

Construindo critérios de lematização para a Língua de Sinais Brasileira

Building criteria to Brazilian sign language - Libras lemmatization

Antionielle Cantarelli Martins*

José Mario De Martino**

Janice Gonçalves Temoteo Marques***

Francielle Cantarelli Martins****

Resumo: A construção de dicionários, glossários e bancos lexicais de línguas de sinais, tanto impressos quanto *on-line*, requer decisões lexicográficas complexas. Um ponto crítico é a lematização, ou seja, o conjunto de critérios para definir quais grupos de sinais podem compor um mesmo lema. Empregando o *Dicionário da Língua de Sinais do Brasil: A Libras em suas mãos* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017) como *corpus* de análise, este estudo propõe critérios de lematização específicos para Libras nos níveis fonético-fonológico, morfológico e semântico. O estudo almeja oferecer subsídios para construção de *corpora* lematizados da Libras. O emprego de lematização favorece a anotação por glosa, bem como a organização das unidades lexicais e sistemas de busca, além de possibilitar avanços significativos na descrição quantitativa e qualitativa do núcleo lexical da Libras.

Palavras-Chave: Lexicografia; Lematização; Dicionários; Língua de Sinais; Língua de Sinais Brasileira.

* Universidade Federal de Pelotas. Email: an.cantarellim@gmail.com

** Universidade Estadual de Campinas. Email: martino@unicamp.br

*** Universidade Estadual de Campinas. Email: janicetm@unicamp.br

**** Universidade Federal de Pelotas. Email: franciellecantarellim@gmail.com

TradTerm, São Paulo, v.45, p. 147-179

Número Especial - Libras, Lexicografia e Cultura

www.revistas.usp.br/tradterm

Abstract: The construction of sign language dictionaries and lexical database requires complex lexicographic decisions. A critical point is the lemmatization, that is, the set of criteria for defining which groups of signs form a lemma. Using *Dicionário da Língua de Sinais do Brasil: A Libras em suas mãos* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017) as a corpus of analysis, this study proposes a system of lemmatization criteria for Libras at phonetic-phonological, morphological, and semantic levels. This study aims to support the development of Libras lemmatized corpora. The lemmatization process facilitates sign language gloss annotation, as well as the organization of lexical units and search systems, also enables significant advances in the quantitative and qualitative description of the Libras lexical core.

Keywords: Lexicography; Lemmatization; Dictionaries; Sign Languages; Libras.

Introdução

Os estudos linguísticos das Línguas de Sinais são relativamente recentes, iniciados a partir da descrição fonológica da Língua de Sinais Americana -- ASL por William Stokoe (1960). No Brasil, os primeiros trabalhos sobre Língua de Sinais Brasileira - Libras são de Ferreira-Brito (1984, 1990, 1995) e, posteriormente, Quadros & Karnopp (2004). Atualmente há uma série de estudos descritivos da Libras.

Concernente ao registro lexicográfico da Libras na modalidade impressa destacam-se as produções do Laboratório de Neuropsicolinguística Cognitiva Experimental da Universidade de São Paulo -- Lance-IP-Usp desde 2001 sendo a mais recente o *Dicionário da Língua de Sinais do Brasil: A Libras em suas mãos* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017).

Atualmente a Libras passa por um período de rápidas transformações. As conquistas legais da Educação Bilíngue de Surdos, as disciplinas de Libras e o acesso das pessoas surdas nas Instituições de Ensino Superior, a expansão dos serviços de Tradutores/Intérpretes de Libras - TILS têm imposto grande demanda da língua, tanto da parte dos TILS e profissionais da área, quanto pelas próprias pessoas surdas.

Para que registros da Libras estejam acessíveis a consulentes com propósitos variados, é importante que se faça uma sistematização do léxico em repertórios lexicográficos e terminográficos, especialmente em bases de dados eletrônicas. Para isso, é necessário estabelecer um campo aglutinador de variantes, permitindo uma busca a todas elas por meio de um lema.

Trata-se por lema a forma canônica da qual outras variantes são derivadas formando novas palavras ou sinais. Em dicionários, no interior do lema ou entrada, podem ser encontrados sinais relacionados entre si em sua forma e significado (MARTINS, 2017). Dicionários da língua oral têm uma prática amplamente adotada de listar as palavras primariamente pelo seu lema, isto é, *pensou, pensando e pensaram*, não costumam ser arrolados como entradas em dicionários, já que são flexões do lema *pensar*. Dessa forma, as flexões derivadas são chamadas de lexemas.

Logo, por lexema, compreende-se a unidade mínima distintiva do sistema semântico de uma língua que reúne todas as flexões de uma mesma palavra ou sinal (WELKER, 2004), isto é, a unidade que faz referência a uma série de palavras ou sinais e que possuem relação umas com as outras em sua forma e significado.

Para que o léxico de uma língua possa ser registrado e apresentado em dicionários ou banco de dados lexicais, é necessário deflexionar a palavra ou sinal para determinar o seu lema, ou seja, sua forma neutra ou de citação, processo esse denominado lematização. Dicionários lematizados devem conter em cada entrada no mínimo um sinal na forma de citação, ou seja, o lema, representando o lexema, juntamente com as variantes fonológicas e morfológicas.

No contexto deste estudo, lematização é compreendida como o conjunto de critérios para agrupamento de variantes fonológicas, morfológicas e semânticas no mesmo lema. Os critérios de lematização para as línguas de sinais, devem seguir, tanto quanto possível, os critérios de lematização para as línguas orais. Em geral, a forma de citação das palavras apresentada em dicionários das línguas orais é o infinitivo dos verbos, e o singular masculino dos substantivos e dos adjetivos. No entanto, para construção de critérios de lematização para línguas de sinais atenção especial é necessária devido à construção visoespacial. A variabilidade na apresentação da forma e ausência de registro escrito difundido acarretam um desafio adicional para construção de critérios de lematização das línguas de sinais. Enquanto nas línguas orais o lexicógrafo pode ignorar as variantes fonéticas ou os “sotaques”, e registrar o

lema segundo a forma padrão registrada nas gramáticas ou em *corpora* lexicais robustos, o lexicógrafo das línguas de sinais não pode ainda contar com esses recursos.

O processo de lematização é naturalmente um campo fecundo para análises contrastivas entre as variantes fonético-fonológicas e morfológicas. Neste trabalho¹, busca-se contribuir levantando discussões para construção de um modelo teórico com critérios de lematização para Libras em nível fonético-fonológico, morfológico e semântico. O intuito é oferecer subsídio para construção de *corpora on-line* lematizados, dicionários, glossários, sistemas de notação, bem como a análises linguísticas e gramaticais da Libras, especialmente no que tange à compreensão do núcleo lexical dessa língua.

Para isso, os critérios de lematização apresentados no estudo de Fenlon, Cormier e Schembri (2015) para *British Sign Language - BSL* foram analisados e tomados como ponto de partida para construção de um sistema de critérios com princípios de lematização específicos para Libras nos níveis fonológico, morfológico e semântico.

Para ilustrar os critérios de lematização, este estudo emprega o *Dic-Brasil: Dicionário da Língua de Sinais do Brasil: A Libras em suas mãos* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017) como *corpus* para análise por ser uma obra de referência com mais de 13 mil entradas.

A segunda seção deste artigo descreve pontos importantes das mudanças de paradigmas na lexicografia das línguas de sinais proporcionadas pela revolução tecnológica como a possibilidade de construção de *corpora* lexicais *on-line* legíveis por máquina, sistemas de busca e ordenação, vídeos, inclusão de novas informações, *hyperlinks*, glosas e anotação.

Na terceira seção, o estudo introduz os conceitos de léxico e lexema, para finalmente, na quarta seção, apresentar a proposta de critérios básicos de lematização para Libras.

1. Lexicografia das línguas de sinais: impactos

¹ Estudo financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP, Processo: 2021/02365-2

da revolução tecnológica no registro de sinais

Um desafio fundamental para construir dicionários impressos de línguas de sinais é tradicionalmente a modalidade, ou seja, o desafio de descrever uma língua visoespacial em um formato estático (JOHNSTON E SCHEMBRI, 1999). Os desafios compreendem a limitação no sistema de busca, registro estático de uma língua tridimensional e rigidez dos dados. A maioria dos dicionários de línguas de sinais impressos são unidirecionais e ordenam os verbetes alfabeticamente pela modalidade escrita língua oral. Alguns dicionários têm parâmetros fonológicos das línguas de sinais como forma de ordenação; no entanto, essa modalidade de busca não é popular entre os consulentes e é tida como complexa para aprendizes de línguas de sinais como segunda língua (L2). Para possibilitar que o consulente encontre facilmente um sinal por meio da palavra escrita, dicionários impressos costumam repetir entradas com os sinônimos mais frequentes da língua escrita, repetindo assim o mesmo sinal em diversos lemas. Além disso, variantes modificadas de um mesmo lema são frequentemente listadas em entradas separadas, às vezes com algum indicativo de sua relação semântica, fonológica ou morfológica com outra entrada.

Atualmente, a revolução tecnológica proporciona a construção de *corpora* linguísticos legíveis e processáveis por computador, com tecnologia de vídeo, configurando-se em um marco para Lexicografia das línguas de sinais. O software de anotação multimídia *ELAN - EUDICO Linguistic Annotator* é amplamente utilizado entre pesquisadores de língua de sinais. Por meio do programa, as anotações e transcrições podem ser alinhadas diretamente no tempo com um arquivo de mídia que representa um determinado segmento. As anotações sobrepostas também podem ser armazenadas em camadas separadas que representam um nível diferente de análise (CRASBORN E SLOETJES, 2008; NIJMEGEN, 2020).

Por meio desse avanço, a construção de produtos lexicográficos pode ser baseada em um *corpus* robusto, anotado e lematizado. Um *corpus* linguístico é entendido como uma grande coleção de dados de língua falada, escrita ou de sinais com metadados associados, legível e processável por computador.

Basear-se em *corpus*, para desenvolvimento de dicionários, análises linguísticas e gramaticais, possibilita delineamentos atrelados aos padrões de uso, ao invés de basear-se tão somente na intuição de sinalizantes nativos.

Para que um *corpus* de língua de sinais seja legível e processável por computador é necessário que seja anotado. Uma vez que as línguas de sinais carecem de um sistema de escrita amplamente aceito, pesquisadores optam por usar glosas com base contextual para anotação. Johnston (2019) enfatiza que dois tipos de anotação são essenciais para todos os *corpora* de língua de sinais: glosa de identificação e tradução escrita.

A glosa de identificação refere-se à prática de usar uma glosa exclusiva para cada sinal em sua forma de citação, ou neutra, juntamente com todas as suas variantes fonológicas e morfológicas, procedimento relacionado à lematização (JOHNSON, 2019). Em Libras, por exemplo, ao verbo OLHAR é sempre atribuída a mesma glosa independente se ele foi modificado espacialmente para pessoa ou número. A anotação por glosas é consideravelmente mais fácil quando princípios de lematização já tiverem sido delineados e aplicados no *corpus*.

Mesmo com o uso de glosas, princípios de lematização e descrições linguísticas pertinentes, características importantes das línguas de sinais podem ser negligenciadas, como Expressões Não Manuais (ENMs). Para minorar as falhas dos sistemas de notação e obter informações adicionais, a Tecnologia de Captura de Movimento (Mocap) pode ser utilizada para registrar a articulação dos sinais. Por meio do Mocap é possível documentar de forma detalhada a movimentação das juntas do corpo humano durante a sinalização, bem como das expressões faciais e outros aspectos não manuais. Dessa forma, dados adicionais minuciosos antes impossíveis de serem analisados podem estar disponíveis em *corpora* de línguas de sinais. O registro por Mocap, com anotação, recuperação de dados e metadados linguísticos possibilita a documentação da Libras de forma bastante consistente.

O Processamento de Língua de Sinais é um campo de conhecimento emergente de laços com a Ciência da Computação, Linguística de *Corpus* e Linguística da Língua de Sinais. Uma abordagem multidisciplinar, com aparatos

tecnológicos e linguísticos, é bastante apropriada para o desenvolvimento de *corpora* de Libras, dicionários *on-line* lematizados e tradutores automáticos, elementos cruciais para acessibilidade linguística e disseminação da Libras.

Um modelo elucidativo para construção de *corpus* paralelos encontra-se no estudo de Paiva, Barbosa, Faria e De Martino (2017). Com o intuito de construir regras adequadas para o processo de Tradução Automática do Português Brasileiro - Libras por meio de informações morfossintáticas, o estudo gerou *corpus* paralelo por meio da tradução de um livro didático de ciências. O *corpus* paralelo inclui para cada frase do livro o conteúdo em língua portuguesa, transcrição em glosas, os sinais com expressões não manuais. Os sinais foram registrados usando Mocap.

No caso dos dicionários de línguas de sinais, além da vantagem da visualização em vídeo, o formato *on-line* possibilita a inclusão constante de novas informações sem que o lexicógrafo precise optar pelos componentes da entrada, lexemas, variações, definições e exemplos de uso para evitar tamanho exacerbado do livro impresso. Outra vantagem é a construção de redes de significados, já que em todos os exemplos de uso é possível criar *hyperlinks* que levam a entrada respectiva do léxico utilizado no exemplo. Essa rede de significados, proporcionada pelo meio *on-line*, reduz o problema macroestrutural de ordenação dos dicionários de línguas de sinais. Os sinais podem ser pesquisados pelo consulente de forma rápida, flexível e intuitiva: por meio de palavras equivalentes em português (glosas principais e secundárias), pelos parâmetros fonológicos do sinal, por categorias semânticas, ou ainda por *tags* para *status* de uso (arcaico, neologismo, informal e raro). *Tags* e *hyperlinks* com informações gramaticais e propriedades do sinal também podem ser adicionados ao conteúdo das entradas (MCKEE E MCKEE, 2013; MCKEE E VALE, 2017; SCHRIVER, 2003). Espera-se que com outros avanços tecnológicos, em particular das técnicas de reconhecimento de imagens, em breve seja possível a construção de dicionários bidirecionais de línguas de sinais, buscando as entradas por meio de reconhecimento automático de sinais.

Mesmo que alguns desafios da Lexicografia da Língua de Sinais possam ser minorados pelos avanços da tecnologia, a tarefa de registrar o léxico de

uma língua sinalizada em um *corpus* paralelo ou dicionário bilíngue requer tomar decisões baseadas em delineamentos linguísticos. Dessa forma, para que seja possível discutir critérios de lematização, a próxima seção é dedicada a definição de léxico e lexema na Libras, para em seguida propor critérios de lematização.

2. Introdução às definições de léxico e lexemas na Libras

Para a construção de um dicionário, seja ele impresso ou *on-line*, de língua oral ou língua de sinais, definir o que constitui o léxico da língua é crucial. Segundo JOHNSTON E SCHEMBRI (2007), usuários de uma língua possuem um léxico mental com palavras e morfemas daquela língua junto ao seu significado e a outras informações linguísticas, como a classe gramatical. Juntamente com a gramática mental, que contém as regras para combinação de palavras e morfemas para formarem itens lexicais complexos e sentenças, o léxico mental possibilita aos falantes entender e produzir a língua. O léxico individual, no entanto, não pode conter todas as palavras de uma língua, visto que novas palavras são criadas e muitas nem chegam a ser itens estabelecidos na língua.

Para selecionar itens lexicais que devem ser arrolados como entradas em dicionários, o primeiro passo é fazer a distinção entre atos visuoespaciais não linguísticos (gesticulações, gestos e mímicas) de atos visuoespaciais linguísticos (sinais). As entradas arroladas em dicionários devem conter apenas sinais (JOHNSTON E SCHEMBRI, 1999). No entanto, alguns emblemas, gestos mais estáveis típicos das línguas orais, podem ser incorporados no léxico das línguas de sinais, como por exemplo o sinal BEM/BOM ou CERTO apresentado na entrada CERTO! (3) retirada do *Dic-Brasil*:

Figura 1. Entrada do sinal CERTO!



Fonte: retirada do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 62)

O emblema CERTO, usado frequentemente em produções sinalizadas como “TUDO BEM? TUDO CERTO?” constitui um emblema típico das línguas orais que foi incorporado pelos usuários da Libras por meio de lexicalização. Pelo caráter altamente icônico das línguas de sinais, e alta produtividade de mímicas e pantomimas em discursos de sinalizantes da Libras, diferenciar atos visuoespaciais linguísticos de gestos e mímicas pode ser um desafio para lexicógrafos. Um critério inicial que pode ser aplicado são os parâmetros formativos, já que mímicas, pantomimas e gestos frequentemente fazem uso de Pontos de Articulação fora do espaço de sinalização e Movimentos mais longos.

Após fazer a distinção entre atos visuoespaciais linguísticos e não linguísticos, se faz necessário a distinção entre sinais potenciais e sinais reais. Sinais reais são aqueles que a maior parte da comunidade surda conhece, ao contrário do ilimitado número de sinais potenciais possíveis. Observe o sinal **ÁRVORE**:

Figura 2. Entradas do sinal **ÁRVORE**

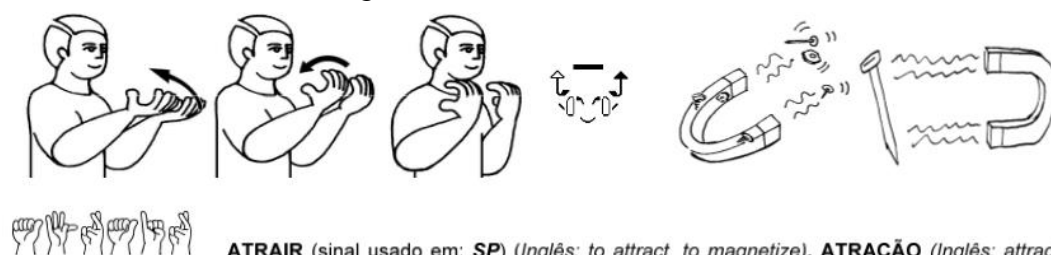


ÁRVORE (sinal usado em: **SP, RJ, MS, DF, PR, SC, CE, RS**) (Inglês: *tree*): s. f. *Vegetal lenhoso e perene, com um tronco principal que se ramifica bem acima do nível do solo (ao contrário do arbusto) e que, em sua maioria, tem mais de 3 metros de altura, podendo chegar até o máximo de 100 metros. É o maior membro da família das plantas. É muito usada para a construção de navios, casas, móveis; assim como na produção do papel. Ex.: Pode-se medir a idade de uma árvore pelo número de anéis concêntricos de que se compõe seu tronco. (Braço esquerdo horizontal dobrado em frente ao corpo, mão aberta, palma para baixo, dedos separados e curvados; cotovelo direito apoiado no dorso da mão esquerda, mão direita aberta, palma para frente, dedos separados. Girar a palma direita para trás, duas vezes.)*

Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 279)

A entrada ÁRVORE se refere indubitavelmente ao referente árvore, o sinal possui um significado identificável mesmo fora de contexto e relativamente estável que é reproduzido com consistência por sinalizadores nativos. Em contraponto, observe o sinal ATRAIR:

Figura 3. Entradas do sinal ATRAIR



ATRAIR (sinal usado em: **SP**) (Inglês: *to attract, to magnetize*), **ATRAÇÃO** (Inglês: *attraction, attractive power, magnetization*): v. t. d. *Exercer atração sobre. Induzir, seduzir, solicitar, trazer para si. Fascinar. Chamar. Prender pela atração. Ex.: Com a sua eficiente explicação, ele atraiu a atenção dos alunos. Atração: s. f. Ação de atrair. Força que atrai. Ex.: O parque de diversões causa atração nas crianças e nos jovens. (Mãos abertas, palmas para cima dedos separados e curvados na altura dos ombros. Mover as mãos para cima, virando as palmas para trás.)*

Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 313)

A entrada ATRAIR, apresenta um sinal que não possui um significado regular e convencionalmente atribuído a ele quando fora de contexto, como puxar ou querer. Pode significar exercer atração sobre algo, puxar um objeto redondo para si, induzir, seduzir, solicitar, trazer para si, fascinar, chamar, prender pela atenção e assim por diante.

Nos sinais caracterizados como potenciais, os aspectos componenciais claramente contribuem com significados específicos do sinal. Embora claramente motivados e muitas vezes icônicos, os componentes não deixam de ser associações convencionais de forma com significado (JOHNSTON E SCHEMBRI, 2007). O sinal apresentado na Figura 3. pode ser classificado como um sinal potencial, que traz desafios adicionais para ser tratado como lexema em dicionários lematizados.

O léxico das línguas de sinais pode ser dividido em subcomponentes que contém vocabulário de sinais nativos e não nativos. Os sinais nativos devem estar de acordo com as regras de nativização, que incluem condição de simetria, de dominância e a tendência a ser monossilábico (ou de no máximo duas configurações de mão na forma de citação). Formas não nativas incluem itens lexicais com soletração manual. O léxico nativo ainda pode ser dividido entre sinais do núcleo e sinais do não-núcleo. Componentes do núcleo são facilmente identificados como lexemas, como os sinais NOITE, AVISAR e AMANHÃ. Já itens lexicais do não-núcleo como os apontamentos e classificadores, impõem desafios adicionais para serem acomodados como entradas em dicionários (JOHNSTON E SCHEMBRI, 2007).

O trabalho seminal de Johnston e Schembri (1999) define **lexema** nas línguas de sinais como um sinal que possui uma forma de citação claramente identificável e replicável, que é fortemente associada a um significado mesmo quando citados fora de contexto, diferente dos sinais potenciais, que possuem significado construído pelos componentes formativos e possuem grande dependência do contexto para construção do significado. Isso quer dizer que um lexema é um sinal que alcança seu significado por meio de convenção. Depreende-se, então, que lexemas possuem divergências entre forma e significado. Essa definição de lexema é bastante estrita, limitando de forma considerável a quantidade de entradas sugeridas a serem arroladas em dicionários.

3. Lematização das línguas de sinais

A tradição de lematização para línguas orais se refere ao processo de agrupar variantes fonológicas e morfológicas em um único lema. É possível que se busque aplicar às línguas de sinais os mesmos princípios de lematização que existem tradicionalmente para as línguas orais, levando em conta as adaptações necessárias a uma língua visuoespacial.

A questão da lematização se torna ainda mais complexa pelo fato de não haver sistema de escrita amplamente utilizado para línguas de sinais. Para línguas orais ortográficas escritas, o lexicógrafo é capaz de ignorar variantes

fonéticas e fonológicas que fazem distinções entre variantes lexicais relacionadas.

Processos de lematização podem parecer simples à primeira vista: se o significado e a forma de dois sinais são diferentes, eles se constituem em dois lemas, portanto, merecem duas entradas diferentes (é o caso de sinais homônimos); se o significado é similar e existem um ou dois parâmetros fonológicos diferentes, os dois sinais estão sob o mesmo lema, portanto, na mesma entrada, sendo considerados sinais polissêmicos. Neste último caso, a diferença dos aspectos formativos pode ser descrita em um subitem na própria entrada (HOCHGESANG, CRASBORN E LILLO-MARTIN, 2018).

O exemplo mais prototípico de banco de dados lexical lematizado é caso do banco de Auslan, disponível primeiramente como um dicionário impresso (JOHNSTON, 1989) lematizado e, em seguida, como um banco de dados *off-line* (JOHNSTON, 2001), e mais recentemente como um dicionário de acesso público *on-line*, e paralelamente como um banco de dados lexical de acesso restrito. O fato de o dicionário de JOHNSTON (1989) seguir princípios de lematização deu ao *corpus* de Auslan uma vantagem que resultou no maior *corpus* de língua de sinais anotado baseado em um banco de dados lexical lematizado. Já os dicionários disponíveis para *British Sign Language - BSL* (por exemplo, BRIEN, 1992) não seguiram práticas de lematização. Como resultado, o trabalho de lematização ocorreu simultaneamente com o trabalho de anotação de um *corpus* de aproximadamente 25.000 sinais (tokens). Isso resultou em um banco de dados de aproximadamente 1.800 entradas (CORMIER, ET AL., 2012). Outros dicionários de Língua de Sinais descritos como fontes lematizadas são: *Dicionário de Língua de Sinais Dinamarquesa* (CENTRE FOR TEGNSPROG, 2008), *Dicionário on-line de Língua de Sinais da Nova Zelândia* (MCKEE E MCKEE, 2013), o *Dicionário de Língua de Sinais Alemã (Deutsche Gebardensprache, DGS)*, o *corpus* lexical em desenvolvimento da Língua Holandesa de Sinais (*Nederlandse Gebarentaal, NGT*) baseado em Konrad et al (2012).

3.1 Construção de Critérios de lematização para Libras

Para a elaboração dos princípios apresentados a seguir, os critérios de lematização do estudo de Fenlon, Cormier e Schembri (2015) foram empregados para desenvolvimento de um sistema de critérios com princípios de lematização específicos para Libras nos níveis fonológico, morfológico e semântico. Com abordagem de pesquisa baseada em Linguística de *Corpus*, o *Dic-Brasil* (2017) foi analisado por meio de inspeção visual para testar, comprovar e exemplificar os critérios previamente elaborados.

Mesmo que se separem os critérios em categorias de formação fonológica, morfológica e de significação dos sinais, é sempre necessário levar em consideração ambos: forma e significado para determinar se dado sinal é um lexema que deve constituir uma entrada com outro lema existente ou constituir uma entrada separada (FENLON, CORMIER E SCHEMBRI, 2015).

3.1.2 Descrição de critérios para análise no nível fonético-fonológicos

Para se determinar o grau com que duas variantes se diferem uma da outra fonologicamente é necessário referir-se aos parâmetros especificados para cada sinal. A partir dos estudos de Stokoe (1960) os pesquisadores de línguas de sinais se referem a Configuração de Mãos (CM), Ponto de Articulação (PA), Movimento (M), Orientação da Palma (OP) e Expressões Não Manuais (ENM) como os parâmetros contrastivos mínimos fonético-fonológicos que se combinam para formação de sinais.

Outros pontos que devem ser levados em conta nesse nível de análise são as *condições de simetria e de dominância* apontadas por Brentari (1998). A *condição de simetria* estabelece que quando as duas mãos se movem e são ativas na realização de um sinal, isso acontece de forma simétrica, sendo que as duas mãos devem ter a mesma CM e movimento espelhado. Nessa condição, alguns sinais podem sofrer a perda de uso da mão não dominante (M2) e em outros sinais pode acontecer ganho de uso da M2. Já na *condição de dominância*, as mãos não partilham as mesmas especificações. Enquanto a mão dominante (M1) pode se mover e possui maior carga de significado, a M2 frequentemente

serve como PA. Em alguns casos, a M2 pode ser omitida durante a articulação do sinal.

Para cada lexema é necessário avaliar se a sua articulação difere significativamente em termos fonético-fonológicos para que se constitua em uma nova entrada. São raros os casos de decisões que dizem respeito a lexemas separados, que sejam feitos em relação a forma unicamente. Majoritariamente, as decisões são feitas baseadas em forma e significado. A seguir, a descrição de critérios para análise no nível fonético-fonológico: a) Variantes fonológicas e variantes lexicais e b) Processos fonológicos.

- a) **Variantes fonológicas e variantes lexicais:** Se dois sinais hipotéticos A e B diferem em somente um parâmetro fonológico (CM, PA, M, OP, ENM) e os significados são os mesmos ou similares, A e B são considerados *variantes fonológicas* e podem constar na mesma entrada. Variações na CM da M2, não precisam ser levadas em consideração para separar sinais em diferentes lemas, caso não altere o significado do sinal. Por exemplo, no *Dic-Brasil (2017)* constam seis entradas para o verbete ABACAXI. Por meio de princípios de lematização, as entradas podem ser agrupadas em duas entradas, ou até mesmo em apenas uma, já que todas as entradas do verbete ABACAXI possuem exatamente o mesmo significado, com variações sutis no parâmetro fonológico: a CM ou OP da M2, que serve apenas como PA, além de alguma variação no movimento, como apresentado na Figura 4, a seguir.

Figura 4. Entradas do sinal ABACAXI

**ABACAXI (1)**

(sinal usado em: **MG, RJ, SP**) (Inglês: *pineapple*): s. m. Fruto grande e espinhento, amarelo avermelhado e de casca dura, muito aromático e saboroso. Originário da América tropical e subtropical, sendo hoje um dos frutos tropicais que alcançam maior difusão

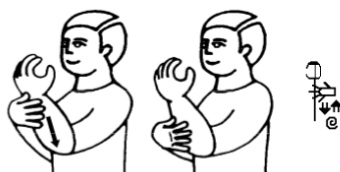
mundial. É indicado para a melhoria geral do quadro respiratório em casos de constipação broncopulmonar. Ex.: O Havai é responsável por três quartos da produção mundial de abacaxi. (Braço esquerdo dobrado diante do corpo, mão esquerda aberta, palma para baixo, dedos separados e curvados; braço direito vertical, cotovelo apoiado no dorso da mão esquerda, mão aberta, palma para cima, dedos separados e curvados.)

**ABACAXI (2)**

(sinal usado em: **BA, PA, RS**): *Idem ABACAXI (1)*. (Mão esquerda vertical aberta, palma para trás; mão direita em **B**, horizontal, palma para trás, tocando o pulso esquerdo. Mover a mão direita para baixo, em direção ao cotovelo.)

**ABACAXI (3)**

(sinal usado em: **MS, RS**): *Idem ABACAXI (1)*. (Mão esquerda vertical aberta, palma para frente; mão direita horizontal aberta, palma para trás, dedos para a esquerda, tocando o braço esquerdo. Mover a mão direita para baixo, em direção ao cotovelo esquerdo, duas vezes.)

**ABACAXI (4)**

(sinal usado em: **SC**): *Idem ABACAXI (1)*. (Mão esquerda em **C**, palma para trás; mão direita horizontal aberta, palma para trás, tocando o dorso do pulso esquerdo. Mover a mão direita para baixo.)



ABACAXI (5) (sinal usado em: **CE**): *Idem ABACAXI (1)*. (Mão esquerda vertical em **S**, palma para trás, mão direita horizontal aberta, palma para trás, tocando o braço esquerdo. Mover a mão direita para baixo, em direção ao cotovelo esquerdo, duas vezes.)

vertical, cotovelo direito apoiado no dorso da mão esquerda, mão aberta, palma para frente, dedos separados. Mover os braços levemente para baixo e para cima, oscilando os dedos da mão direita.)



ABACAXI (6) (sinal usado em: **CE**): *Idem ABACAXI (1)*. (Braço esquerdo dobrado diante do corpo, mão esquerda fechada, palma para baixo; braço direito

vertical, cotovelo direito apoiado no dorso da mão esquerda, mão aberta, palma para frente, dedos separados. Mover os braços levemente para baixo e para cima, oscilando os dedos da mão direita.)

Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 48 E 49)

Como pode ser observado nas entradas (1) e (6) do sinal ABACAXI, a diferença ocorre na CM da M2, sendo a CM da M2 da entrada (1) mão aberta e a CM da entrada (2) em S. Nesses sinais, a M2 é passiva, e serve apenas como PA. Nas duas entradas a CM da mão dominante (M1) é a mesma, bem como o PA

TradTerm, São Paulo, v.45, p. 147-179

Número Especial - Libras, Lexicografia e Cultura

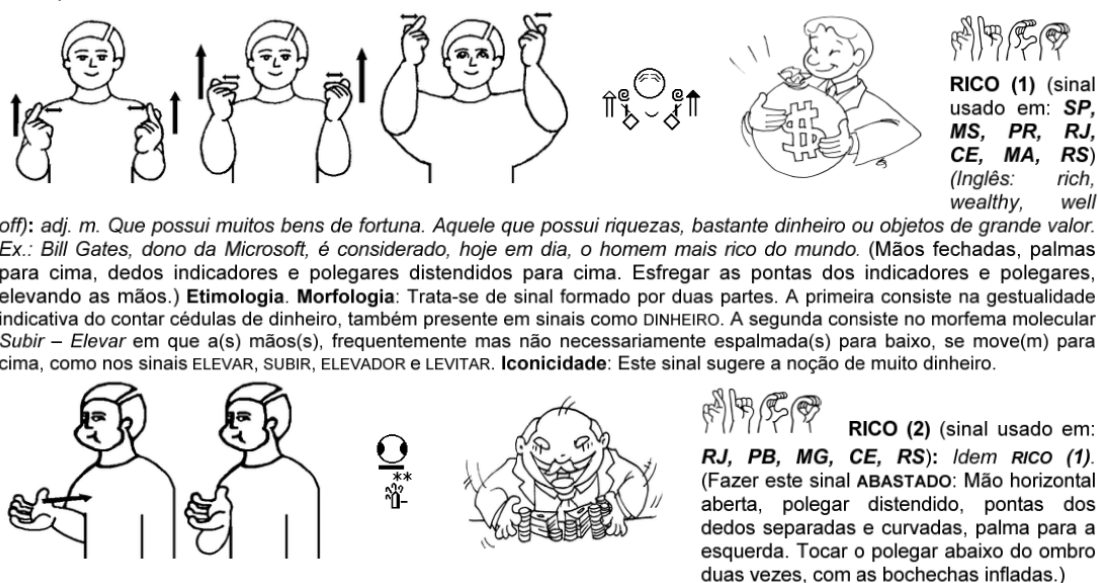
www.revistas.usp.br/tradterm

e OP. Na entrada (6) o sinal apresenta um leve movimento dos braços e oscilação dos dedos, já na entrada (1), o sinal é estático. Portanto, a diferença formativa dessas duas entradas é sutil e poderia ser apontada dentro da mesma entrada. As entradas (2), (3), (4) e (5) também apresentam diferenças fonológicas sutis. O movimento de todos os sinais é o mesmo. A diferença está novamente na CM da M2. Portanto, as diferenças formativas acontecem em apenas um parâmetro fonológico, e essas diferenças fonológicas podem ser mencionadas na descrição da entrada no *corpus* lexical. Dessa forma, por meio de critérios de lematização, as entradas (1) e (6) poderiam ser aglutinadas em uma entrada, e as entradas (2), (3), (4) e (5) em outra.

Dessa forma, se os sinais A e B diferem em mais de um parâmetro fonológico, e o significado é o mesmo ou similar, então, A e B são *variantes lexicais*, ou seja, podem constar em entradas separadas no dicionário ou banco de dados lexical. Pares de sinais com significados relacionados podem também diferenciar-se em relação a outros parâmetros como Movimento (M) e Ponto de Articulação (PA). Como mencionado anteriormente, variações na CM da M2, não precisam ser levadas em consideração para separar os sinais em diferentes lemas, caso não altere o significado do sinal

Na Figura 5, duas entradas do sinal RICO, que possuem parâmetros formativos distintos devidamente apresentados em entradas diferentes no *Dic-Brasil*.

Figura 5. Entradas (1) e (2) do sinal RICO

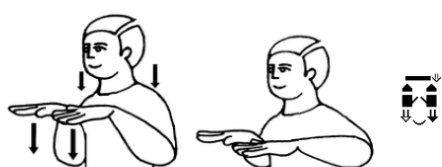


Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 2.458)

Sinais com significado relacionado, mas parâmetros formativos distintos devem ser apresentados em entradas separadas em bancos lexicais.

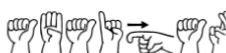
- b) Processos fonológicos:** Neste item, discutimos princípios de lematização relacionados a processos fonológicos típicos das línguas de sinais. Como mencionado anteriormente, na *condição de simetria*, alguns sinais podem sofrer a perda de uso da M2 em outros sinais pode acontecer ganho de uso da M2. Alguns sinais que permitem esse processo podem ser articulados com uma mão ou com ambas com ou sem mudança de significado. Conforme os princípios de lematização de Fenlon, Cormier e Schembri (2015), sinais articulados com uma mão ou com duas mãos são variantes consideradas como constituinte de um mesmo lema, quando o significado não é alterado. Observe a entrada (1) e (2) do sinal ABAIXAR no *Dic-Brasil* conforme mostrado na Figura 6.

Figura 6. Entradas (1) e (2) do sinal ABAIXAR



ABAIXAR (1) (CL) (sinal usado em: **CE, DF, PR, RS, SC SP**) (Inglês: *to lower, to bring down*), **ABAIXAR-SE** (Inglês: *to duck, to stoop down*): Abaixar: v. t. d. Tornar baixo ou mais baixo; baixar. Tornar menos alto. Pôr em lugar mais baixo. Descer. Ex.: Abaixar as

persianas para reduzir a claridade. Abaixar-se: v. pr. Tornar-se baixo ou mais baixo. Baixar-se. Tornar-se menos alto. Pôr-se em lugar mais baixo. Descer. Ex.: Não fique em pé, abaixe-se! (Mãos abertas, palmas para baixo, na altura do peito. Movê-las para baixo, baixando ligeiramente o corpo.) **Etimologia. Morfologia:** Trata-se de sinal formado por morfema metafórico molar que representa atributos físicos de coisas e de relações entre coisas, como nos sinais PEQUENO, ESTREITO, MOLE e DURO. **Iconicidade:** Este sinal consiste na pantomima do abaixar, um gesto de amplo uso. Nele, o sinalizador tem as mãos abertas, então ele as move para baixo, enquanto baixa ligeiramente o corpo.



ABAIXAR (2)

(CL) (sinal usado em: **CE, DF, MG, MS PR, RS, SC**): *Idem ABAIXAR (1)*. (Mão aberta, palma para baixo, acima do ombro direito. Movê-la para baixo.) **Etimologia. Morfologia:** *Idem ABAIXAR (1)*. **Iconicidade:** Este sinal consiste em emblema de uso universal. Nele, o

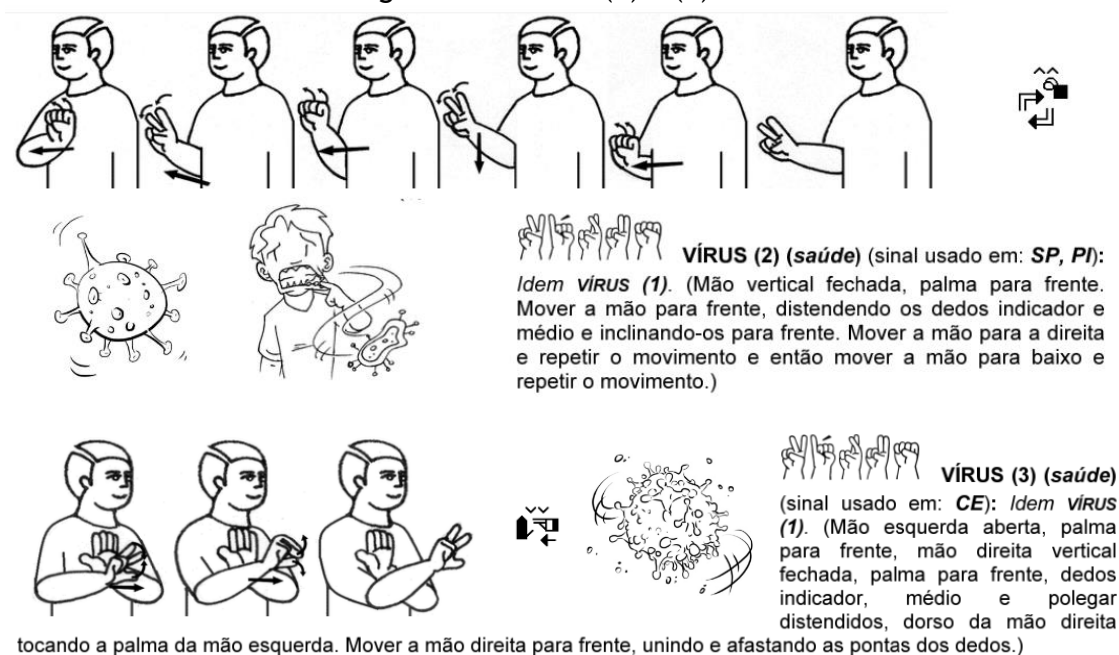
sinalizador começa com a mão aberta, então ele a move para baixo, dando a ideia de reduzir e abaixar.

Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 50)

Na entrada ABAIXAR (1) o sinal é articulado com ambas as mãos, enquanto na entrada ABAIXAR (2) o sinal é articulado com uma mão. O significado de ambos é no mínimo altamente relacionado, se não o mesmo. Dessa forma, seguindo princípios de lematização, em banco de dados lexicais *on-line*, os lexemas podem fazer parte da mesma entrada. Ressalta-se a importância de que os sinais sejam analisados caso a caso, já que não é possível afirmar que todos os pares de sinais dessa categoria sejam intercambiáveis em todos os contextos.

Outro princípio de lematização de processos fonológicos está relacionado a *regra de dominância* apontada por Brentari (1998). Em alguns casos, a M2 pode ser omitida durante a articulação do sinal. Exemplo desse processo é apresentado na entrada de VÍRUS (2) E VÍRUS (3), conforme apresentado na Figura 7.

Figura 7. Entradas (2) e (3) do sinal VÍRUS



Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO & MARTINS, 2017: 2877 e 2878)

Pode ser observado que todos os parâmetros formativos da M1 de ambos os sinais são iguais. No entanto, na entrada Vírus (3) a mão de apoio ou M2 não aparece na articulação do sinal. Ambos os sinais podem constar na mesma entrada em bancos de dados lexicais, já que o significado é exatamente o mesmo.

3.2. Descrição de critérios para análise no nível morfológico

O nível morfológico foi analisado em duas grandes categorias com respectivos critérios de análise com base nos estudos de Fenlon, Cormier e Schembri (2015): modificação de sinal e formação de sinal.

A prática de lematização por **modificação de sinal** envolve o agrupamento de variantes morfológicas que são resultado de modificações de sinais convencionais existentes que aumentam/mudam o significado do radical de maneira previsível, tanto nos aspectos da forma quanto semanticamente, em conjuntos de sinais. A lematização por modificação do sinal é apresentada

em quatro categorias: a) Marcação de aspecto, b) Modificação direcional, c) Número e marcação distributiva, e d) Intensificação.

Já a **formação de sinal** se refere às formas de sinais que são claramente relacionadas a outros sinais. No entanto, o efeito dessa modificação pode expressar um significado que pode não ser totalmente previsível. Nesses casos, frequentemente se justifica reconhecer dois lemas diferentes. A lematização por formação do sinal é apresentada em quatro categorias: a) Distinção substantivo e verbo, b) Incorporação numeral, c) Incorporação negativa, d) Compostos e colocações.

3.2.1. Descrição de critérios para análise no nível morfológico: Modificação de sinal

- a) **Marcação de aspecto:** A marcação de aspecto é normalmente usada com verbos ou outros predicados para caracterizar como uma ação, evento ou estado acontece ao longo do tempo. Por exemplo, o sinal “andar” pode ser modificado para algo como “andar à toa”, e ainda “andar cambaleando”. Essas modificações são previsíveis; em bancos lexicais *on-line* esses sinais podem pertencer ao mesmo lema. Na Figura 8, são apresentadas as entradas de ANDAR (1), na sua forma de citação, e suas possíveis modificações nas entradas ANDAR À TOA e ANDAR CAMBALEANDO.

Figura 8. Entradas do sinal ANDAR, ANDAR A TOA e ANDAR CAMBALEANDO



saudável andar. (Mão em V invertido, palma para trás, dedos para baixo. Mover a mão para frente, oscilando alternadamente os dedos.) **Etimologia. Morfologia:** Trata-se de sinal formado por: 1) um morfema metafórico molar que representa concretamente características conspicuas do comportamento humano em atividades de jogos, esportes e lazer, emulando a pantomima envolvida, 2) o morfema *Pessoa*, e 3) o morfema *Irradiação* (*Propagação - Fluxo*). O morfema metafórico molar está presente nos sinais PESCAR e NAVEGAR. O morfema *Pessoa* é codificado pela primeira articulação (mão em 2 com os dedos apontando para baixo aparece sempre associada a movimento dos dedos ou da mão), como nos sinais DE PÉ, ESCORREGAR e BÉBADO. O morfema *Irradiação* é codificado pela(s) mão(s) aberta(s) balançando e (ou) dedos balançando, frequentemente enquanto a(s) mão(s) se move(m) para uma direção, como nos sinais MULTIDÃO e NEBLINA. **Iconicidade:** Neste sinal, a mão se move para frente enquanto os dedos indicador e médio balançam para frente e para trás como se representasse o movimento das pernas ao caminhar. Este sinal é típico da gestualidade universal.



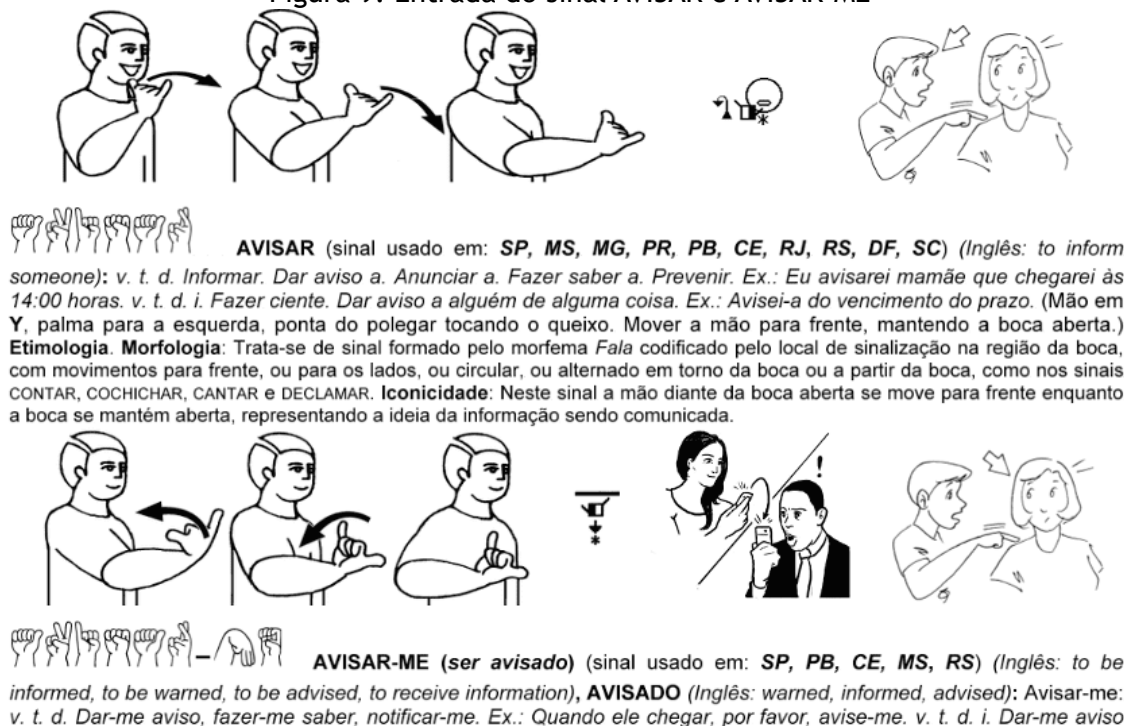
SP (Inglês: to walk a staggering way, to barge along): expressão. Andar de maneira desengonçada. Ex.: O mendigo andava cambaleando pela rua. (Fazer este sinal ANDAR, balançando o corpo para os lados enquanto deixa cair os ombros alternadamente, com expressão displicente: Mão esquerda aberta, palma para cima; mão direita em V invertido, palma para trás, dedos apontando para baixo, tocando a base da palma esquerda. Mover a mão direita para frente, sobre e além da esquerda, oscilando os dedos.) **Etimologia. Morfologia:** Trata-se de sinal formado pelo morfema *Pessoa* em sua primeira articulação (mão em 2 com os dedos apontando para baixo), que aparece sempre associada a movimento dos dedos ou da mão, como nos sinais DEITAR-SE, AJOELHAR-SE, PASSARELA ELEVADA, e ATROPELAR. **Iconicidade:** Este sinal representa a ideia do andar do bêbado.

Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 196 E 197)

Como pode ser observado na Figura 8, as entradas ANDAR À TOA e ANDAR CAMBALEANDO são modificações previsíveis do verbo andar, que podem constar na mesma entrada em bancos lexicais *on-line*. Salienta-se que a decisão lexicográfica de registrar os sinais em entradas diferentes merece atenção especial.

b) **Modificação direcional:** Alguns verbos podem ser espacialmente modificados para que sejam direcionados para os referentes de seus argumentos fisicamente presentes ou a locais associados aos referentes. Nesses casos, a mudança de direção geralmente não altera o significado geral do sinal (ou seja, qualquer que seja sua direção, o significado do sinal permanece o mesmo), não há necessidade de separar o *status* do lexema para cada variante direcional (FENLON, CORMIER E SCHEMBRI, 2015). Na Figura 9, há o exemplo do sinal AVISAR e AVISAR-ME. Essas duas variantes podem estar na mesma entrada seguindo princípios de lematização em banco de dados lexicais *on-line*, já que as alterações são previsíveis.

Figura 9. Entrada do sinal AVISAR e AVISAR-ME

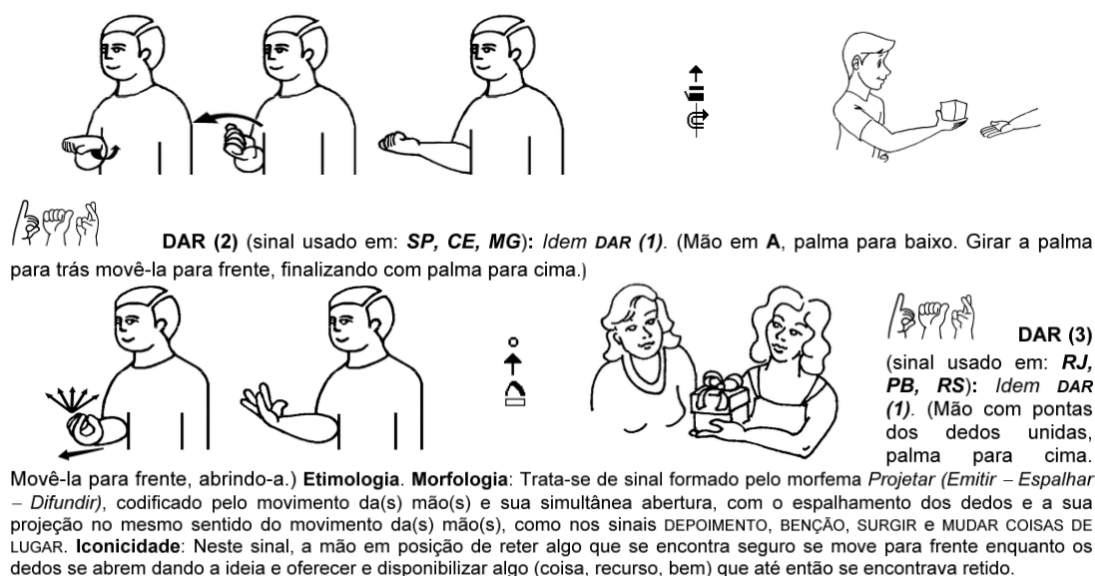


Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 333)

c) **Número e marcação distributiva:** As línguas de sinais têm várias formas disponíveis para marcar número. Para substantivos, isso inclui a repetição do sinal em questão com um movimento lateral curto. Verbos também podem ser modificados por meio de dois tipos diferentes de marcação: múltipla ou exaustiva. A múltipla envolve repetições da articulação do verbo em várias localizações no espaço de forma

distributiva. Já a exaustiva ocorre em um único movimento longo pelo espaço de sinalização e normalmente indica plural. (QUADROS E KARNOPP, 2004; JOHNSTON E SCHEMBRI, 2007). A Figura 10 evidencia o exemplo das entradas DAR (2) e DAR (3).

Figura 10. Entrada do sinal DAR (2) E DAR (3)

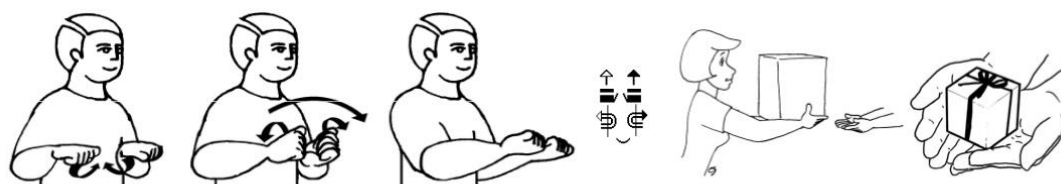


Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 849)

Pode-se perceber pela diferença de articulação dos sinais que a entrada DAR (3) transmite a ideia de dar para mais de um objeto. A modificação é previsível; portanto, não necessariamente devem estar em uma entrada separada em banco de dados lexicais *on-line*.

Outra maneira pela qual os sinalizadores podem marcar vários referentes é usando as duas mãos em vez de uma (JOHNSTON E SCHEMBRI, 1999). Alguns sinais articulados com apenas uma mão (normalmente verbos) podem ser produzidos usando ambas as mãos para indicar mais de um referente (e, em alguns casos, para dar um significado recíproco ao verbo); também nesse caso, já que a mudança de significado não é significativa, não devem ser tratadas como diferentes lemas. Assim, o *Dic-Brasil* registra ainda outra entrada para o sinal DAR (Figura 11) articulado com as duas mãos expressando o significado de dar algo para mais de uma pessoa.

Figura 11. Entrada do sinal DAR (1)

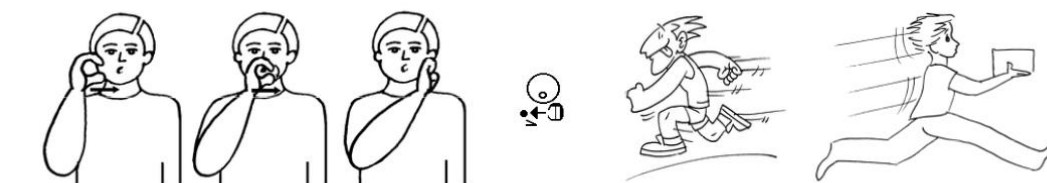


DAR (1) (sinal usado em: **SP, CE, PR, RS**) (Inglês: *to give, to offer, to present, to grant*): v. t. d. *Ceder gratuitamente, fazer doação. Entregar. Permutar. Doar, oferecer. Ex.: Dei o dicionário de presente para as crianças da creche.* (Mãos em **A**, palmas para baixo, lado a lado. Girar as palmas para trás, e movê-las em um arco para frente (sentido horário), finalizando-as com as palmas para cima.)

Fonte: retirada do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 849)

d) Intensificação: O movimento em alguns sinais pode ser modificado de forma a intensificar o significado. Por exemplo, o sinal RÁPIDO (Figura 12) pode ser modificado para que o início do sinal seja mantido por mais tempo do que o normal, significando que algo é "muito rápido". De um modo geral, esse tipo de modificação não precisa ser inserida em um novo lema, já que a mudança de significado é previsível. Em outros casos, o significado de um sinal pode ser intensificado usando as duas mãos, e como mencionado anteriormente, o uso de ambas as mãos como marcador de intensificação não é tratado aqui como um diferente lema.

Figura 12. Entrada do sinal RÁPIDO (1)



RÁPIDO (1) (sinal usado em: **SP, MS, PR, RJ, CE, PB, RS, DF**) (Inglês: *fast, quick, swift, speedy, rapid*): adj. m. *Veloz, ligeiro, breve. Que se move, que corre ou que passa rapidamente. Ex.: Nosso corredor foi o mais rápido de todo o torneio.* (Mão em **C**, palma para a esquerda ao lado direito da boca. Mover a mão para a esquerda, rapidamente, fechando-a em **S** vertical.)

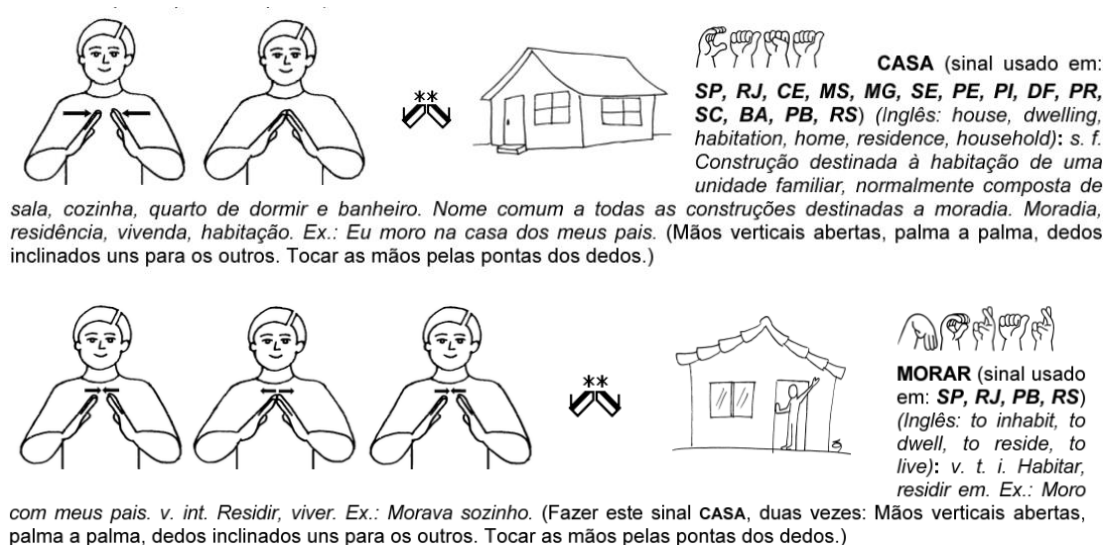
Fonte: retirada do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 2385)

3.2.2. Descrição de critérios para análise no nível morfológico: Formação de sinal

a) Distinção substantivo e verbo: Um exemplo bem conhecido de formação de sinal referido na literatura é a diferenciação de movimento observada em pares de substantivos e verbos relacionados em algumas

línguas de sinais incluindo a Libras (QUADROS E KARNOPP, 2004). Nesses pares, o movimento associado ao substantivo é frequentemente encurtado e repetido, enquanto o movimento associado ao verbo é mais longo e não repetido. Pode-se observar que as entradas, CASA e MORAR, diferem-se apenas pela repetição do movimento, como mostra a Figura 13.

Figura 13. Entrada do sinal CASA e MORAR



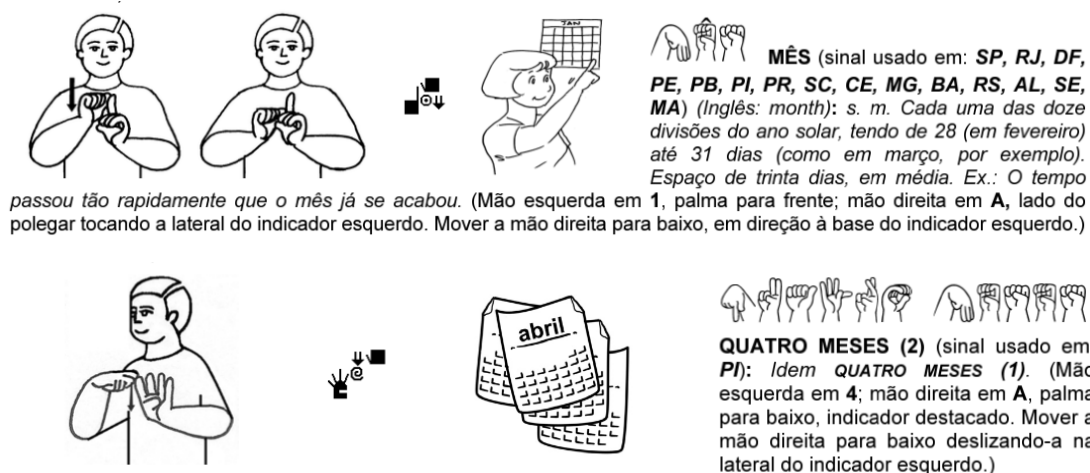
Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 590 E 1890, respectivamente)

Essa é uma decisão lexicográfica complexa. Alguns pares de substantivos e verbos parecem diferir sistematicamente com base apenas no movimento, outros não parecem ter uma diferença sistemática, como VOAR e AVIÃO. A literatura aponta que análises mais aprofundadas utilizando dados de *corpus* são necessárias para tomar decisões de lematização para alternância substantivo e verbo (FENLON, CORMIER E SCHEMBRI, 2015).

b) Incorporação numeral: Alguns sinais podem incorporar sinais numéricos (QUADROS E KARNOPP, 2004). Em alguns casos, o sinal subjacente é considerado um morfema vinculado combinado com um número para ser um sinal totalmente articulado. Exemplos desse fenômeno na Libras é o sinal MÊS, um mês, dois meses, sendo possível até quatro meses. Os

critérios comumente utilizados são de não separar os lemas.

Figura 14. Entrada do sinal MÊS e QUATRO MESES

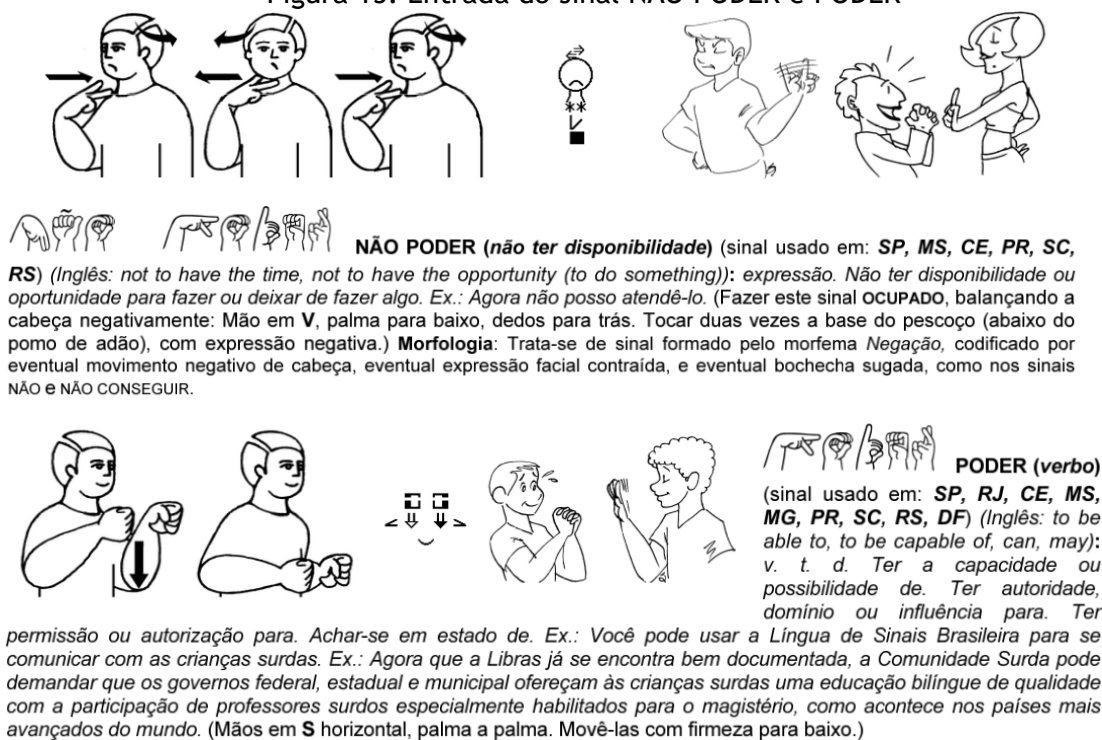


Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 1834 E 2361, respectivamente)

Existem diversas entradas no *Dic-Brasil* (2017) com incorporações numerais, como HORA, DIA, MÊS e ANO. Essas incorporações podem aparecer na mesma entrada de bancos lexicais.

- c) **Incorporação negativa:** Outro processo formativo nas línguas de sinais é a incorporação negativa. Uma das formas é o movimento de cabeça, a outra é o indicador distendido, palma para frente, balançando para os lados, gestos típicos das línguas orais. Nesses casos, pela previsibilidade da modificação, os sinais podem ocupar a mesma entrada. No entanto, em sinais como PODER e NÃO PODER que são articulados de formas completamente diferentes, entradas separadas podem ser criadas, visto que as transformações não são previsíveis. A Figura 15 ilustram essas entradas devidamente separadas.

Figura 15. Entrada do sinal NÃO PODER e PODER



Fonte: retiradas do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 1929 E 2212, respectivamente)

Existe ainda outro grupo de sinais, em que a alteração da direção do movimento parece ser indicativa de negação, como **NÃO QUERER** ou **NÃO TER**. Estudos com *corpora* robustos são necessários para encontrar previsibilidade nesse grupo de sinais, e dessa forma aglutiná-los em um mesmo lema.

- d) **Compostos e colocações:** Compostos nas línguas de sinais se referem ao processo de criação de novos sinais com dois ou mais morfemas livres e é um processo comum bem documentado (JOHNSTON E SCHEMBRI, 1999; RODERO-TAKAHIRA, 2015). Em muitos casos, os compostos são considerados lemas separados. Na Figura 16 há o exemplo de compostos em Libras: o sinal de **SANGUE** (vermelho + escorrer). Nesse caso, o sinal sangue é altamente lexicalizado, portanto, deve ter uma entrada à parte.

Figura 16. Entrada do sinal SANGUE



SANGUE (1) (sinal usado em: **SP, SC, RS**) (Inglês: *blood*): s. m. Líquido espesso e vermelho composto de plasma, glóbulos vermelhos e brancos, além de plaquetas, e que circula pelo sistema vascular do corpo carregado de nutrientes e oxigênio, provenientes dos aparelhos digestivo e respiratório, para os tecidos por meio das artérias e, de volta dos tecidos até os pulmões, por meio das veias. Ex.: O pobre homem acidentado perdeu muito sangue enquanto esperava pelo resgate. Ex.: Doando sangue podemos salvar muitas vidas. Ex.: O sangue é produzido pela medula óssea, bombeado pelo coração, oxigenado pelos pulmões e filtrado pelos rins. Ex.: Qual é seu tipo sanguíneo? O meu tipo é O Rh negativo, e sou doador universal. (Fazer este sinal VERMELHO: Mão em 1, palma para trás, ponta do indicador tocando abaixo do lábio inferior. Movê-la, ligeiramente, para baixo, curvando o dedo indicador, duas vezes. Em seguida, braço esquerdo distendido, mão esquerda fechada, palma para a direita; mão direita em D, palma para baixo. Tocar a ponta do indicador direito no pulso esquerdo, baixar a mão direita, abrindo-a, e oscilar os dedos.) **Etimologia.**

Fonte: retirada do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 2514)

O desafio para o lexicógrafo das línguas de sinais é determinar quando dois sinais representam um composto lexicalizado e, portanto, ocupam um lema único, e quando eles representam uma colocação, ou seja, dois sinais que aparecem ao lado de um outro com frequência, mas não formam um composto lexicalizado.

3.3 Descrição de critério de lematização baseado em nível de significado

Além dos critérios com base na forma dos sinais, os autores (FENLON, CORMIER E SCHEMBRI, 2015) também apontam critérios baseados especificamente no significado do sinal, determinando quando duas variantes constituem um mesmo lema ou não, em duas categorias distintas: a) Homônimos e b) Polissêmicos.

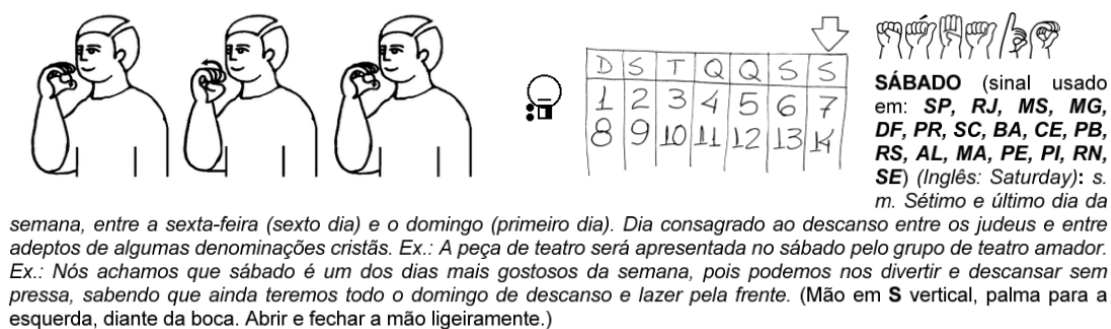
- a) **Homônimos:** São sinais com parâmetros formativos idênticos, porém, com significados distintos, isto é, pares de sinais que possuem a mesma construção fonológica, mas diferem semanticamente. Por exemplo, na Libras os sinais SÁBADO e LARANJA são fonologicamente idênticos, porém, com significados completamente distintos, sendo, assim, considerados homônimos, como mostrado nas Figura 17 e Figura 18.

Figura 17. Entrada dos sinais LARANJA (cor) e LARANJA (fruta)



Fonte: retiradas do *Dic-Brasil* (Capovilla, Raphael, Temoteo & Martins, 2017, p. 1843)

Figura 18. Entrada do sinal SÁBADO



Fonte: retirada do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 2488)

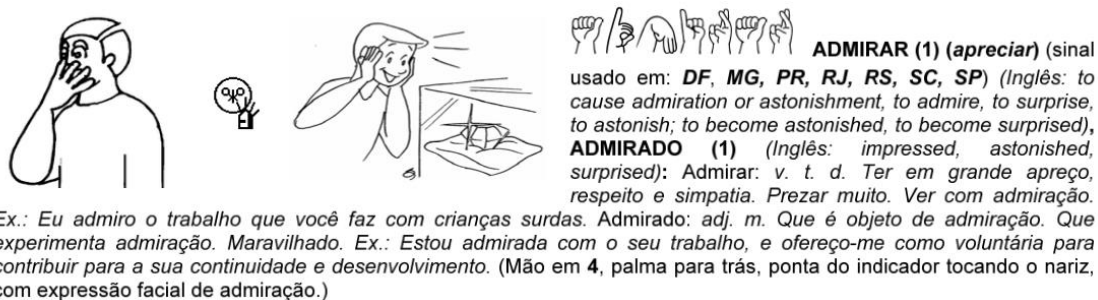
Como pode ser observado, as entradas LARANJA e SÁBADO no *Dic-Brasil* (2017) estão devidamente separadas já que possuem significados completamente distintos. Nesse caso, sugere-se que homônimos constituam lemas separados em dicionários ou bancos lexicais.

Outra decisão lexicográfica que merece atenção são as entradas distintas de LARANJA (fruta) e LARANJA (cor). Pelos critérios de lematização por significado, ambas poderiam ser registradas na mesma entrada, já que a forma do sinal é exatamente a mesma e o significado é similar.

b) Polissêmicos: Outro critério semântico está relacionado a sinais com significados amplos e próximos, classificados como polissêmicos. O

sinal apresentado na entrada ADMIRAR em Libras pode servir como exemplo (Figura 19), pois pode ter significados diversos relacionados como contemplar, apreciar, observar e assim por diante. Existem muitos casos na Libras que um único sinal tem uma gama de significados associados.

Figura 19. Entrada do sinal ADMIRAR



Fonte: retirada do *Dic-Brasil* (CAPOVILLA, RAPHAEL, TEMOTEO E MARTINS, 2017: 104)

Segundo Fenlon, Cormier e Schembri (2015) a abrangência de sinais polissêmicos pode se estender tanto para conceitos abstratos quanto para conceitos concretos, e essa questão, com frequência, é relativa à extensão metafórica de conceitos concretos para conceitos abstratos. Por exemplo, palavras-chave associadas ao sinal ESTRADA podem indicar tanto conceitos concretos como avenida, caminho, rodovia ou rua, quanto conceitos abstratos como método ou caminho.

Em alguns casos, o significado de alguns sinais tem uma relação metonímica uns com os outros, como por exemplo, o sinal DISTINTIVO, como delegado, oficial. Seguindo rigorosamente princípios de lematização apontados na literatura, se os parâmetros formativos são idênticos e os significados relacionados, então se referem ao mesmo lema. No entanto, exceções são apontadas por Fenlon, Cormier e Schembri (2015), alguns sinais com os mesmos parâmetros formativos e significados relacionados, mas altamente lexicalizados podem ser reconhecidos como entrada a parte.

Na Libras, por exemplo o sinal da cidade AMERICANA é exatamente o mesmo sinal para ESTADOS UNIDOS ou AMERICANO. Uma vez que o sinal da cidade de Americana é lexicalizado, foi reconhecido como um sinal registrado em entrada própria no *Dic-Brasil*.

Considerações finais

Este artigo descreve práticas para lematização nos níveis fonético-fonológico, morfológico e semântico específicos para Libras. Os princípios desenvolvidos podem ser considerados, como sugestão, na criação de *corpora*, dicionários, glossários e banco de dados lexicais da Libras. O desenvolvimento de produtos lexicográficos lematizados da Libras possibilita um avanço significativo nos estudos linguísticos e gramaticais, possibilitando generalizações sobre o léxico e descrições quantitativas e qualitativas do núcleo lexical dessa língua.

A construção de dicionários, *corpus* e banco de dados lexicais requer decisões complexas sobre inclusão e agrupamento de sinais, além de recursos financeiros, humanos, científicos e tecnológicos. O *Dic-Brasil* é um *corpus* profícuo para ilustrar possíveis critérios de lematização. É importante ressaltar que a decisão de lexicógrafos das línguas de sinais de registrar em entradas separadas variantes de um mesmo lexema, dá-se também pelo fato de não existirem ainda estudos linguísticos suficientes que possam dar embasamento para as transformações previsíveis dos sinais da Libras.

Enquanto *corpora* servem majoritariamente para o desenvolvimento de ferramentas tecnológicas e pesquisas linguísticas, dicionários possuem um propósito pragmático, especialmente direcionado para aprendizes. Dessa forma, os princípios de lematização empregados na construção de dicionários devem levar em conta o pragmatismo do ponto de vista do consulente. Dicionários lematizados devem conter em cada entrada no mínimo um sinal na forma de citação, ou seja, o lema, juntamente com as variantes fonológicas e morfológicas. O procedimento pode ser feito por meio de diferentes camadas no software *ELAN*.

Referências

BRENTANI, D. *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press, 1998.

- BRIEN, D. (ED.). *Dictionary of British Sign Language/English*. Boston: Faber & Faber, 1992.
- CAPOVILLA, F. C.; RAPHAEL, W. D; TEMOTEO, J. G.; MARTINS, A. C. *Dicionário da Língua de Sinais do Brasil: A Libras em Suas Mãos*. Volume 1: Sinais de A a D. São Paulo: Edusp, 2017a, p. 1037.
- _____. *Dicionário da Língua de Sinais do Brasil: A Libras em Suas Mãos*. Volume 2: Sinais de E a O. São Paulo: Edusp. 2017b. p. 1100.
- _____. *Dicionário da Língua de Sinais do Brasil: A Libras em Suas Mãos*. Volume 3: Sinais de P a Z. São Paulo: Edusp. 2017c. p. 862.
- CENTRE FOR TEGNSPROG. *Ordbog over Dansk Tegnsprog. (the Danish Sign Language Dictionary)*, 2008.
- CORMIER, K. et al. *From corpus to lexical database to online dictionary: Issues in annotation of the BSL Corpus and the development of BSL SignBank* In CRASBORN, O. et al. (Eds.), *Proceedings of the 5th Workshop on the representation and processing of sign languages: Interactions between corpus and lexicon* (pp. 7-12). Paris: ELRA, 2012.
- CRASBORN, O., & SLOETJES, H. *Enhanced ELAN functionality for sign language corpora*. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora* (pp. 39-43), 2008.
- FENLON, J.; CORMIER K. SCHEMBRI, A. *Building BSL SignBank: The lemma dilemma revisited*. *International Journal of Lexicography* 28:2, 169-206.
- FEIREIRA-BRITO, L. *Similarities and Differences in Two Sign Languages*. *Sign Language Studies*. 42: 45-46. Linstok Press, In: Silver Spring, USA, 1984.
- FEIREIRA-BRITO, L. *Epistemic, Alethic, and Deontic Modalities in a Brazilian Sign Language*. In: S.D. Fisher and P. Siple (eds.) *Theoretical Issues in Sign Language Research*. Vol. 1. University of Chicago Press, 1990.
- FEIREIRA-BRITO, L. *Por uma gramática de línguas de sinais*. Rio de Janeiro: Tempo brasileiro, 1995, 273 p.
- JOHNSTON, T. *The Sign Language of the Australian Deaf Community*. Doctoral dissertation, University of Sydney, 1989.
- JOHNSTON, T.; SCHEMBRI, A. *On defining lexeme in a signed language*. *Sign Language and Linguistics*, 2(2), p. 115-185, 1999.
- JOHNSTON, T., SCHEMBRI A. *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge: Cambridge University Press, 2007.
- JOHNSTON, T. *The Lexical Database of Auslan (Australian Sign Language)*. *Sign Language e Linguistics*, 4.1/2: 145-169. 2001.
- JOHNSTON, T. *Auslan Corpus annotation guidelines* [Relatório], 2019.
- HOCHGESANG, J., CRASBORN, O. E LILO-MARTIN, D. *Building the ASL Signbank: Lemmatization Principles for ASL*. In Bono, M. et Al. (Eds.), *Proceedings of the Language Resources and Evaluation Conference 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*; 69-74, 2018.
- KONRAD, R., et al. 'From Form to Function. A Database Approach to Handle Lexicon Building and Spotting Token Forms in Sign Languages' In Crasborn, O. et. Al. (Eds), *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between*

- Corpus and Lexicon*. Language Resources and Evaluation Conference (LREC), Istanbul, 87-94 May 2012.
- MARTINS, A.C. *Lexicografia, Metalexigrafia e Natureza da Iconicidade da Língua de Sinais Brasileira (Libras)*. Tese (Doutorado) - Instituto de Psicologia da Universidade de São Paulo, São Paulo, 2017.
- MCKEE, R. L.; MCKEE, D. *Making an Online Dictionary of New Zealand Sign Language*. *Lexikos*, 23.1: 500-531, 2013.
- MCKEE, R.L; VALE, M. *Sign Language Lexicography*. In P. Hanks e G. de Schryver (Eds.), *International Handbook of Modern Lexis and Lexicography* (pp. 22). Germany: Springer-Verlag GmbH, 2017.
- NIJMEGEN: Max Planck Institute for Psycholinguistics. *The Language Archive ELAN (Version 6.2)* [Computer software], 2020.
- PAIVA, F. A. S. et al. *Towards machine translation from Brazilian portuguese-to-libras: A corpus-based, morphosyntactic analysis*. In: XIV Encontro de Linguística de Corpus (ELC 2017). São Leopoldo, RS, Brasil, 2017.
- QUADROS, R. M; KARNOPP, L. *Língua de Sinais Brasileira: estudos linguísticos*. Porto Alegre: Artmed, 2004.
- RODERO-TAKAHIRA, A. G. *Compostos na Língua de Sinais Brasileira*. Tese de Doutorado em Semiótica e Linguística Geral. Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015.
- SCHRYVER, G. *Lexicographers' dreams in the electronic dictionary age*. In *The International Journal of Lexicography* 15(2):143-199, 2003.
- STOKOE, W. *Sign language structure: An outline of the visual communication systems of the American Deaf*. *Studies in Linguistics, Occasional Papers* 8. Buffalo: University of Buffalo Press, 1960.
- WELKER, H. A. *Dicionários: Uma pequena introdução à Lexicografia*. Brasília, DF: Thesaurus, 2004.