveis, a depender do objetivo da pesquisa. Para tanto, faz-se necessário tomar decisões sobre como tratar o *corpus* trabalhado.

A exemplo disso, o grupo de pesquisa do *Corpus* de Libras⁸, ao realizar as transcrições da Libras no escopo do Inventário Nacional de Libras, precisou tomar algumas decisões, que compreendem basicamente os seguintes aspectos:

- 1) O uso do ELAN para a transcrição de dados do *Corpus* de Libras;
- 2) A anotação apenas por meio de glosas de sinais produzidos, exclusão de informações morfológicas com a utilização do ID de cada sinal (evita-se o problema em definir o que constituiria a sentença na língua de sinais);
- 3) A anotação de sinal por sinal de ambas as mãos: mão direita e mão esquerda;
- 4) A tradução livre do texto em Libras para a Língua Portuguesa, no formato de texto com segmentação por meio de sentenças enquanto unidades de sentido (aqui a questão da sentença é determinada pelo sentido, e não por razões sintáticas) (QUADROS 2016:19-20).

Embora adequada, aos olhos das pesquisadoras envolvidas, essa metodologia possuía poucos sinalizadores naturais nos vídeos em Libras. Estamos considerando aqui a representatividade no âmbito institucional e geográfico, haja vista que ainda há poucos docentes, discentes e técnicos surdos, o que indica que o levantamento seria feito por não usuários nativos da língua de sinais. Diante desse cenário, deliberamos por deixar de lado a Metodologia 4 e iniciar uma nova tentativa, que culminou na Metodologia 5.

2.5 Metodologia 5 - *Corpus* “SEFALS”

Após quatro tentativas, válidas para o amadurecimento metodológico da pesquisa, iniciamos uma nova proposta que atendesse ao objetivo da pesquisa, considerando sobretudo a questão da espontaneidade da produção discursiva em língua de sinais e a representatividade do *corpus*.

⁸ O *Corpus* da Língua Brasileira de Sinais (Libras) teve início em 1995, e envolve diferentes projetos de diferentes fontes de fomento, tais como: CNPq, CAPES, IPHAN e NIH, compreendendo dados de fontes diversas e diretrizes para o registro dos dados e metadados em Libras. O *Corpus* de Libras é constantemente alimentado e encontra-se disponível Portal de Libras, www.libras.ufsc.br, no link do *corpus* que pode ser acessado diretamente em www.corpuslibras.ufsc.br.

A ideia da Metodologia 5 surgiu em meio à organização de um evento acadêmico pela Universidade Federal de Santa Catarina (UFSC), uma das instituições proponentes do presente estudo. Em uma das reuniões de alinhamento das pesquisadoras, comentou-se sobre o *1º Seminário de Formação Acadêmica em Libras e Língua de sinais internacional para Doutores e Doutorandos Surdos*, o SEFALS. O evento, como o próprio nome diz, teve como objetivo reunir acadêmicos surdos em nível de doutorado (concluído ou em andamento) para discutir e problematizar questões inerentes à formação e atuação (profissional e acadêmica) dos doutores(as) e doutorandos(as) surdos(as), e sua relação com as suas respectivas instituições, pares ouvintes e surdos (colegas de trabalho) e intérpretes de Libras.

O SEFALS foi realizado presencialmente na UFSC de 14 a 18 de março de 2022. Embora nesse período ainda estivéssemos sofrendo os efeitos e danos causados pela pandemia da covid-19, considerando a liberação das vacinas, a diminuição dos casos e óbitos em âmbito nacional, e os protocolos de segurança adotados nos estabelecimentos e instituições, o evento contou com a participação de 65 professores surdos (doutores/doutorandos) oriundos de diferentes universidades e institutos federais.

Do ponto de vista metodológico, esse era o cenário ideal para identificarmos os sinais-termo utilizados por essa parcela significativa e representativa da comunidade surda acadêmica. Assim, confiantes de que essa seria a metodologia que culminaria em um resultado mais satisfatório, confiável e representativo da terminologia utilizada nas universidades e institutos federais para se referir aos jargões da academia, percorremos as etapas a seguir.

3. Etapas percorridas - Metodologia 5

3.1 Procedimentos pré-coleta

Conforme explicitado, o SEFALS foi escolhido para servir de *locus* para a coleta do material que viria a compor o novo *corpus*. Diante disso, os participantes do evento tomaram conhecimento da pesquisa de forma prévia. Para apresentá-la aos inscitos do evento, foi elaborado um vídeo em língua de

sinais e enviado a cada participante o Termo de Consentimento Livre e Esclarecido - TCLE. Além disso, foi oferecido um canal via WhatsApp para possíveis dúvidas. Após a assinatura dos TCLE, as pesquisadoras iniciaram a organização dos equipamentos necessários para a realização da coleta. Fizeram parte dos materiais:

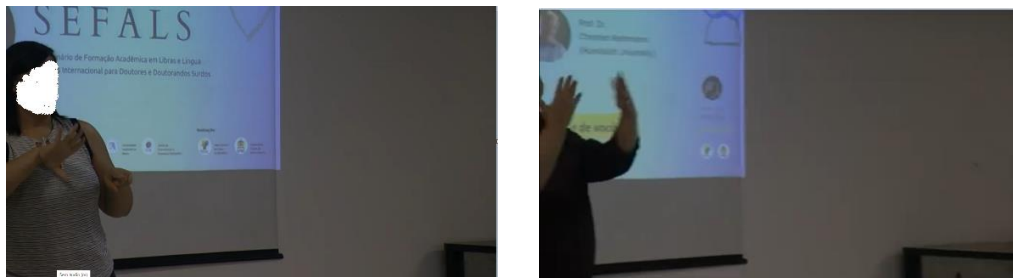
1. Dois tripés para filmadora;
2. Duas máquinas filmadoras de mão; e
3. Três computadores para baixar os vídeos.

Diferentemente das metodologias aplicadas anteriormente, desta vez o *corpus* foi construído e não selecionado. Tal construção demandou uma organização prévia para que, no dia da coleta, todos os recursos estivessem disponíveis para uso. O material foi testado e o seu emprego foi deferido pelos participantes.

Outro fator que vale destacar quando se pretende construir um *corpus* com materiais produzidos por quem realiza a pesquisa é o conhecimento técnico para manusear os equipamentos. Saber explorar os recursos e funcionalidades dos equipamentos impacta diretamente na qualidade do material a ser produzido. No caso de uma pesquisa cuja língua objeto é uma língua de sinais, deve-se redobrar a atenção para diferentes questões, entre elas:

1. **iluminação** - o equipamento de iluminação deve ser favorável para a captura da imagem;
2. **posicionamento das câmeras** - o equipamento de captura de imagem deve estar posicionado em ângulos que possibilitem uma boa e clara visualização do sinalizante. No caso de um evento, como o SEFALS, em que os palestrantes e participantes se movimentam durante suas falas, deslocando-se de um lado para outro, ou mesmo quando se tem mais de uma pessoa alternando os turnos de fala, o olhar atento do pesquisador ou auxiliar responsável pela captura das imagens e vídeos é fundamental; caso contrário, é possível que, ao conferir o material ou analisá-lo, o resultado seja este (Figura 5):

Figura 5 - Fotos do sinalizante fora do enquadramento



Fonte: Acervo das autoras (2022).

3. **conferência dos materiais** - a conferência dos equipamentos envolvidos é essencial. Por exemplo, no caso das câmeras, deve-se verificar cabos (é necessário ter uma extensão?), bateria (está carregada? Há bateria reserva ou não?), tomadas (proximidade e funcionalidade) etc.

Realizados todos os procedimentos necessários para a pré-coleta, o próximo passo para a construção de um *corpus* é a coleta dos dados propriamente ditos. Feito isso, na sequência o pesquisador deve se ocupar com a etapa do pós-coleta.

3.2 Procedimentos pós-coleta

O 1º SEFALS ocorreu dentro do esperado. Tínhamos em mãos um *corpus* composto por dezessete (17) vídeos produzidos em língua de sinais, de forma contínua, espontânea, sem nenhuma interferência da língua portuguesa ou de interpretação.

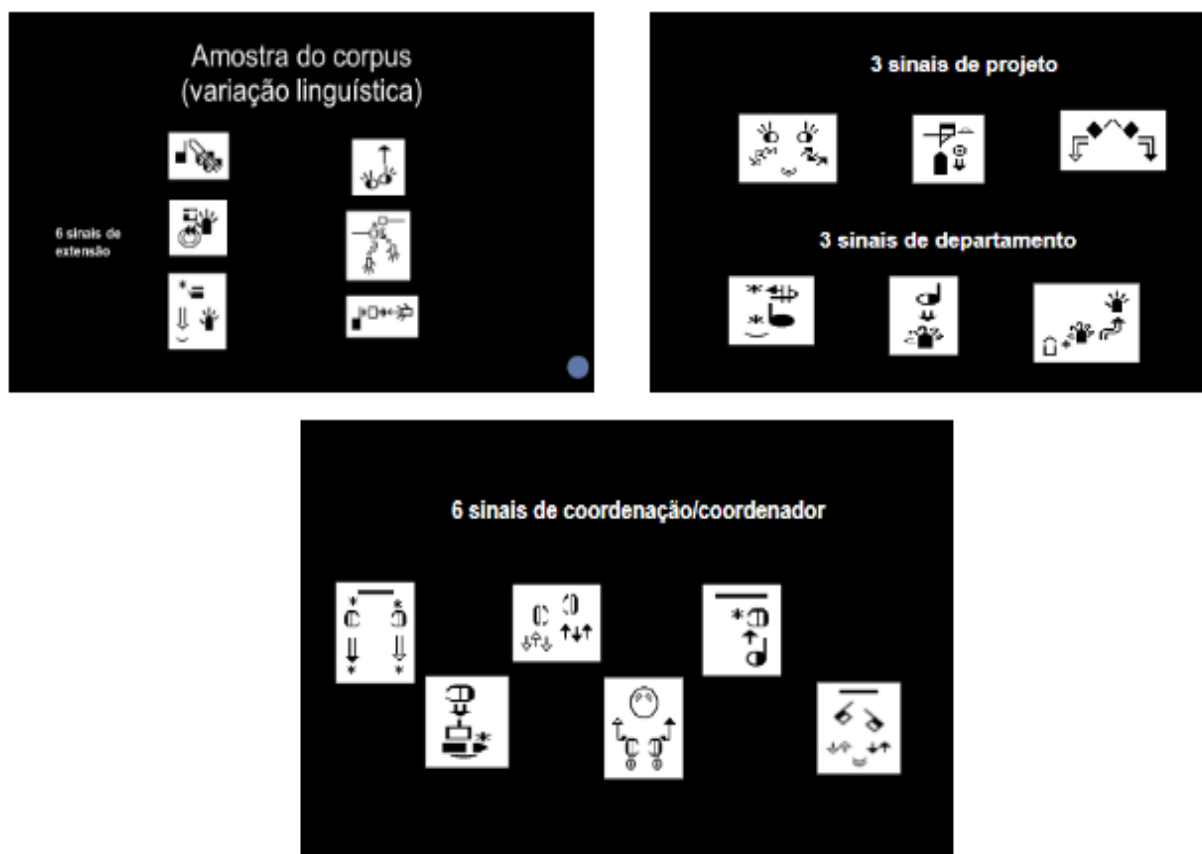
Durante todo o processo foi utilizada a Língua Brasileira de Sinais e Língua de Sinais Internacional - IntSL. O corpo de participantes foi formado por doutores, doutorandos e professores de IES já concursados. O início das gravações ocorreu desde o primeiro dia do seminário, envolvendo as apresentações, os momentos de palestra, discussões e também depoimentos dos integrantes do evento.

Todo o material foi gravado para em seguida ser transferido para os computadores das pesquisadoras. Com os vídeos em mãos, iniciamos o processo de compilação e armazenamento para posterior análise e identificação do conteúdo do material.

As pesquisadoras fizeram anotações com base no *corpus* compilado para buscar sinais-termo dentro dos vídeos armazenados. A partir da identificação de sinais-termo, ocorreram a extração, a seleção e a listagem dos mesmos. Feita a análise do *corpus*, foram encontrados cento e sessenta e três (163) sinais-termo. Considerando a natureza e a riqueza dos dados, pode-se dizer que tais fenômenos não estão circunscritos apenas ao campo da Terminologia. *Corpus* como o do SEFALS servem de insumo para vários tipos de pesquisa, em diferentes níveis de análise (fonológico, morfológico, sintático, semântico, pragmático, discursivo etc.).

Durante a análise linguística do *corpus*, constatamos tanto o uso de sinais-termo com conceitos ambíguos quanto a ocorrência de variação de sinais-termo referindo-se a um mesmo conceito. Neste último caso, notamos que, à medida em que os participantes percebiam o emprego de uma nova forma do sinal-termo, eles o incorporavam na sinalização para promover a compreensão dos demais interlocutores presentes no evento. Em outras palavras, se o sinal-termo não fosse conhecido de todos, era feita uma expansão lexical em referência ao conceito do termo e, a partir disso, passava-se ao uso apenas do sinal-termo. Caso isto não bastasse, lançava-se mão do uso da datilologia ou mesmo da articulação labial (*mouthing*) para identificar o termo (em português) e, assim, oferecer suporte para a compreensão de determinados sinais-termo utilizados.

Figura 8 - Amostra do *corpus* - variação linguística



Fonte: Elaborada pelas autoras (2023).

Realizar pesquisa com uma metodologia que possibilite o cumprimento dos objetivos propostos é a meta de todo pesquisador. No caso da Metodologia 5, conseguimos identificar 163 sinais-termo a partir da base epistemológica preconizada pela LC e a aplicamos, na medida do possível, para observar o comportamento linguístico e terminológico dos enunciados proferidos pelos participantes e palestrantes em língua de sinais em um evento acadêmico. Além da variação linguística, outro fenômeno também identificado foi a criação social de sinais-termo.

3.3 Criação social de sinais-termo na área

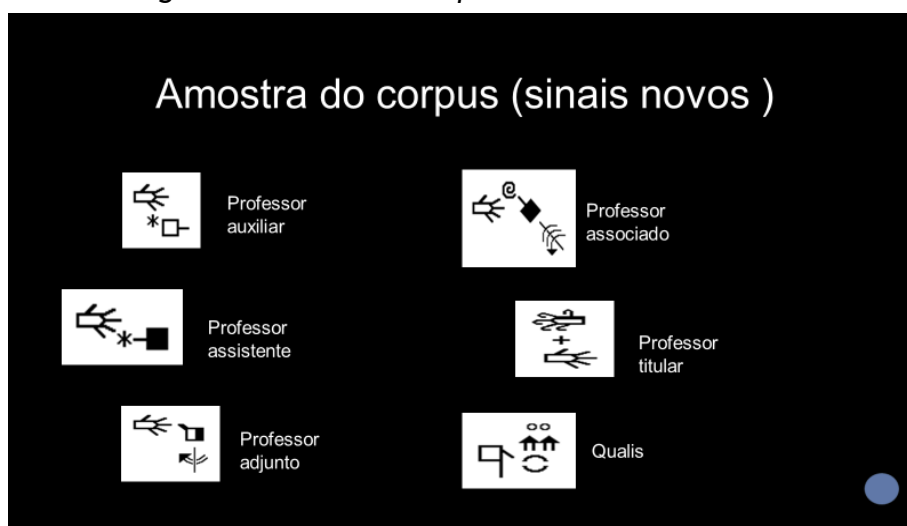
Durante o 1º SEFALS, constatou-se o uso de sinais-termo variados acerca do mesmo conceito, o que indica a criação social, pautada na necessidade de uso, ou seja, na funcionalidade da língua nos espaços acadêmicos e institucionais de ensino.

Segundo Barbosa (1978:185), com base na orientação sociológica, “a origem do signo, assim como sua função, está estreitamente ligada às necessidades sociais do grupo”. O grupo social ao qual estamos nos referindo é

um grupo relativamente seletivo, de doutores e doutorandos surdos, cuja maioria atua profissionalmente como docente em uma universidade ou instituto federal. A experiência social e linguística, bem como a dinâmica dessas instituições de ensino, e as demandas cotidianas das comunidades surdas acadêmicas, juntas, culminam em fontes quase inesgotáveis de unidades lexicais e terminológicas, que nos servem de subsídios para pesquisas linguísticas.

Dentre as unidades terminológicas identificadas no contexto de criação social, segue abaixo, na Figura 9, uma pequena amostra do *corpus* com novos sinais-termo, até então desconhecidos pela maioria dos presentes na atividade.

Figura 9: Amostra do *corpus* - sinais novos



Fonte: Elaborada pelas autoras (2023).

Apesar de a pesquisa ter apresentado resultados significativos, ainda há muitos dados a serem analisados nas gravações realizadas. Como já mencionado na seção introdutória, apresentamos neste trabalho cinco possibilidades para a realização de pesquisas baseadas em *corpus*, da Língua Brasileira de Sinais, que culminará na organização de um instrumento linguístico dedicado ao léxico especializado da Libras, que intitulamos *Dicionário Especializado de Termos Acadêmicos Bilíngue* (DETAB). Tal dicionário será o primeiro a ser produzido envolvendo a Libras, com base em um *corpus* que promove a língua partindo da própria língua e de seus falantes, tendo como princípio a valorização do uso que o falante faz dela em contextos reais de comunicação.

Considerações Finais

Este artigo teve como foco principal explorar percursos metodológicos para o fazer terminológico em língua de sinais com o apoio dos preceitos da Linguística de Corpus, de modo a observar o comportamento linguístico da comunidade surda acadêmica por meio de evidências empíricas.

Fundamentadas nos pressupostos da LC, apresentamos neste trabalho cinco propostas metodológicas para a realização de pesquisas e organização de instrumentos linguísticos (glossários e dicionários) dedicados às terminologias em Libras, tendo como princípio a valorização do uso que o falante faz da língua em contextos reais de comunicação enquanto corpus linguístico autêntico.

Diante das cinco metodologias apresentadas, observamos que existem diferentes formas para se trabalhar com um *corpus*. A eficácia de uma metodologia ou outra depende dos objetivos propostos para o trabalho.

A Metodologia 1 seguiu a proposta teórica da LC, ou seja, com o uso de textos e de softwares que geram listas de palavras. Contudo, essas listas estão em português, que não é a língua objeto de nosso trabalho. Porém, os resultados obtidos a partir dessa metodologia poderão ser a base para um trabalho em Terminologia ou Terminografia bilíngue comparável (*corpus* de textos enunciados por nativos) que envolva Libras e Português em contexto acadêmico, caso de nosso dicionário futuro.

Já a Metodologia 2, com resultados apenas em língua de sinais, teve como instrumento uma entrevista por meio da sinalização em Libras. Apesar dos protocolos relacionados, incluindo as perguntas direcionadas aos objetivos da pesquisa, não houve um registro significativo de sinais-termo. Contudo, foi observado o uso do léxico comum em Libras, incluindo o registro mais informal. Tais dados indicam que essa metodologia poderá ser utilizada para o levantamento, a descrição e o registro de sinais relacionados à língua geral.

Em relação à Metodologia 3, foi organizada uma fonte de vídeos do *Youtube* de reuniões colegiadas em que surdos docentes, discentes e técnicos estavam presentes. Contudo, como em sua maioria os demais membros não dominavam a língua de sinais, a mediação linguística era realizada por Tradutores e Intérpretes de Libras. Tal fato mostra que a sinalização não era,

em grande parte, natural, com uso excessivo de datilologia, o que não era interessante para os nossos objetivos. No entanto, os dados indicam que essa metodologia poderá ser utilizada para pesquisas relacionadas à interpretação, tanto em língua geral como em linguagens especializadas.

Muito similares aos resultados obtidos pela Metodologia 3, aqueles da Metodologia 4 vêm de *corpus* composto por reuniões em IES com mediação por interpretações de ouvintes, que tem Libras como segunda língua ou como língua estrangeira. Poucos dados coletados foram registrados em Libras faladas por surdos. Assim, mesmo que não seja útil para a pesquisa que ora desenvolvemos, a Metodologia 4 também poderá ser utilizada para outros fins, principalmente em contexto bilíngue paralelo, ou seja, de tradução e interpretação.

Finalmente, a Metodologia 5 é a que mais se adapta aos nossos objetivos. Por ela, foram levantados sinais-termo que já circulam na comunidade surda e outros criados socialmente por esses indivíduos, os quais também variam terminologicamente em alguns casos, do ponto de vista denominativo ou conceitual. Tais dados foram gerados a partir de enunciados proferidos por falantes nativos de Libras em contextos reais de comunicação, cujos textos compuseram nosso *corpus* linguístico autêntico no âmbito da Academia.

Desse modo, mostramos primeiramente que é possível e necessário ter como ponto de partida a própria língua de sinais e não o português, histórica e tradicionalmente privilegiado para realizar as pesquisas no campo da Terminologia da Libras. Assim, promovem-se também estratégias de política linguística em prol da manutenção linguística da Libras no contexto vivenciado pelas comunidades surdas acadêmicas. Além disso, viabilizam-se novos mecanismos para a realização de pesquisas em todos os níveis linguísticos (fonológico, morfológico, semântico, sintático, pragmático, discursivo e lexical).

É certo que os resultados aqui apontados devem ser mais explorados em estudos mais aprofundados que possibilitem uma melhor compreensão de como os sinais-termo circulam na comunidade surda. Também a forma de registro e organização deve ser realizada por meio de um *corpus* em Libras. Porém, as metodologias aqui apresentadas podem ser caminhos possíveis rumo a uma nova linha de estudo, a Linguística de Corpus das línguas de sinais. Para tanto,

reiteramos a necessidade de mais pesquisas dedicadas a esse campo de conhecimento e reflexões sobre a temática, haja vista a sua incipiência.

Referências

- ALUÍSIO, S. M., & Almeida, G. M. de B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa linguística. *Calidoscópico*, 4(3), 155-177, 2006.
- BARBOSA, M.A (1978) Aspectos da dinâmica do neologismo. *Revista Língua e Literatura* nº 7. USP.
<https://www.revistas.usp.br/linguaeliteratura/issue/view/10161>
- BRASIL, Decreto n. 5.626, de 22 de dezembro de 2005. Lei n. 10.436, de 24 de abril de 2002, Língua Brasileira de Sinais - Libras [art. 18 da Lei n. 10.098, de 19 de dezembro de 2000].
https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/decreto/d5626.htm.
- BRASIL, Lei n. 5.296/2004 - Regulamenta as Leis nºs 10.048, de 8 de novembro de 2000, que dá prioridade de atendimento às pessoas que especifica, e 10.098, de 19 de dezembro de 2000, que estabelece normas gerais e critérios básicos para a promoção da acessibilidade das pessoas portadoras de deficiência ou com mobilidade reduzida, e dá outras providências. https://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Decreto/D5296.htm
- BRASIL, Lei 10.436, de 24 de abril de 2002. Língua brasileira de sinais - Libras e dá outras providências.
https://www.planalto.gov.br/ccivil_03/leis/2002/l10436.htm
- BIBER, D., Conrad, S., & Reppen, R. *Corpus Linguistics: Investigating Language Structure and use*. New York: Cambridge University Press, 1998.
<http://dx.doi.org/10.1017/CBO9780511804489>.
- BIDERMAN, M.T.C (2001) Léxico e vocabulário fundamental. **ALFA: Revista de Linguística**, São Paulo, v. 40, 2001. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/3994>. Acesso em: 20 out. 2023.
- CALVET, L.-J. *As políticas linguísticas* (l. de O. Duarte, J. Tenfen & M. Bagno, Trans.). São Paulo: Parábola; Florianópolis: IPOL, 2007.
- CASTRO JÚNIOR, G. de. *Projeto Varlibras*. 2014. 259 f. Tese (Doutorado em Linguística) – Universidade de Brasília, Brasília, 2014.
- CASTRO JÚNIOR, G. de. *Variação linguística em língua de sinais brasileira: foco no léxico*. Dissertação (Mestrado em Linguística) - Universidade de Brasília, Brasília, 2011.
- CORREA, D. A., & Güths, T. R. Por um constante repensar de nossas visões sobre língua: revisitando o conceito de política linguística. *Cadernos De Linguagem E Sociedade*, 16(2), 140-159, 2015.
<https://doi.org/10.26512/les.v16i2.7482>.
- FERREIRA-BRITO, L. *Por uma gramática de língua de sinais*. 2. ed. Rio de Janeiro: Tempo Brasileiro, 2010 [1995].
- HUNDT, M., Nesselhauf, N., & Biewer, C. (ed.). *Corpus Linguistics and the Web*. Amsterdam/New York: Rodopi, 2007.

- KRIEGER, M. da G., & Maciel, A. M. B. *Temas de terminologia*. Porto Alegre/São Paulo: Ed. Universidade/UFRGS/Humanitas/USP, 2001.
- LAKATOS, E. M., & Marconi, M. de A. *Fundamentos de metodologia científica*. 5. ed. São Paulo: Atlas, 2003 [1985].
- OTHERO, G. de Á., & Menuzzi, S. de M. *Linguística computacional: teoria & prática*. São Paulo: Parábola Editorial, 2005.
- QUADROS, R. M. de. A transcrição de textos do Corpus de Libras. *Revista Leitura*, Maceió, v. 1, n. 57: 8-34, jan./jun. 2016. Disponível em: <https://www.seer.ufal.br/ojs2-somente-consulta/index.php/revistaleitura/article/view/3618/2853>. Acesso em: 29 nov. 2023.
- QUADROS, R. M. de (org.). *Letras LIBRAS: ontem, hoje e amanhã*. Florianópolis: Ed. da UFSC, 2014.
- QUADROS, R. M. de, & Pizzio, A. L. *Aquisição da língua de sinais*. Florianópolis: UFSC, 2011. https://www.libras.ufsc.br/colecaoLetrasLibras/eixoFormacaoEspecificica/aquisicaoDeLinguaDeSinais/assets/748/Texto_Base_Aquisi_o_de_Linguas_de_sinais_.pdf
- QUADROS, STUMPF (2014). orgs. Série. Estudos de Língua de Sinais. V.II. Florianópolis: Insular. 2014.
- SARDINHA, T. B. *Linguística de Corpus*. Barueri: Manole, 2004.
- SILVA, R. C. da. *Gêneros emergentes em Libras da esfera acadêmica: a prova como foco de análise*. 241f. 2019. Tese (Doutorado em Linguística) - Universidade Federal de Santa Catarina, Florianópolis, 2019. https://repositorio.ufsc.br/bitstream/handle/123456789/214869/PLL_G0782-D.pdf?sequence=-1&isAllowed=y
- SINCLAIR, J. Corpus and Text - Basic Principles. In: Wynne, M. (ed.). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books, 2005: 1-16. <https://ota.bodleian.ox.ac.uk/repository/xmlui/bitstream/handle/20.500.12024/2951/chapter1.htm>.
- TUXI, P. *A terminologia na língua de sinais brasileira: proposta de organização e de registro de termos técnicos e administrativos do meio acadêmico em glossário bilíngue*. 232 f. Tese (Doutorado em Linguística) – Universidade de Brasília, Brasília, 2017.
- VASILÉVSKI, V. Aspectos histórico-teóricos da linguística de corpus: surgimento, abandono, levante e uso. In: Gerber, R. M.; Vasilévski, V. (org.). *Um percurso para pesquisas com base em corpus*. Florianópolis: Editora da Ufsc, 2007: 1-166.