

A DESAMBIGUAÇÃO DE PALAVRAS HOMÔNIMAS EM SENTENÇAS POR APLICATIVOS DE TRADUÇÃO AUTOMÁTICA PORTUGUÊS BRASILEIRO-LIBRAS

THE DISAMBIGUATION OF HOMONYMOUS WORDS IN SENTENCES BY BRAZILIAN PORTUGUESE-LIBRAS AUTOMATIC TRANSLATION APPLICATIONS

Ygor Corrêa*
Rafael Peduzzi Gomes**
Carina Rebello Cruz***

RESUMO

Este estudo analisa a Tradução Automática (TA) de palavras homônimas isoladas e inseridas em sentenças, realizada pelos aplicativos Hand Talk (HT) e ProDeaf Móvel (PDM), ambos tradutores automáticos do Português Brasileiro (PB) para Língua Brasileira de Sinais (Libras), uma vez que estudos anteriores têm apontado para a ausência de estratégias de desambiguação, uso de datilologia e erros de tradução. Esta é uma pesquisa qualitativa exploratória fundamentada em estudos sobre o léxico da Libras, Tradução Automática, Processamento de Línguas Naturais e Desambiguação, que tem por finalidade aprofundar achados iniciais, sob o mesmo viés teórico, com vistas a propor melhorias para a TA de PB-Libras. Assim, esta pesquisa analisou a TA de 38 palavras homônimas isoladas e contextualizadas em 38 pares de sentenças (dois significados em cada par de sentença). Os resultados revelaram que na TA de palavras isoladas os aplicativos geram a tradução de apenas um dos significados das palavras homônimas (HT: 89% e PDM: 63%). Nas sentenças, o percentual de acerto do significado das palavras, em ambos os aplicativos, é menor (HT: 82% e PDM: 60%) do que na TA por palavra isolada e ainda mais reduzido (HT: 13% e PDM: 11%) na desambiguação de palavras homônimas. Os achados desta pesquisa indicam a necessidade de aperfeiçoamento do sistema de Processamento de Línguas Naturais dos aplicativos quanto à desambiguação de palavras homônimas. Infere-se que uma TA mais adequada em nível de palavra isolada e de sentença pode oferecer aos usuários traduções mais análogas à Libras em níveis semântico e sintático, enquanto língua natural e oficial da comunidade surda no Brasil.

Palavras-chave: aplicativos de tradução automática; libras; palavras homônimas.

* Centro Universitário Ritter dos Reis (UniRitter), Porto Alegre (RS), Brasil. correaygorprof@gmail.com

** Centro Universitário Ritter dos Reis (UniRitter), Porto Alegre (RS), Brasil. rafaelpeduzzi@gmail.com

*** Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, (RS), Brasil. carina.cruz@ufrgs.br

ABSTRACT

This study analyzes the Automatic Translation (AT) of homonymous words isolated and inserted in sentences, produced by Hand Talk and ProDeaf Mobile applications, both automatic translators from Brazilian Portuguese to Brazilian Sign Language (Libras), once previous studies have pointed to the absence of disambiguation strategies, use of fingerspelling and translation errors. This is an exploratory qualitative research based on studies of the lexicon of Libras, Automatic Translation, Natural Language Processing and Disambiguation, whose purpose is to deepen initial findings, under the same theoretical bias, in order to propose improvements on the AT of Portuguese Language-Libras. Thus, this study analyzed the AT of 38 homonymous words isolated and contextualized in 38 sentence pairs (two distinct meanings in each sentence pair). The results revealed that in AT of isolated words the applications generate the translation of only one of the meanings of the homonymous words (HT: 89% and PDM: 63%). In sentences, the percentage of correct word meanings, in both applications, is lower (HT: 82% and PDM: 60%) than in AT per isolated word and even lower (HT: 13% and PDM: 11 %) in the disambiguation of homonymous words. The findings of this research indicate the need to improve the Natural Language Processing system of the applications for the disambiguation of homonymous words. It is inferred that a more adequate AT at the level of isolated word and sentence can offer users more analogous translations to Libras, as the natural and official language of the deaf community in Brazil.

Keywords: automatic translation applications; libras; homonymous words.

INTRODUÇÃO

Este estudo analisa a Tradução Automática (TA) do Português Brasileiro (PB) para Língua Brasileira de Sinais (Libras) de palavras homônimas nos aplicativos Hand Talk (HT) e ProDeaf Móvel (PDM)¹. A condução do presente estudo está associada, inicialmente, ao surgimento e à disponibilização gratuita dos aplicativos a partir de 2013, os quais visam ao rompimento de barreiras comunicacionais entre ouvintes e surdos, assim como ao fato de que estes têm se voltado cada vez mais ao cenário de educação bilíngue. No entanto, estes aplicativos, enquanto tecnologias digitais contemporâneas, que fazem uso de agentes animados em 3D, para fins de apresentação da sinalização realizada do PB para Libras, têm sido motivo de referência a fragilidades linguísticas. Estudos anteriores têm apontado para traduções com a ausência de estratégias de desambiguação (CORRÊA; GOMES; RIBEIRO, 2017), uso de datilologia (alfabeto manual)² e erros de tradução (CORRÊA et al., 2014b; VIEIRA et al., 2014a, 2014b). De acordo com Specia (2007, p. 2),

1. Informações sobre os aplicativos e downloads - HT em: <https://www.handtalk.me/app> e PDM em <http://prodeaf.net/pt-br/prodeaf-movel>. Acesso em 13 fev. de 2018.
2. Na datilologia há a representação de letras de palavras faladas ou escritas em um código alfabético sendo usado em situações específicas. Nenhuma comunidade linguística utiliza exclusivamente tal código para comunicar-se (QUADROS; KARNOPP, 2004).

mesmo após décadas de pesquisa e com o registro de grandes avanços no campo da TA e de Processamento de Línguas Naturais (PLN)³, “alguns problemas ainda representam grandes barreiras para o desenvolvimento de aplicações visando traduções de qualidade”. Inclusive, uma das principais dificuldades no campo da TA é ambiguidade semântica das línguas, ou seja, gerar uma tradução quando há diversas interpretações para uma mesma proposição.

Neste horizonte, depreende-se, a partir dos autores mencionados, a efetiva demanda por aprofundamentos investigativos, que permitam uma melhor compreensão da TA gerada por aplicativos, do ponto de vista linguístico. Isso porque, conforme postulado por Corrêa, Gomes e Ribeiro (2017, p. 2), se está inferindo que “traduções incorretas podem gerar problemas de comunicação entre surdos e ouvintes”, assim como problemas na aprendizagem de Libras. Diante do exposto, esta pesquisa adota em sua fundamentação teórica os estudos sobre a formação do léxico e dos sinais da Libras (QUADROS; KARNOPP, 2004; FERREIRA-BRITO, 1995), Processamento de Línguas Naturais (WINOGRAD, 1972; BIDERMAN, 1978; DIAS-DA-SILVA, 1996; DI FELIPPO; DIAS-DA-SILVA, 2009), Tradução Automática (DORR; JORDAN; BENOIT, 2000; MELO, 2013), Estudos de Tradução (HOLMES, 1988; JAKOBSON, 1992) incluindo os de textos escritos para a Libras oral (SEGALA; QUADROS, 2015) e Desambiguação (HUTCHINS; SOMERS, 1992; BIDARRA, 2004; CABEZUDO, 2015; SILVA, 2016; SPECIA, 2007), de modo a propor melhorias ao campo da TA de PB-Libras. Esta é uma pesquisa qualitativa (SAMPIERI; COLLADO; LUCIO, 2013) exploratória (SANTOS, 2010), que se insere em uma temática ainda pouco investigada, pois os tradutores automáticos PB-Libras, disponibilizados em dispositivos móveis do tipo *smartphone* e *tablet*, são recentes no Brasil.

A partir da base teórica mencionada, o estudo retoma os achados iniciais de um estudo realizado por Corrêa, Gomes e Ribeiro (2017) sobre a análise de um conjunto de 38 palavras homônimas, as quais foram analisadas isoladamente com uso dos referidos aplicativos e se propõe a contextualização destas em 76 sentenças, constituídas por pares homônimos, possibilitando verificar como os significados de cada palavra homônima são traduzidos para Libras e se há desambiguação das palavras nas sentenças.

3. Dias-Da-Silva (1996) afirma que, em geral, os estudos se referem ao Processamento de Linguagem Natural, pois isso se deve à tradução equivocada do “substantivo *“language”* do inglês, que é notadamente ambíguo”, tendo este sido traduzido como *“linguagem”*. Para fins de esclarecimento, o autor enfatiza que *“Processamento de Línguas Naturais”* refere-se *“a línguas e não à capacidade humana da linguagem (p. 33)”*, sendo estes campos de pesquisa distintos.

A análise dos dados coletados nos dois aplicativos está dividida em três etapas, sendo verificado: (1) o significado de cada palavra homônima traduzida; (2) o significado de cada palavra traduzida em cada uma das 76 sentenças e no par de sentenças com palavras homônimas; (3) se o significado traduzido na consulta por palavra (correto ou incorreto) seria mantido na tradução em sentenças. Acredita-se que este estudo pode contribuir para estreitar o diálogo entre as áreas da Linguística Aplicada e das Ciências da Computação, relativo ao uso de aplicativos de tradução em contextos educacionais bilíngues para alunos surdos e ouvintes. O artigo está dividido em: 1. Língua Brasileira de Sinais: formação do léxico e dos sinais; 2. Processamento de Línguas Naturais, Tradução Automática e Desambiguação; 3. Investigações sobre aplicativos de TA: Hand Talk e ProDeaf Móvel; 4. Metodologia; 5. Análise e discussão de dados; 6. Considerações Finais; e Referências. Na próxima seção, são apresentadas postulações sobre a Libras e a formação de sinais.

1. LÍNGUA BRASILEIRA DE SINAIS: FORMAÇÃO DO LÉXICO E DOS SINAIS

Comunidades surdas em todo o mundo utilizam línguas de sinais que são ininteligíveis entre si e independentes das línguas orais que estão a sua volta (MACSWEENEY et al., 2008). Desta forma, a Língua Brasileira de Sinais (Libras) é utilizada no Brasil, a Língua Gestual Portuguesa (LGP) é utilizada em Portugal, a Língua de Sinais Americana (ASL)⁴ nos Estados Unidos, a Língua de Sinais Britânica (BSL)⁵ no Reino Unido, a Língua de Sinais Sueca (SSL)⁶ na Suécia, e assim por diante. Do mesmo modo que as línguas orais, as línguas de sinais são línguas naturais que passam de geração a geração e são adquiridas pelas crianças sem qualquer ensino formal, ou seja, sua aquisição ocorre por meio de exposição natural e interação (SANDLER, 2005; QUADROS, 2017). A língua de sinais é a primeira língua ou a língua materna de muitas pessoas surdas, mas também é a primeira língua ou língua materna de pessoas ouvintes ou surdas-cegas.

As línguas de sinais podem ser analisadas nos mesmos níveis estruturais linguísticos das línguas orais, a saber: fonológico, morfológico, sintático, semântico e pragmático, e são tão completas e complexas quanto as línguas orais. Embora as línguas orais e as línguas de sinais possam transmitir a mesma informação, essas se diferenciam em relação à modalidade: visual-espacial e auditivo-oral, respectivamente. Nas línguas orais a recepção é pela da audição e a produção é por

4. ASL: *American Sign Language*.

5. BSL: *British Sign Language*.

6. SSL: *Swedish Sign Language*.

meio do trato vocal e órgãos fonoarticulatórios; nas línguas de sinais a recepção ocorre pela visão e a produção por meio dos movimentos das mãos, face e corpo e do uso do espaço (QUADROS; KARNOPP, 2004). Considerando que o presente artigo analisa a TA (PB-Libras) de palavras homônimas do PB isoladas e em sentenças, a formação do léxico e dos sinais na Libras é abordada de forma breve nas próximas duas subseções.

1.1 Formação do léxico na Libras

Diferentes tipos de sinais são utilizados durante a interação na língua de sinais. Na Libras, Quadros e Karnopp (2004) seguem o modelo de Brentari e Padden (2001), dividindo o léxico em: léxico nativo (classificadores), núcleo e soletração manual ou datilologia (alfabeto). A Figura 1, a seguir, mostra a adaptação do modelo Brentari e Padden (2001) por Quadros e Karnopp, para Libras. O léxico nativo está marcado em negrito.

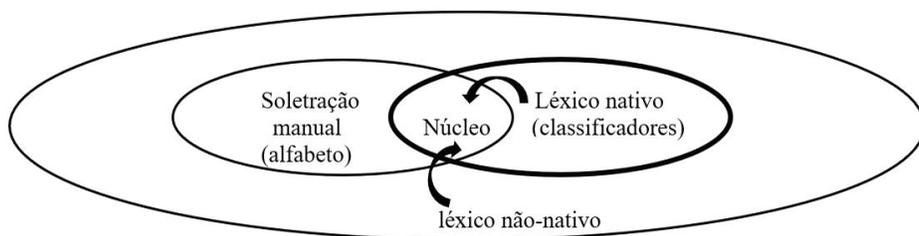


Figura 1. Léxico na Libras. Fonte: Quadros e Karnopp (2004, p. 88).

Segundo Quadros e Karnopp (2004), no léxico nativo da Libras (representado no círculo em negrito na Figura 1) há sinais que obedecem às restrições de boa-formação de sinais e os sinais que usam classificadores⁷, que podem violar restrições formacionais do núcleo lexical. Os sinais que formam o léxico nativo (e que não usam classificadores), conforme Kormier e Brentari (2018), são compostos por unidades sublexicais sem significado e apresentam uma forma e um significado fortemente estabelecidos, ou seja, esses sinais frequentemente são encontrados em um dicionário de língua de sinais.

Em relação ao léxico não-nativo, há palavras que são soletradas manualmente como o sinal AZUL que deriva da soletração 'AL', pois algumas palavras da língua

7. Os classificadores são formados por unidades com significado e são, geralmente, usados para especificar o movimento e a posição dos objetos e das pessoas ou para descrever o tamanho e a forma dos objetos (QUADROS; KARNOPP, 2004).

oral podem ser emprestadas à Libras por meio da soletração. Além disso, no léxico não-nativo a soletração manual de palavras completas em português é utilizada por sinalizadores em uma variedade de contextos, como, por exemplo, para introduzir uma palavra técnica (QUADROS; KARNOPP, 2004). A seguir, nas Figuras 2 e 3, apresenta-se a produção do sinal AMANHÃ (léxico nativo), em Escrita de Sinais (SW)⁸, que registra a produção exata do sinal citado, e em foto. Na Figura 4, apresenta-se a soletração manual da palavra ‘flúor’.



Figura 2. Sinal AMANHÃ⁹ na Libras, escrito em SW.



Figura 3. SINAL AMANHÃ¹⁰ em Libras¹¹.



Figura 4. Soletração manual da palavra ‘flúor’.

8. Informações sobre o Sistema *SignWriting* em: <http://www.signwriting.org/>. Acesso em 10 fev. de 2018.

9. Imagem do sinal AMANHÃ disponível no Dicionário Brasil do *SignPuddle online*: <http://www.signbank.org/signpuddle2.0/index.php?ui=12&sgn=46>. Acesso em 10 fev. de 2018.

10. O sinal AMANHÃ é um sinal nativo na Libras. A configuração de mão utilizada não representa uma letra do português no alfabeto manual da Libras.

11. Os autores agradecem ao colega Nelson Goettert, que, gentilmente, disponibilizou o registro dos sinais AMANHÃ e FONOLOGIA e autorizou o uso de sua imagem neste artigo.

1.2 Formação dos sinais na Libras

Os estudos sobre a fonologia de línguas de sinais são recentes se comparados aos estudos sobre línguas orais. Os primeiros estudos linguísticos foram conduzidos, por Stokoe (1960) na ASL. Stokoe (1960) observou que os sinais eram formados por um pequeno e finito conjunto de componentes sem significado e propôs a decomposição dos sinais em três parâmetros, a saber: configuração de mão, locação e movimento (SANDLER, 2008). Posteriormente, Battison (1978) introduziu mais um parâmetro: a orientação da palma da mão. Segundo Sandler (2012), há autores que incluem a orientação da palma da mão como uma quarta categoria ou parâmetro. Entretanto, outros modelos representam a orientação da palma da mão como uma categoria subordinada à configuração de mão. Além dos parâmetros que formam os sinais há a produção de expressões não-manuais que podem acompanhar a produção de alguns sinais.

Em relação aos parâmetros que formam os sinais, segundo Ferreira-Brito (1995), o parâmetro configuração de mão (CM) se refere às diversas formas que a(s) mão(s) toma(m) na realização do sinal. Há muitas possibilidades anatômicas, embora cada língua de sinais possua o seu inventário de configurações de mão. Na Libras, Ferreira-Brito e Langevin (FERREIRA-BRITO, 1995) identificaram 46 configurações de mão. Há, entretanto, outras propostas com mais configurações de mão: 73 configurações no Dicionário da Língua Brasileira de Sinais, desenvolvido por Lira e Souza (2008), disponível *online*, e 75 configurações de mão no estudo de Faria-Nascimento (2009).

O parâmetro locação (L) ou ponto de articulação, por sua vez, segundo Ferreira-Brito (1995), é o espaço em frente ao corpo ou uma região do próprio corpo, onde os sinais são articulados. O movimento (M) é um parâmetro complexo que pode envolver uma vasta rede de formas e direções, desde os movimentos internos da mão, os movimentos do pulso, os movimentos direcionais no espaço até conjuntos de movimentos no mesmo sinal (KLIMA; BELLUGI, 1979). Finalmente, a orientação da mão (Or.) é a direção da palma da mão durante a produção do sinal. A palma da mão pode estar voltada para cima, para baixo, para o corpo, para frente, para esquerda ou para a direita. Além das unidades sublexicais referidas, nas línguas de sinais há muitos itens lexicais que são produzidos com expressões não-manuais (movimentos da face, dos olhos, da cabeça ou do tronco). As expressões não-manuais além de possuírem a função de diferenciação entre itens lexicais, são empregadas para expressar marcações sintáticas, como nas sentenças interrogativas, orações relativas, topicalização, concordância e foco (QUADROS; KARNOPP, 2004).

As Figuras 5 e 6, a seguir, mostram as unidades sublexicais que formam o sinal FONOLOGIA escrito no sistema *SignWriting* e em foto. No sinal FONOLOGIA, há uso das duas mãos. A mão que tem CM somente com indicador estendido na horizontal é a mão de apoio, sendo utilizada como a locação. A mão dominante, com os cinco dedos estendidos, produz movimentos alternados dos dedos para a direita, sobre a mão de apoio.



Figura 5. Sinal FONOLOGIA¹², na Libras, escrito no Sistema SW.



Figura 6. Sinal FONOLOGIA na Libras.

Nesta seção, abordou-se brevemente a formação do léxico da Libras e das unidades que formam os sinais da Libras, sendo demonstrado que análises em níveis estruturais linguísticos independem da modalidade de língua. Desta forma, pode-se esperar que os sinais e as palavras em línguas de sinais e orais, respectivamente, explorem as vantagens disponíveis para seus respectivos sistemas. Como a fonologia é o nível da gramática que está diretamente relacionada com os sistemas fonéticos perceptivos e articulatórios, seja visual-espacial ou oral-auditivo, é esperado que surjam diferenças entre as modalidades de línguas e em sua organização de elementos (FENLON; CORMIER; BRENTARI, 2018). Na próxima seção são apresentados estudos sobre o Processamento de Línguas Naturais, Tradução Automática e Desambiguação.

12. Sinal escrito extraído do Glossário Libras, da Universidade Federal de Santa Catarina (UFSC), disponível em: <http://glossario.libras.ufsc.br/exibir/sinal/exibir/sinal/id/62>. Acesso em 10 fev. de 2018.

2. PROCESSAMENTO DE LÍNGUAS NATURAIS, TRADUÇÃO AUTOMÁTICA E DESAMBIGUAÇÃO

Nesta seção são abordados conceitos relacionados aos estudos sobre Processamento de Línguas Naturais (PLN) e Tradução Automática (TA) em diálogo com a Desambiguação, especificando as noções adotadas em relação às palavras homônimas.

2.1 Processamento de Línguas Naturais (PLN)

Inicialmente, o conceito de Processamento de Línguas Naturais (PLN) é apresentado enquanto “engenharia do conhecimento da linguagem” (DIAS-DA-SILVA, 1996; DI FELIPPO; DIAS-DA-SILVA, 2009) que, contemporaneamente, visa ao estabelecimento de um diálogo mais consistente entre pesquisadores das áreas da Linguística Aplicada e das Ciências da Computação. Isso porque, seja do ponto de vista teórico ou metodológico, infere-se que a já sugerida cooperação entre estas áreas do conhecimento (BIDERMAN, 1978), outrora mais dissociadas pela questão tecnológica (BERBER SARDINHA, 2005), pode representar uma contribuição mútua, no que tange ao desenvolvimento e à utilização de sistemas de PLN.

No contexto da inserção do computador na vida cotidiana, desde 1940 até os tempos atuais, a comunicação entre um sistema informatizado e seus usuários humanos se apresenta como um desafio. Dentre algumas soluções encontradas, como as linguagens de programação e as interfaces gráficas, o PLN foi impulsionado pela tentativa de compreender, manipular ou gerar texto/fala em língua natural, socialmente compreendida e usada por humanos (CHOWDHURY, 2003), com os primeiros sistemas de Tradução Automática (DI FELIPPO; DIAS-DA-SILVA, 2009).

Um sistema de PLN, segundo Winograd (1972), pode ser considerado como um “sistema especialista”, que usa uma fração específica do conhecimento humano e regras definidas por especialistas para tomar decisões e resolver problemas em um determinado campo. Há diferentes tipos de sistemas de PLN, sendo que a tradução automática (TA) faz parte deste conjunto. No Quadro 1 são apresentados alguns tipos de sistemas (sist.) de PLN, conforme a compilação de Di Felippo e Dias-da-Silva (2009).

Quadro 1 Sistemas de Processamento de Línguas Naturais desenvolvidos ao longo de 50 anos de pesquisas.

Dicionários, <i>thesaurus</i> e enciclopédias eletrônicas	Sist. de recuperação de informação	Sist. de extração de informação
Potencializam relações lexicais morfológicas, sintagmáticas, semânticas e paradigmáticas.	Buscam textos ou partes relevantes a uma dada consulta em uma coleção de textos ou documentos.	Buscam encontrar certa informação, em resposta a dada pergunta de entrada em um ou mais documentos.
Sist. de tradução automática	Sist. de sumarização automática	Sist. de correção ortográfica
Produzem, de um texto-fonte, escrito em uma língua natural x , um texto-alvo, versão do texto-fonte em uma língua y .	Geram “extratos” ou “resumos” de um ou mais textos, de acordo, por exemplo, com uma determinada taxa de compressão.	Processam um texto numa dada língua natural para identificar os erros de ortografia cometidos e sugerir alternativas prováveis.
Sist. de correção gramatical	Analísadores semânticos	Analísadores discursivos
Detectam desvios gramaticais em um texto, como os de concordância nominal ou verbal, pontuação e regência nominal.	Extraem automaticamente parcelas do conhecimento semântico subjacente a um texto.	Extraem automaticamente parcelas do conhecimento discursivo subjacente a um texto.
Sist. de categorização de textos	Sist. de diálogos	Sist. de auxílio à escrita
Classificam, mesmo que de modo parcial, textos de acordo com algum critério, como domínio, gênero, estilo, etc.	Englobam a interpretação e a participação em um diálogo, geralmente com um usuário, como em sistemas de compra de passagens.	Auxiliam à produção de texto com recursos para o usuário construir textos bem estruturados, de um gênero/domínio, etc.

Fonte: adaptado de Di Felippo e Dias-da-Silva (2009).

Mesmo com 50 anos de pesquisas sobre o PLN, considerado como um campo fértil (DI FELIPPO; DIAS-DA-SILVA, 2009), os sistemas de Tradução Automática (TA) feita por máquinas, segundo Chowdhury (2003), ainda se configuram como um complexo desafio não resolvido. Nesse sentido, os sistemas de PLN (CHOWDHURY, 2003) devem, para compreender uma língua natural, considerar a natureza léxico-sintática de um termo e seu significado, a gramática e a ordem da frase, o sentido da sentença como um todo e as relações contextuais para seleccionar, enfim, o termo adequado ao contexto ou compreender o sentido dado de um determinado termo.

Dessa forma, neste estudo, o PLN está tomado como uma área que se ocupa de aspectos linguístico-computacionais, voltados ao processamento de línguas naturais em interfaces digitais, quanto ao processamento do Português Brasileiro (PB) de modo escrito para a Língua Brasileira de Sinais (Libras), feito pelos aplicativos de tradução automática HT e PDM. A seguir, é abordado o conceito de Tradução Automática.

2.2 Tradução Automática (TA)

Gauche (2013) destaca a escassez de trabalhos voltados para a Tradução Automática (TA) de línguas orais para línguas de sinais, porém, essa é uma área que tem despertado o interesse global por parte de pesquisadores (BIDARRA, 2015), o que a torna num campo promissor de pesquisa. Desta forma, entende-se que a recente popularização das ferramentas de TA – para fins de comunicação, de leitura e de tradução por usuários da Internet – resulta de uma mudança de posicionamento, tendo em vista que na década de 1980, essas eram pouco usadas (MELO, 2013). Isso porque, anteriormente, considerava-se as limitações da tradução realizada apenas por máquina, sem intervenção humana. Em suas postulações, Melo (2013) aponta que, já a partir da década de 1990, a indústria buscou melhorar as máquinas, pautando-as em modelos de tradução baseados em fundamentos teóricos, nos quais, atualmente, os pesquisadores reconhecem tanto os progressos, quanto os desafios da TA, em um escopo chamado de tecnologias de apoio à tradução.

Convém ressaltar que, conquanto o presente estudo não objetiva identificar o tipo de estratégia gerada na TA dos aplicativos em análise, Hand Talk e ProDeaf Móvel, optou-se por denominar a tradução destes apenas como TA. No entanto, a literatura da área apresenta a TA dividindo-a em três estratégias, a saber: TA baseada em regras (*Rule-based Machine Translation* ou RBMT – GALLEY et al., 2004; GÜVENIR; CICEKLI, 1998); TA estatística (*Statistical Machine Translation* ou SMT – EL MAAZOUZI; EL MOHAJIR; AL ACHHAB, 2017; GULCEHRE et al., 2015) e, mais recentemente, TA neural (*Neural Machine Translation* ou NMT – KALCHBRENNER; BLUNSOM, 2013; CHO et al., 2014; SUTSKEVER et al., 2014; LUONG et al., 2015a, 2015b; JEAN et al., 2015; SENNRICH et al., 2016; LUONG; MANNING, 2016; JOHNSON et al. 2016).

Quanto aos sistemas de TA, Dorr, Jordan e Benoit (2000) distinguem três paradigmas predominantes: (1) a tradução baseada em conhecimento linguístico - dicionários e gramáticas (*Language-Based Machine Translation* – LBMT); (2) a tradução baseada em conhecimento - dicionários, gramáticas, enciclopédias e bases de conhecimento (*Knowledge-Based Machine Translation* – KBMT); e (3) a tradução baseada

em exemplos - dicionários, gramáticas e corpora (*Example-Based Machine Translation – EBMT*). Enquanto o primeiro paradigma se apresenta como de baixo custo e mais genérico, sem ater-se a outros conhecimentos além do que se apresenta no conteúdo a ser traduzido, os outros dois fornecem subsídios externos às questões linguísticas da frase para enriquecer o processo de tradução, combatendo ambiguidades e dificuldades em geral das línguas naturais.

Dessa forma, (2) e (3) se propõem a representar e a emular o conhecimento que seria usado por um leitor humano no processo de tradução, por meio de recursos linguístico-computacionais que devem ser representados de forma compreensível a uma máquina, os quais segundo Di Felippo e Dias-da-Silva (2009), podem ser descritos como: (a) *corpora* (textuais), coleções de textos úteis para o conhecimento linguístico (lexical, sintático, semântico), cujo levantamento pode ser realizado por linguistas, com o suporte de *softwares* para manipulação de *corpus*, ou por métodos estatísticos; (b) *léxicos*, estoques de unidades lexicais descritas junto a seus traços morfológicos, sintáticos, semânticos e/ou pragmático-discursivos, organizados segundo critérios; (c) *ontologias* (bases de conhecimento), inventários de conceitos, propriedades e relações conceituais que representam uma interpretação da realidade, baseada no conhecimento de uma comunidade linguística; e (d) *gramáticas*, sistemas de regras representados por sistemas formais, que descrevem estruturas de sentenças em uma língua e possibilitam, junto do léxico, a identificação e a geração de sentenças nessa língua.

Este estudo está enquadrado nos Estudos de Tradução, conforme as postulações de Holmes (1988), como um estudo de natureza aplicada, relacionado à Crítica de Tradução, sendo o objeto de análise a Tradução Automática via aplicativos. Neste sentido, optou-se por classificar a tradução abordada, como uma tradução intermodal na tradução interlingual (SEGALA; QUADROS, 2015), ainda que o tradutor não seja humano, mas um agente animado virtual em 3D, com forma humana que se expressa na Libras, visualmente, do PB para Libras. Compreendeu-se, neste caso, que a TA é o resultado da tradução do português escrito (língua fonte) para a Libras (língua alvo), ou seja, de línguas que se apresentam em diferentes modalidades. A seguir, são abordados os conceitos de desambiguação e tipos de ambiguidade lexical, como polissemia e homonímia

2.3 Desambiguação lexical

Os paradigmas citados (DORR; JORDAN; BENOIT, 2000) buscam lidar com o problema da ambiguidade das palavras, sempre presente nas aplicações de TA e sistemas de PLN. Nesse tocante, tratam do que é definido na literatura

como ambiguidade lexical na tradução, nos casos em que não é possível transpor simetricamente a multiplicidade de significados de um termo em uma língua para outra (HUTCHINS; SOMERS, 1992).

Specia (2007) acrescenta que a ambiguidade lexical pode ser categorial, com relação à categoria gramatical da palavra, ou semântica. Frente a essa questão, a Desambiguação Lexical surge como uma etapa da TA (COLLING, BOSCARIOLI, 2014; CABEZUDO, 2015) e de outras aplicações, sendo a tarefa computacional de resolver o problema da ambiguidade das palavras (SILVA, 2016). Nesse sentido, Silva (2016) afirma que a área ainda não atendeu a todos os seus desafios teóricos e práticos, e atenta para a escassez de trabalhos de desambiguação para o Português Brasileiro. Quanto à execução da desambiguação, Mihalcea (2006) aponta para três heurísticas, a fim de tornar menos complexos seus métodos: (1) *sentido mais frequente* - usar o sentido que ocorre com maior frequência em um *corpus* anotado ou em um repositório de sentidos; (2) *um sentido por colocação* - usar palavras que, frequentemente, co-ocorrem com a palavra a desambiguar como uma pista do sentido correto; e (3) *um sentido por discurso* - usar o sentido de uma palavra em um componente discursivo (sentença, parágrafo ou texto) como o mesmo em todas as ocorrências.

No contexto da ambiguidade lexical, segundo Specia (2007), os autores referenciam a homonímia e a polissemia como as mais relevantes, sendo que a distinção entre esses não é consensual. Vê-se que nos trabalhos de PLN, geralmente, são englobados os sistemas de Desambiguação Lexical, por receberem o mesmo tratamento, em essência: selecionar o sentido de uma palavra conforme seu contexto (SPECIA, 2007). Entretanto, esses conceitos se diferem, segundo Bidarra (2004), pela relação entre os significados possíveis: no caso da polissemia, os significados de uma palavra apresentam um relacionamento semântico, que os torna muito próximos uns dos outros; já no caso da homonímia, os significados possíveis de uma palavra são tão díspares que se tornam palavras completamente distintas, mas “acidentalmente” escritas da mesma forma.

Nesta perspectiva, compreendendo a distinção por meio de exemplos, conforme a classificação de Amaral et al. (2010), a palavra polissêmica apresenta significados diferentes, mas relacionados, e pode ser exemplificada pela palavra *rede*, que pode ser elétrica ou de computadores, mas se mantém como representação de um conjunto de objetos interconectados.

A palavra homônima, cujos significados são diferentes e sem relação entre si, ainda pode ser dividida em homonímia perfeita, homofonia e homografia. As homônimas perfeitas apresentam a mesma grafia e som, como no caso de *manga* da camisa ou a fruta, também *manga*. Já as homófonas possuem mesmo som, mas

grafias diferentes, por exemplo, *concerto* musical e *conserto* de um carro. Por último, as homógrafas apresentam a mesma grafia, mas com som diferente, como no caso da palavra *olho* (substantivo e verbo) nas seguintes sentenças: “Caiu um cisco no meu olho” e “Quando olho para você, lembro do meu pai”. A partir dessas categorias, neste estudo, é usada sempre a palavra *homônimo* para referir-se às palavras abordadas, pois sendo perfeitos, homófonos ou homógrafos, preservam sua classificação de homonímia. Concluída a fundamentação acerca de PLN, TA e Desambiguação Lexical, são apresentados, a seguir, os aplicativos HT e PDM, bem como estudos relacionados e o estudo que deu origem a esta pesquisa.

3. INVESTIGAÇÕES SOBRE APLICATIVOS DE TA: HAND TALK E PRODEAF MÓVEL

Nesta seção, inicialmente, são apresentados, brevemente, na subseção 3.1 os aplicativos Hand Talk e ProDeaf Móvel quanto à origem e às funcionalidades destes, em seguida, na subseção 3.2 estudos desenvolvidos com os referidos aplicativos, e, por fim, na subseção 3.3 um estudo sobre homonímia na TA dos aplicativos Hand Talk e ProDeaf Móvel, sendo este último, o estudo que deu origem à presente pesquisa, permitindo, a partir de seus dados iniciais, o aprofundamento aqui proposto.

3.1 Os aplicativos Hand Talk e ProDeaf Móvel

Hand Talk (HT) e ProDeaf Móvel (PDM) são aplicativos de tradução automática, ou seja, sistemas computacionais de processamento de línguas naturais, que traduzem palavras, frases ou pequenos parágrafos do Português Brasileiro (PB), escrito ou falado, para Língua Brasileira de Sinais (Libras) por meio de agentes animados virtual em 3D, comumente denominados como avatares. Estes aplicativos surgiram, inicialmente, como resultado de pesquisas acadêmicas realizadas no nordeste brasileiro (HT – Universidade Federal de Alagoas; PDM – Universidade Federal de Pernambuco) e, posteriormente, foram lançados no mercado brasileiro de dispositivos móveis por suas respectivas empresas.

Os aplicativos HT e PDM processam texto e voz, digitados ou falados, conteúdos esses capturados pela interface digital inicial (Figura 7) e traduzidos para Libras por meio da sinalização feita por de agente animado virtual em 3D. Assim, o usuário pode optar por escrever ou falar a palavra, sentença ou pequeno texto em PB para que a mensagem seja traduzida automaticamente para Libras. Os referidos aplicativos possuem sistemas operacionais compatíveis com boa parte dos aparelhos

smartphones e *tablets* em circulação no Brasil (ver CORRÊA et al., 2014b) e, além disso, podem operar parcialmente sem conexão à Internet, sendo instalados por meio de *download* em dispositivos móveis via lojas virtuais de aplicativos. Essas ferramentas digitais permitem selecionar a velocidade da sinalização, assim como diferentes ângulos de rotação do agente animado, para fins de uma melhor visualização da sinalização em Libras e disponibilizam dicionário, vocabulário e vídeos temáticos (apenas HT).

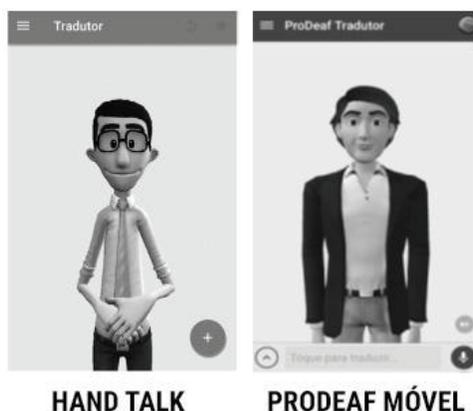


Figura 7. Interface digital dos aplicativos.

Fonte: capturas de tela.

Neste tocante, HT e PDM configuram-se como ferramentas digitais gratuitas, que visam à inclusão sociodigital, objetivando o rompimento de barreiras comunicacionais entre as comunidades ouvinte e surda no Brasil (VIEIRA et al., 2014b). Diante da presença massiva de dispositivos móveis, principalmente, do tipo *smartphone* (FGV-SP, 2016), em uma sociedade contemporânea hiperconectada à Internet (CASTELLS, 1999), a utilização dos referidos aplicativos foi impulsionada. Esta condição potencializou a inserção destes aplicativos não apenas em contextos interacionais sociais, mas, também, diante de suas funcionalidades, em contextos educacionais bilíngues (VIEIRA et al., 2014a). No entanto, conforme mencionado na introdução deste artigo, ao serem utilizados em contextos educacionais bilíngues os aplicativos trouxeram à tona questões pontuais, sendo estas relativas à qualidade da tradução automática feita por estes, no que se refere a fragilidades linguísticas, as quais têm sido reveladas por pesquisadores da área, quanto a traduções com a ausência de estratégias de desambiguação (CORRÊA; GOMES; RIBEIRO, 2017), uso de datilologia (alfabeto manual) e erros de tradução (CORRÊA et al., 2014a, 2014b; VIEIRA et al., 2014a, 2014b). A seguir são apresentados estudos

realizados sobre os aplicativos HT e PDM, que visavam a verificar a validade social e inclusiva destes, do ponto de vista tecnológico e linguístico, em escolas regulares, em Atendimento Educacional Especializado (AEE) e em escolas bilíngues, que contaram com a participação de sujeitos de pesquisa ouvintes, surdos, intérpretes/tradutores e professores fluentes em Libras, surdos e ouvintes.

3.2 Estudos em contexto educacional com os aplicativos Hand Talk e ProDeaf Móvel

Um dos estudos iniciais sobre as ferramentas em pauta foi o de Corrêa et al., (2014a), que investigou a utilização dos aplicativos HT e PDM, quanto à inserção destes na promoção de uma melhor comunicação entre surdos e ouvintes. Os autores enfatizam aspectos como: a possibilidade de ampliação do arcabouço linguístico de sujeitos surdos e ouvintes, dado o caráter bilíngue das ferramentas em questão; a socialização da pessoa surda em contextos orais; facilidade de uso dos aplicativos, do ponto de vista da usabilidade; autonomia da pessoa surda; e aspectos negativos, como a necessidade de possuir um dispositivo *smartphone*, que ainda tem custo elevado no Brasil; e, por fim, a falta de acesso à Internet. Em relação a isso, Corrêa et al., (2014a, p. 8) inferem que "(...) muitas escolas brasileiras ainda não possuem acesso à Internet ou não disponibilizam acesso a todos os seus espaços", entendendo que possuir um aparelho *smartphone*, por exemplo, para uso dos aplicativos HT e PDM, pode representar não restringir-se às limitações da escola, desde que este uso seja consentido.

Já Vieira et al. (2014a), ao investigarem as contribuições da Teoria da Aprendizagem Multimídia e da Usabilidade para aprendizagem de Libras e Língua Portuguesa, em relação aos referidos aplicativos móveis, apontaram aspectos relacionados: à validade social destes enquanto Tecnologias Assistivas, mediadoras da comunicação entre sujeitos surdos e ouvintes acerca da aprendizagem de Língua Portuguesa e/ou Libras; às contribuições de teorias voltadas à aprendizagem multimídia, apoiada em sistemas computacionais e à usabilidade, quanto ao desenvolvimento e ao aprimoramento de sistemas tradutores automáticos de Língua Portuguesa para Libras, no que tange à redução de problemas de interface e de interação, com vistas a contemplar as demandas dos usuários e a garantir a qualidade de uso.

Ao conduzir uma análise das expressões não manuais de negação e interrogação dos agentes animados em 3D dos aplicativos HT e PDM, Vieira et al. (2014c) afirmam a existência de fragilidades quanto às expressões corporais e faciais, as quais podem resultar em processos linguísticos deficitários. Segundo

Vieira et al. (2014c), o estudo possibilitou a identificação de fragilidades, que foram ao encontro das narrativas dos sujeitos surdos e ouvintes, o que levou os autores à proposição de diretrizes compostas por 11 variáveis de diferentes expressões faciais com 8 combinações de movimentos não manuais. Os autores consideram que as diretrizes podem contribuir para futuras implementações, quando do desenvolvimento de agentes animados virtuais em 3D sinalizadores de Libras. Diante do exposto, é possível considerar que as pesquisas acadêmicas voltadas a aplicativos de tradução automática de uma língua oral-escrita (Português Brasileiro) para a uma língua sinalizada (Libras) no Brasil têm evidenciado aspectos que podem fomentar a sofisticação destes sistemas computacionais. Convém enfatizar, que os três estudos mencionados reforçam a necessidade de aprimoramento tecnológico e linguísticos por parte dos aplicativos investigados. Na próxima subseção, apresenta-se, isoladamente, o estudo a partir do qual a presente pesquisa toma emprestado dados iniciais a respeito das dificuldades enfrentadas por ambos os aplicativos em relação ao processamento de palavras homônimas, quando da TA de Língua Portuguesa para Libras.

3.3 Um estudo sobre homonímia na TA dos aplicativos Hand Talk e ProDeaf Móvel

Conforme referido, a presente pesquisa é a continuidade do estudo intitulado ‘Aplicativos de Tradução Português-Libras na Educação Bilíngue: desafios frente à desambiguação’, desenvolvido por Corrêa, Gomes e Ribeiro (2017). Este estudo avaliou a qualidade da tradução automática (TA), do PB para a Libras, de 38 termos homônimos via aplicativos HT e PDM. A relevância da investigação, segundo os autores, estava apoiada no entendimento de que o uso dos aplicativos em contexto educacional bilíngue para surdos devia levar em consideração que “traduções incorretas podem gerar problemas de comunicação e de internalização de parâmetros linguísticos de maneira equivocada” (CORRÊA; GOMES; RIBEIRO, 2017, p. 1).

O referencial teórico pautou-se em estudos sobre a Libras e seus dispositivos legais, Educação Bilíngue de Surdos e na área do Processamento de Linguagem Natural (PLN), mais especificamente quanto à Desambiguação de termos polissêmicos homônimos, a fim de identificar se essas tecnologias digitais contemporâneas de TA apresentavam traduções incorretas. Além disso, no referencial teórico são apresentados os aplicativos HT e PDM, sob a condição de inovações disruptivas, posto que esses visam ao preenchimento de uma lacuna relativa à oferta de tecnologias digitais de semelhante função no cenário educacional bilíngue para surdos. Para conduzir o estudo, os pesquisadores selecionaram 38

termos homônimos, sendo 36 homônimos perfeitos e 2 homógrafos (colher e olho) no PB. O conjunto de dados foi extraído de quatro apostilas impressas de nível básico de Libras (I, II, III, IV)¹³ (GOIÁS, 2011a; 2011b; 2011c; 2011d), que apresentam a palavra no PB e a imagem (foto) do sinal equivalente em Libras. O Quadro 2 a seguir, mostra o conjunto de 38 palavras homônimas no PB, sendo que em negrito estão os termos extraídos das apostilas e entre parênteses o significado atribuído em Libras nas mesmas.

Quadro 2. Conjunto de 38 termos homônimos selecionados por Corrêa, Gomes e Ribeiro (2017).

Andar (<i>caminhar</i>)	Cunhado (<i>parente</i>)	Heroína (<i>droga</i>)	Quarto (<i>numeral</i>)
Baixo (<i>altura</i>)	Derrame (<i>doença</i>)	Língua (<i>parte do corpo</i>)	Revista (<i>publicação</i>)
Bala (<i>doce</i>)	Disciplina (<i>matéria</i>)	Meia (<i>metade</i>)	Rosa (<i>cor</i>)
Bonito (<i>beleza</i>)	Estado (<i>do Brasil</i>)	Negro (<i>raça</i>)	Saia (<i>vestuário</i>)
Brasília (<i>cidade</i>)	Fino (<i>delgado</i>)	Nós (<i>pronome</i>)	Só (<i>conjunção</i>)
Cinza (<i>cor</i>)	Fiscal (<i>profissão</i>)	Olho (<i>parte do corpo</i>)	Tarde (<i>parte do dia</i>)
Colher (<i>talher</i>)	Fortaleza (<i>estado</i>)	Ouvido (<i>parte do corpo</i>)	Tênis (<i>esporte</i>)
Como (<i>pronome</i>)	Graça (<i>divina</i>)	Palmas (<i>cidade</i>)	Vestido (<i>vestuário</i>)
Contar (<i>números</i>)	Gravata (<i>vestuário</i>)	Peru (<i>país</i>)	-----
Cuba (<i>país</i>)	Grosso (<i>espesso</i>)	Primo (<i>parente</i>)	-----

Fonte: CORRÊA; GOMES; RIBEIRO (2017)

De acordo com Corrêa, Gomes e Ribeiro (2017, p. 6), a avaliação da TA das 38 palavras isoladas contou com a *expertise* de três pesquisadores: um linguista, um designer de interfaces e um intérprete de Libras com certificação ProLibras¹⁴, número coerente de avaliadores para uma inspeção de caráter tecnológico e linguístico. A tradução automática dos termos homônimos foi classificada nas seguintes categorias: a) Convergência: traduções em acordo com o sinal referido nas apostilas; b) Divergência: traduções em desacordo com o sinal referido nas apostilas; c) Datilologia (uso do alfabeto manual da Libras); d) Errado: traduções que não representavam o sinal referido nas apostilas e nem mesmo um par homônimo.

Na análise de dados das TAs foi constatado que a tradução feita pelo aplicativo HT alcançou maior índice de convergência (71%) em relação aos termos extraídos das apostilas do que o aplicativo PDM (39%). Comparativamente, o aplicativo HT apresentou menor índice de divergência (29%), posto que PDM mostrou-se mais inclinado à incidência de termos em divergência (53%). Contudo, infere-se que

13. As apostilas foram elaboradas a partir de parceria entre o Ministério da Educação e Cultura (MEC), a Secretaria de Educação Especial (SEESP) e a Secretaria de Estado da Educação de Goiás (SEE) através da Superintendência de Ensino Especial-SUIEE.

14. O exame ProLibras é uma combinação de um exame de proficiência e uma certificação profissional proposto pelo Ministério da Educação, por tempo determinado, conforme previsto no Decreto 5626/2005, decreto que regulamenta a Lei nº 10.436/2002 (QUADROS et al., 2009).

tais percentuais, mesmo o de menor índice, demandam tarefas de desambiguação lexical para palavras homônimas. Os autores evidenciaram que o uso de datilografia foi superior no aplicativo PDM (13%) em comparação a HT (3%). As traduções erradas foram identificadas apenas no aplicativo PDM (5%)¹⁵.

Tendo em vista os resultados obtidos, os autores sugerem a implementação de estratégias de desambiguação para ambos os aplicativos. Reiteram que os termos em divergência ao significado das apostilas não são propriamente erros de tradução e sim, sob a perspectiva do Processamento de Linguagem Natural, pares homônimos das palavras analisadas, aspecto que revela certa desconsideração por parte dos desenvolvedores dos aplicativos quanto à homonímia existente em uma língua natural. O estudo mostrou que para a desambiguação de termos isolados talvez seja “relevante o uso de mais de uma animação (entrada) para cada termo, de modo a demonstrar mais de um significado para o mesmo termo” (CORRÊA; GOMES; RIBEIRO, 2017, p. 10). Os autores depreendem que a referida estratégia de classificação dos termos homônimos é uma prática realizada em dicionários e tradutores de Língua Portuguesa, sejam esses na modalidade impressa ou digital. Os pesquisadores ainda afirmam que “se está distante de um Processamento de Linguagem Natural e de estratégias de desambiguação que abarquem amplamente a polissemia presente na tradução automática do PB para Libras” (CORRÊA; GOMES; RIBEIRO, 2017, p. 10). Por fim, os autores reforçam a presença de limitações e desafios linguísticos, no que concerne à implementação de sistemas de desambiguação de homônimos perfeitos e homógrafos. Na próxima seção, apresenta-se a metodologia de pesquisa.

4. METODOLOGIA

Esta é uma pesquisa qualitativa (SAMPLERI; COLLADO; LUCIO, 2013) exploratória (SANTOS, 2010) que investiga a TA de palavras homônimas do PB para Libras, em sentenças, pelos aplicativos Hand Talk (HT) e ProDeaf Móvel (PDM), que se insere em uma temática contemporânea pouco explorada. Haja vista se tratar de tecnologias digitais de constituição bilíngue em caráter de mobilidade digital, infere-se que a abordagem exploratória pode indicar melhorias que apontem para o aperfeiçoamento da TA de tecnologias digitais de função semelhante. Não

15. Quando a datilografia foi produzida diferente do termo digitado, os autores consideraram que a tradução pertencia a duas categorias, a saber: datilografia e divergente. O mesmo ocorreu com a soma percentual, incidindo na análise de dados um percentual superior a 100% em relação aos 38 termos (CORRÊA; GOMES; RIBEIRO, 2017).

obstante, o estudo pode estreitar o diálogo entre as áreas da Linguística Aplicada e das Ciências da Computação, a saber, pela utilização deste tipo de tecnologia digital em contextos educacionais bilíngues para alunos surdos e ouvintes. Para esta investigação foram elaboradas 76 sentenças em PB a partir do conjunto de 38 palavras homônimas previamente utilizados por Corrêa, Gomes e Ribeiro (2017) em um estudo sobre a TA PB-Libras de palavras homônimas via aplicativos HT e PDM. As 76 sentenças elaboradas para o estudo são apresentadas no Quadro 3 com a palavra homônima marcada em negrito em cada sentença.

Quadro 3. Pares de sentenças com 38 palavras homônimas. Fonte: elaborado pelos autores.

Palavras	Sentenças	Palavras	Sentenças
Andar	Vou andar nesta rua até a avenida. O senhor deve ir ao terceiro andar .	Grosso	Este pedaço está muito grosso . Não seja grosso com sua namorada.
Baixo	Baixo e alto são antônimos. Há quanto tempo você toca baixo ?	Heroína	Consegui me livrar do vício em heroína . Ela foi uma verdadeira heroína ontem.
Bala	Esse menino adora chupar bala . No tiro acertei somente a última bala .	Língua	Quase queimei minha língua . Estou aprendendo uma nova língua .
Bonito	Nunca fui um garoto bonito . A cidade de Bonito é muito legal.	Meia	Depois de meia hora, você pode ligar. Percebi que minha meia está rasgada.
Brasília	Estive em Brasília semana passada. Seu carro é uma Brasília muito velha.	Negro	Um negro alto entrou no restaurante. Ela tinha um lindo cabelo negro .
Cinza	Gostei da sua camiseta cinza . Odeio o cheiro de cinza de cigarro.	Nós	Nós não somos mais os mesmos. Preciso desatar estes nós .
Colher	A colher de sopa está suja. Preciso colher os morangos.	Olho	Meu olho está com alergia. Eu olho para você e lembro do seu pai.
Como	Como você soube disso? Eu como frutas no café da manhã.	Ouvido	Estou com uma forte dor no ouvido . Tenho ouvido coisas estranhas.
Contar	Vamos contar até dez. Quero contar uma história.	Palmas	Eu me mudei para Palmas . Uma salva de palmas para eles!
Cuba	Os médicos de Cuba estão aqui. A cuba do seu banheiro é linda.	Peru	Viajei para o Peru no ano passado. No natal comemos peru recheado.
Cunhado	Eu moro com meu cunhado . Este selo foi cunhado em bronze.	Primo	Meu primo disse que você é amigo dele. Dezessete é um número primo .
Derrame	Este senhor sofreu um derrame . Cuidado e não derrame esse suco.	Quarto	Ele foi o quarto colocado na maratona Adoro dormir no meu quarto .

Disciplina	Vou cursar uma nova disciplina . As crianças precisam de disciplina .	Revista	Trabalho em uma revista mensal. Fizeram a revista nos torcedores.
Estado	Meu estado é o Rio Grande do Sul. Tenho estado muito distraído.	Rosa	Sua cor preferida é rosa . Ganhei uma rosa de Dia dos Namorados.
Fino	Quero uma fatia de bolo mais fina . Ele é um homem elegante e fino .	Saia	O que você achou dessa saia bordada? Saia daqui agora!
Fiscal	Trabalho como fiscal sanitário. O presidente propôs um ajuste fiscal .	Só	Só você respondeu ao questionário. Ando muito só e deprimido aqui.
Fortaleza	Trabalho na cidade de Fortaleza . O castelo era uma fortaleza .	Tarde	Às cinco horas da tarde , vou embora. Não tarde a voltar para casa.
Graça	O milagre foi uma graça divina. Achei sua piada sem graça .	Tênis	Ela foi uma grande jogadora de tênis . Você viu meu tênis novo?
Gravata	Você sabe dar nó em gravata ? O golpe chamado de gravata é usado para defesa pessoal.	Vestido	Esse vestido é o mais lindo. Eu havia vestido essa camisa.

É importante ressaltar que as 38 palavras homônimas no PB possuem equivalentes na Libras (sinais), pertencentes ao léxico nativo ou não-nativo e, portanto, a datilografia nestes casos geralmente não é utilizada para produção do significado dos mesmos. Além disso, o sinal PERU tem a mesma forma de produção e dois sentidos diferentes (país e animal), portanto, é um homônimo na Libras e no PB. Os autores optaram por mantê-lo na análise pois, conforme será apresentado na próxima seção, os aplicativos usaram mecanismos diferentes para TA de palavras homônimas isoladas e palavras homônimas em sentenças. A TA das 76 sentenças recebeu o mesmo tratamento teórico-metodológico adotado no estudo anterior (CORRÊA; GOMES; RIBEIRO, 2017), possibilitando evidenciar a relação causada pela contextualização de palavras homônimas em sentenças. Neste estudo, optou-se por não descrever a estrutura sintática das sentenças utilizada nas TAs, não discutir se as sentenças traduzidas estão bem formadas em Libras nem verificar, especificamente, a TA das outras palavras não pertencentes ao conjunto estabelecido para análise nas sentenças. No entanto, considera-se fundamental o desenvolvimento destas investigações futuramente. Em relação à coleta de dados, nos aplicativos digitou-se cada palavra e, posteriormente, cada sentença, e registrou-se o significado da TA gerada em português.

A análise dos dados coletados nos dois aplicativos está dividida em três etapas. Na TA (PB - Libras) foi verificado: (1) o significado de cada palavra homônima traduzida; (2) o significado de cada palavra traduzida em cada uma das 76 sentenças e no par de sentenças com palavras homônimas; (3) se o significado traduzido na

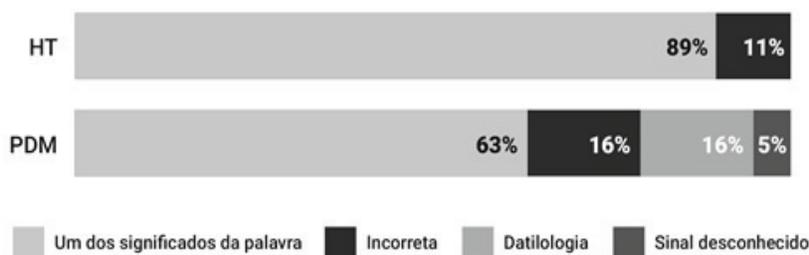
consulta por palavra (correto ou incorreto) seria mantido na tradução em sentenças. Na seção seguinte, são apresentados os dados analisados e os resultados obtidos.

5. ANÁLISE E DISCUSSÃO DE DADOS

Na primeira etapa, verificou-se a tradução automática (TA) de cada uma das 38 palavras homônimas do Português Brasileiro (PB) para Língua Brasileira de Sinais (Libras) em cada aplicativo. A digitação de cada uma das palavras do conjunto de 38 palavras homônimas do PB (CORRÊA; GOMES; RIBEIRO, 2017) na interface digital (tela inicial), de ambos os aplicativos, gerou a TA de apenas um de dois ou mais significados que cada palavra digitada expressa. A palavra ‘andar’, por exemplo, foi traduzida com diferentes significados em cada aplicativo. O Hand Talk (HT) gerou um sinal com o significado ‘caminhar’ e o ProDeaf Móvel (PDM) gerou um sinal com o significado de ‘divisória de um prédio’.

Em relação ao desempenho de cada aplicativo, na TA das 38 palavras homônimas do PB para Libras, constatou-se que o HT traduziu 34 palavras com apenas um dos dois ou mais significados que a palavra expressa e 4 palavras incorretas, a saber: cinza, disciplina, quarto e tênis. Os sinais utilizados na tradução destas palavras, foram: SUMIR, EXIGIR, DORMIR e MÊS, respectivamente. O aplicativo PDM, por sua vez, traduziu 24 palavras com apenas um dos dois ou mais significados que a palavra expressa, 6 palavras incorretamente, a saber: contar, graça, fino, heroína, fiscal, estado. Os sinais utilizados na tradução destas palavras, foram: SOMAR, GRÁTIS, MAGRO, FAMOSA, OPINIÃO QUALIDADE, TER VIVO (expressão usada para dizer que uma pessoa está no local). Além disso, 6 palavras foram traduzidas com o uso de datilologia, sendo que uma delas não faz parte do léxico (colher, derramar, revistar, cunhar, finar e cinzar) e 2 palavras não foram traduzidas com nenhum dos significados da palavra homônima e os sinais produzidos não foram identificados pelos pesquisadores (Palmas e primo). No Gráfico 1, é apresentado o desempenho de cada aplicativo em relação à TA (PB – Libras) das 38 palavras homônimas do PB, em porcentagem.

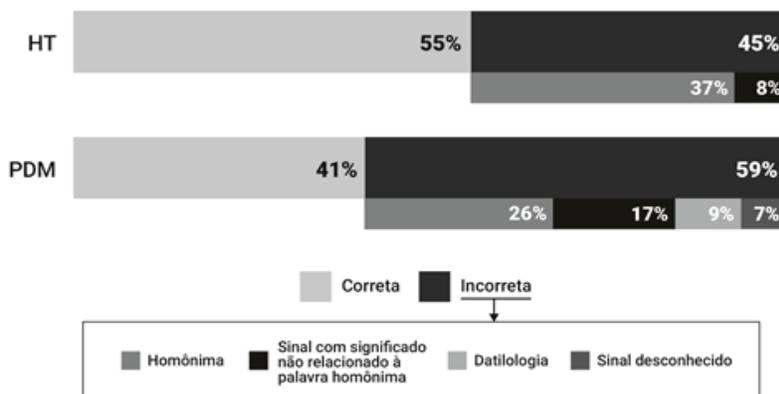
Gráfico 1. Desempenho dos aplicativos Hand Talk (HT) e ProDeaf Móvel (PDM) na Tradução Automática de 38 palavras homônimas do Português Brasileiro.



Conforme comentado, no conjunto de 38 palavras homônimas investigadas somente a palavra ‘peru’ é homônima em PB e em Libras. Em relação ao desempenho do aplicativo, o HT apresenta uma vantagem de acerto na TA de palavras homônimas, quando comparado ao PDM, que além de traduções incorretas, faz uso de datilologia ou produz um sinal com significado desconhecido.

Na segunda etapa da análise, verificou-se como cada aplicativo realizou a TA das 38 palavras homônimas em cada sentença (76) e, posteriormente, em cada par de sentenças (38), visando a identificar a porcentagem de TA com desambiguação. Considerando a TA correta de palavras homônimas em cada sentença, o aplicativo HT realizou a tradução de 42 sentenças e o PDM de 31 sentenças. Nas traduções incorretas verificou-se o uso de um sinal cujo significado corresponde a um dos significados da palavra homônima em PB, sinais que não correspondiam a nenhum sentido da palavra homônima, datilologia e sinais cuja produção não é conhecida pelos pesquisadores (sinal desconhecido). O gráfico a seguir, apresenta a porcentagem de TA correta e incorreta de palavras homônimas em cada sentença, assim como os tipos de erros identificados.

Gráfico 2. Desempenho dos aplicativos Hand Talk (HT) e ProDeaf Móvel (PDM) na Tradução Automática de 38 palavras homônimas do Português Brasileiro nas 76 sentenças.



Constatou-se que a porcentagem de acerto na TA (PB-Libras) de palavras homônimas é superior, quando a consulta é realizada por palavra. Os significados das palavras homônimas traduzidas isoladamente não foram mantidos em pelo menos uma das sentenças. Além disso, apesar de a porcentagem de acerto da TA de palavras homônimas em sentenças alcançarem porcentagem de acerto próxima ou um pouco acima da metade do total de sentenças, como veremos a seguir, a porcentagem de desambiguação ainda é reduzida.

O aplicativo HT gerou a TA do significado do par homônimo em 5 pares de sentenças, ou seja, houve desambiguação das palavras em cada sentença. Em relação às traduções sem desambiguação, em 31 pares de sentenças o significado das palavras homônimas foi traduzido corretamente em uma das sentenças e incorretamente na outra e, neste caso, o mesmo sinal foi utilizado nas duas sentenças (a palavra 'saia' nas duas sentenças foi traduzido para Libras com o significado 'ir embora', sendo gerado o sinal SAIR). Além disso, houve a tradução incorreta das palavras homônimas em 2 pares de sentenças. As palavras 'estado' e 'como' foram traduzidas para Libras com significado não relacionado à palavra homônima e com um movimento de ombros (não reconhecido pelos pesquisadores como equivalente ao sinal COMO).

O aplicativo PDM gerou a TA correta do significado do par homônimo em 4 pares de sentenças, ou seja, houve desambiguação das palavras em cada sentença. Em relação às traduções sem desambiguação, em 23 pares de sentenças o significado das palavras homônimas foi traduzido corretamente em uma das sentenças e incorretamente na outra. Neste caso, o mesmo sinal foi utilizado nas duas sentenças, datilografia ou um sinal que não se refere a nenhum dos sentidos da palavra homônima ou um sinal não conhecido pelos pesquisadores. Além disso, houve a tradução incorreta das palavras homônimas em 11 pares de sentenças. Na tradução para Libras, a TA gerou um sinal com um significado não relacionado à palavra homônima em PB, um sinal não conhecido pelos pesquisadores ou foi utilizada a datilografia de palavras do léxico do PB (colher) ou não (cinzar).

No Gráfico 3, é apresentado o desempenho de cada aplicativo em relação à TA (PB-Libras) dos 38 pares de sentenças, com palavras homônimas do PB, em porcentagem. Apesar da visível dificuldade dos aplicativos realizarem a TA (PB-Libras) de palavras homônimas em sentenças com desambiguação, ambos conseguem uma porcentagem de acerto muito semelhante.

Gráfico 3. Desempenho dos aplicativos Hand Talk (HT) e ProDeaf Móvel (PDM) na Tradução Automática de palavras homônimas do Português Brasileiro em 38 pares de sentenças.



Na terceira etapa, comparou-se a TA das 38 palavras homônimas, traduzidas por cada aplicativo, com a TA nas sentenças, visando verificar se o significado seria mantido nas duas condições analisadas. A análise de dados revelou que o desempenho dos aplicativos na TA de palavras homônimas, na busca por palavra, geram traduções diferentes quando a mesma palavra está inserida em uma sentença. No aplicativo HT as palavras ‘como’ e ‘estado’ foram traduzidas para Libras na consulta por palavra com um dos sentidos (pronome interrogativo e divisão político-geográfica de uma república federativa), porém na TA em sentenças houve omissão do sinal COMO, no par de sentenças, e uso de um sinal com um significado não relacionado à palavra ‘estado’ (o sinal TESTE tem pelo menos três significados: ‘fazer um teste’, ‘exemplo’ ou ‘ficar com alguém’). Além disso, a tradução incorreta das palavras na consulta por palavra gerou tradução de um dos significados da palavra homônima ou uso de datilologia em sentenças.

Os aplicativos HT e PDM traduziram incorretamente o significado das palavras quarto e tênis (HT), colher, cunhado, disciplina, estado e revista (PDM), na consulta por palavra, mas em uma das sentenças utilizaram datilologia ou geraram um sinal cujo significado está relacionado à palavra homônima do PB. Além disso, a tradução incorreta do significado na consulta por palavra foi mantida em sentenças. O HT, por exemplo, traduziu incorretamente as palavras ‘cinza’ e ‘disciplina’ na consulta por palavra e em sentenças. Inclusive, na sentença: ‘Odeio o cheiro de cinza de cigarro’ na TA a palavra ‘cinza’ foi traduzida para CINZA, com o sentido ‘cor’ (na consulta por palavra a TA da palavra cinza gerou o sinal SUMIR, conforme descrito nos resultados da primeira etapa das análises). O PDM, por sua vez, traduziu 9 palavras incorretamente na consulta por palavra e manteve a tradução incorreta ou fez datilologia nas sentenças.

Os achados deste estudo apresentam importantes desafios, os quais os sistemas de PLN linguístico-computacionais dos aplicativos analisados, enquanto tradutores automáticos, enfrentam para gerar traduções adequadas de palavras

homônimas isoladas e/ou sentenças. No entanto, observou-se que, considerando o conjunto de 38 palavras verificadas, ambos os aplicativos disponibilizam apenas um dos significados que as palavras homônimas podem expressar. É fundamental que ambos os aplicativos futuramente possibilitem a seus usuários o acesso a diferentes significados, em se tratando de palavras homônimas, em geral o PB e a Libras, não compartilham do mesmo conjunto de palavras/sinais homônimos. Em nível de sentença o significado da palavra homônima foi traduzido adequadamente em parte das sentenças e há sucesso na desambiguação de algumas sentenças. Acredita-se que as potencialidades e as limitações observadas podem contribuir para que os sistemas utilizados sejam aperfeiçoados, de forma que os usuários surdos e ouvintes tenham melhor aproveitamento das ferramentas, seja em contextos sociais ou educacionais bilíngues. A seguir, apresenta-se as considerações finais da pesquisa realizada.

CONSIDERAÇÕES FINAIS

Na análise de TA (PB-Libras) de palavras homônimas em sentenças, a partir de resultados relevantes apresentados em um estudo anterior (CORRÊA, GOMES, RIBEIRO, 2017), identificou-se limitações importantes na TA de palavras isoladas e, conseqüentemente, ainda maiores em nível de sentenças. Os resultados revelaram que na TA de palavras isoladas os aplicativos Hand Talk (HT) e ProDeaf Móvel (PDM) traduziram corretamente apenas um dos significados das palavras homônimas (HT: 89% e PDM: 63%). Nas sentenças, o percentual de acerto do significado das palavras, em ambos os aplicativos, foi menor (HT: 82% e PDM: 60%) do que na TA por palavra isolada e ainda mais reduzido na desambiguação de palavras homônimas (HT: 13% e PDM: 11%). Entre as limitações encontradas constatou-se: (i) a presença de datilologia (alfabeto manual), não sendo considerada a tradução ideal, quando de palavras em PB que possuem seu respectivo sinal em Libras; (ii) traduções incorretas, com uso de sinalização em Libras não correspondente à palavra analisada isoladamente ou em sentenças; (iii) traduções com sinais desconhecidos, os quais, *a priori*, não possuem registro na Libras e (iv) traduções sem tratamento da ambigüidade, característica presente em quaisquer línguas naturais, sejam estas orais ou sinalizadas, sendo que ambos os aplicativos analisados tendem a traduzir corretamente apenas um dos significados de um par homônimo de palavras.

Apesar de os aplicativos de TA serem ferramentas digitais recentes no cenário brasileiro de aplicativos de tradução e apresentarem limitações semelhantes às identificadas em tecnologias de TA de línguas orais, é fundamental o desenvolvimento constante de sistemas linguisticamente mais sofisticados para que

os usuários tenham acesso a traduções de melhor qualidade. Embora a literatura da área também aponte para o entendimento de que, a partir de 1990, a indústria tem se embasado em modelos de TA baseados em fundamentos teóricos (MELO, 2013), este estudo evidenciou a necessidade de inserção de sistemas de desambiguação lexical (SILVA, 2016) nos aplicativos HT e PDM. Infere-se, contudo, que os sistemas de PLN (CHOWDHURY, 2003), atualmente, têm sido considerados um campo fértil (DI FELIPPO; DIAS-DA-SILVA, 2009), até mesmo pelo fato de uso constante de tecnologias digitais na sociedade contemporânea, todavia entende-se que este campo do conhecimento de fato precisa configurar-se como uma área que se constitui como «engenharia do conhecimento da linguagem» (DIAS-DA-SILVA, 1996; DI FELIPPO; DIAS-DA-SILVA, 2009). Embora não tenha sido possível determinar o tipo de método de geração de conhecimento dos sistemas de PLN dos aplicativos analisados, sejam estes por meio de *corpora* (textuais), léxicos ou ontologias, não há dúvida de que a ambiguidade lexical na TA (HUTCHINS; SOMERS, 1992) de línguas naturais caracteriza-se como um desafio contínuo a ser enfrentado por meio de métodos de desambiguação (MIHALCEA, 2006), quando do uso de recursos linguístico-computacionais (DI FELIPPO; DIAS-DA-SILVA, 2009; DORR; JORDAN; BENOIT, 2000).

Devido à facilidade de acesso (aplicativo gratuito e uso no próprio *smartphone*) possivelmente muitas pessoas ouvintes ou surdas fazem consultas rápidas nos aplicativos, em seu dia a dia ou até mesmo utilizam estes recursos para complementar a aprendizagem de sinais da Libras. Dessa forma, os achados desta pesquisa indicam a necessidade de aperfeiçoamento dos sistemas, de ambos os aplicativos, para desambiguação lexical, visando a maior qualidade nas traduções, PB-Libras, de palavras homônimas isoladas e em sentenças, mesmo que a qualidade da TA não alcance o nível de qualidade da tradução produzida por tradutores humanos. Quanto à inserção dos aplicativos HT e PDM em contextos sociais e educacionais bilíngues, de ensino e aprendizagem, acredita-se que a utilização destas tecnologias contemporâneas de TA, em sala de aula bilíngue para surdos, a saber pelos dados apresentados neste estudo, demandam o efetivo acompanhamento pedagógico, por partes de professores proficientes em Libras, de modo que sua utilização possa constituir-se de maneira positiva (CORRÊA; GOMES; RIBEIRO, 2017).

O presente estudo visou a contribuir para os Estudos de Tradução e, mais especificamente, para os estudos sobre TA de uma língua oral para uma língua de sinais, ainda emergentes no Brasil, de modo a progredir na produção de conhecimento científico, quando do diálogo entre Linguística Aplicada e Ciências da Computação, a fim de que o processo tradutório de aplicativos, como os aqui analisados, torne-

se dotado de um desempenho cada vez mais sofisticado. Desta forma, acredita-se que será possível oferecer aos usuários traduções mais análogas à Libras em níveis semântico e sintático, enquanto língua natural e oficial da comunidade surda no Brasil. Uma vez que os estudos sobre TA de PB-Libras ainda são recentes, tornam-se relevantes investigações em todos os níveis estruturais linguísticos (fonológico, morfológico, semântico, sintático) visando avanços tecnológicos para melhorar o sistema e a tradução gerada. Como mencionado, entre os possíveis estudos futuros é importante a verificação da estrutura linguística utilizada nas frases, considerando que a Libras possui uma gramática própria e as relações gramaticais ocorrem no espaço de sinalização (sintaxe espacial).

REFERÊNCIAS BIBLIOGRÁFICAS

- AMARAL, E.; FERREIRA, M.; LEITE, R.; SEVERINO, A. (2010). *Língua Portuguesa*. São Paulo: FTD.
- BATTISON, R. (1978). *Lexical borrowing in American Sign Language*. Silver Spring, M.D: Linstok.
- BERBER SARDINHA, A. P. (org.). (2005). *A língua portuguesa no computador*. Campinas/São Paulo: Mercado de Letras/FAPESP, 296 p.
- BIDARRA, J. (2004). *O léxico no processamento da Linguagem Natural*. Cascavel: Edunioeste.
- BIDARRA, J. (2015). First steps towards a bilingual parallel corpus geared to the treatment of lexical ambiguity on the interface Portuguese - Libras. *Cad. Trad.*, Florianópolis, v. 35, n° 1, pp. 225-250, jan-jun.
- BIDERMAN, M. T. C. (1978). *Teoria linguística (linguística quantitativa e computacional)*. Rio de Janeiro/São Paulo: LTC, 356 p.
- BRENTARI, D.; PADDEN, C. (2001). Native and foreign vocabulary in American Sign Language: a lexicon with multiple origins. In: BRENTARI, D. (ed). *Foreign vocabulary in sign languages*. Mahwah, New Jersey: Laurence Erlbaum Associates, pp. 87-119.
- CABEZUDO, M. A. S. (2015). *Investigação de métodos de desambiguação lexical de sentidos de verbos do português do Brasil*. Dissertação de Mestrado em Ciências - Ciências de Computação e Matemática Computacional, ICMC, USP - São Carlos.
- CASTELLS, M. (1999). *A sociedade em rede*. Volume 1. São Paulo: Paz e Terra.

- CHO, K.; MERRIENBOER, B. V.; GULCEHRE, C.; BAHDABAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: *Conference Of Empirical Methods In Natural Language Processing*, Doha. *Anais...*, Doha: EMNLP, pp. 1724-1734.
- CHOWDHURY, G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37.
- COLLING, J. P.; BOSCARIOLI, C. (2014). Avaliação de tecnologias de tradução Português-Libras visando o uso no ensino de crianças surdas. *RENOTE - Revista Novas Tecnologias na Educação*, v. 12, nº 2.
- CORRÊA, Y.; GOMES, R. P.; RIBEIRO, V. G. (2017). Aplicativos de Tradução Português-Libras na Educação Bilíngue: desafios frente à desambiguação. *RENOTE - Revista Novas Tecnologias na Educação*, v. 15, nº 2, pp. 1-10.
- CORRÊA, Y.; VIEIRA, M. C.; SANTAROSA, L. M. C.; BIAZUS, M. C. V. (2014a). Tecnologia Assistiva: a inserção de aplicativos de tradução na promoção de uma melhor comunicação entre surdos e ouvintes. *RENOTE - Revista Novas Tecnologias na Educação*, v. 12, pp. 1-10.
- CORRÊA, Y.; VIEIRA, M. C.; SANTAROSA, L. M. C.; BIAZUS, M. C. V. (2014b). Aplicativos de tradução para Libras e a busca pela validade social da Tecnologia Assistiva. In: *XXV Simpósio Brasileiro de Informática na Educação (SBIE 2014)*, Dourados. Sociedade Brasileira de Computação SBC, pp. 164-173.
- DIAS-DA-SILVA, B. C. (1996). *A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais*. Araraquara, SP. Tese de Doutorado. Universidade Estadual Paulista, 272 p.
- DI FELIPPO, A.; DIAS-DA-SILVA, B. C. (2009). O processamento automático de línguas naturais enquanto engenharia do conhecimento linguístico. *Calidoscópico*, v. 7, nº 3, pp. 183-191.
- DORR, B. J.; JORDAN, P. W.; BENOIT, J. W. (2000). A survey of current research in machine translation. In: ZELKOWITZ, M. (Ed). *Advances in Computers*, London: Academic Press, pp. 1-68.
- EL MAAZOUZI, E. Z.; EL MOJAHIR, B. E.; AL ACHHAB, M. (2017). A systematic reading in statistical translation: from the statistical machine translation to the neural translation models. *JourJournal of ICT*, v. 16, nº 2, dez. 2017, pp. 408-441.

- FARIA-NASCIMENTO, S. (2009). *Representações lexicais da língua de sinais brasileira: uma proposta lexicográfica*. Tese de Doutorado em Linguística. Universidade de Brasília, Brasília.
- FERREIRA-BRITO, L. (1995). *Por uma gramática de língua de sinais*. Rio de Janeiro: Tempo Brasileiro.
- FENLON, J.; CORMIER, K.; BRENTARI, D. (2018). *The phonology of sign languages*. In: HANNAHS, S. J.; BOSCH, A. (Eds.). *The Routledge Handbook of Phonological Theory*. London and New York: Routledge. pp. 453-475.
- FUNDAÇÃO GETÚLIO VARGAS DE SÃO PAULO (FGV-SP). (2016). 27ª Pesquisa Anual de Administração e Uso de Tecnologia da Informação nas Empresas. Disponível em: <<http://eaesp.fgvsp.br/ensinoconhecimento/centros/cia/pesquisa>>. Acesso em: 17 fev. 2017.
- GALLEY, M.; HOPKINS, M.; KNIGHT, K.; MARCU, D. (2004). What's in a translation rule? In: *Human Language Technology Conference and Meeting of the North American Chapter of the Association for Computational Linguistics, 4*, Edmonton. *Anais...*, Edmonton: HLT-NAACL, pp. 273-280.
- GAUCHE, S. (2013). *Aspectos linguísticos da tradução automática da língua portuguesa para a língua brasileira de sinais (Libras) - uma reflexão inicial*. Monografia de Pós-Graduação Lato Sensu em Revisão de Textos. Centro Universitário de Brasília. Brasília.
- GOIÁS. (2011a). Secretaria da Educação. Centro de Capacitação de Profissionais da Educação e de Atendimento às Pessoas com Surdez. *Apostila Libras I*. Goiás, 51 p.
- GOIÁS. (2011b). Secretaria da Educação. Centro de Capacitação de Profissionais da Educação e de Atendimento às Pessoas com Surdez. *Apostila Libras II*. Goiás, 52 p.
- GOIÁS. (2011c). Secretaria da Educação. Centro de Capacitação de Profissionais da Educação e de Atendimento às Pessoas com Surdez. *Apostila Libras III*. Goiás, 63 p.
- GOIÁS. (2011d). Secretaria da Educação. Centro de Capacitação de Profissionais da Educação e de Atendimento às Pessoas com Surdez. *Apostila Libras IV*. Goiás, 84 p.
- GULCEHRE, C.; FIRAT, O.; XU, K.; CHO, K.; BARRAULT, L.; LIN, H. C.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. (2015). *On using monolingual corpora in neural machine translation*. Disponível em: <<https://arXiv preprint arXiv:1503.03535>> Acesso em: fev. 2018.

- GÜVENIR, H. A.; CICEKLI, I. (1998). Learning translation templates from examples. *Information Systems*, v. 23, n° 6, pp. 353-363.
- HOLMES, J. (1988). *The name and nature of translation studies*. Translation Studies Section, Department of General Literary Studies, University of Amsterdam.
- HUTCHINS, J.; SOMERS, H. L. (1992). *An introduction to machine translation*. San Diego: Academic Press.
- JAKOBSON, R. (1992). *Linguística e Comunicação*. Trad.: Izidoro Blikstein e José Paulo Paes. São Paulo: Cultrix, 1992.
- JEAN, S.; CHO, K.; MEMISEVIC, R.; BENGIO, Y. (2015). On using very large vocabulary for neural machine translation. In: *Annual Meeting Of The Association For Computational Linguistics*, 53, Pequim. *Anais...*, ACL, pp. 1-10.
- JOHNSON, M.; SCHUSTER, M.; LE, Q. V.; KRIKUN, M.; WU, Y.; CHEN, Z.; THORAT, N.; VIÉGAS, F.; WATTENBERG, M.; CORRADO, G.; HUGHES, M.; DEAN, J. (2016). *Google's multilingual neural machine translation system: Enabling zero-shot translation*. Disponível em: <<https://arXiv preprint arXiv:1611.04558>>. Acesso em: jan. 2018.
- KALCHBRENNER, N.; BLUNSOM, P. (2013). Recurrent continuous translation models. In: *Conference Of Empirical Methods In Natural Language Processing*, Washington. *Anais...*, EMNLP, pp. 1700-1709.
- KLIMA, E.; BELLUGI, U. (1979). *The signs of language*. Cambridge, Massachusetts: Harvard University Press.
- LIRA, G. A.; SOUZA, T. A. F de. (2008). Dicionário da Língua Brasileira de Sinais. Versão 2.1 - web. Disponível em: <<http://www.acessibilidadebrasil.org.br/libras/>>. Acesso em: 14 jan. 2018.
- LUONG, M.; SUTSKEVER, I.; LE, V. Q.; VINYALS, O.; ZAREMBA, W. (2015a). Addressing the rare word problem in neural machine translation. In: *Annual Meeting Of The Association For Computational Linguistics*, 53, Lisboa. *Anais...*, ACL, pp. 11-19.
- LUONG, M.; PHAM, H.; MANNING, C. D. (2015b). Effective approaches to attention-based neural machine translation. In: *Annual Meeting Of The Association For Computational Linguistics*, 53, Lisboa. *Anais...*, ACL, pp. 1412-1421.

- LUONG, M.; MANNING, C. D. (2016). Achieving open vocabulary neural machine translation with hybrid word-character models. In: *Annual Meeting Of The Association For Computational Linguistics*, 54, Berlim. *Anais...*, ACL, pp. 1054-1063.
- MACSWEENEY, M.; CAPEK, C. M.; CAMPBELL, R.; WOLL, B. (2008). The signing brain: the neurobiology of sign language. *Trends in Cognitive Sciences*, v. 12, n° 1, pp. 432-440.
- MELO, S. S. C. (2013). Tradução automática e competência tradutória: repensando interseções. *Rónai: Revista de Estudos Clássicos e Tradutórios*, v. 1., n° 1, pp. 60-72.
- MIHALCEA, R. (2006). Knowledge-Based Methods for WSD. In: AGIRRE, E.; EDMONDS, P. (Eds.). *Word Sense Disambiguation: Algorithms and Applications*, pp. 107-132. Springer.
- QUADROS, R. M (2017). *Língua de Herança*. Porto Alegre: Penso.
- QUADROS, R. M.; KARNOPP, L. B. (2004). *Língua de sinais brasileira: estudos lingüísticos*. Porto Alegre: Artmed.
- QUADROS, R. M.; SZEREMETA, J. F.; COSTA, E.; FERRARO, M. L.; FURTADO, O.; SILVA, J. C. (2009). Exame PROLIBRAS. Florianópolis: Universidade Federal de Santa Catarina.
- SAMPIERI, R H.; COLLADO, C.; LUCIO, M. P. B. (2013). *Metodologia de Pesquisa*. Porto Alegre: Penso.
- SANDLER, W. (2005). An overview of sign language linguistics. In: BROWN, K. (Ed.). *Encyclopedia of Language and linguistics*, 2. ed. v. 11. Oxford, UK: Elsevier, pp. 328-338.
- SANDLER, W. (2008). The syllable in sign language: Considering the other natural modality. In: DAVIS, B.; ZAJDO, K. (Eds.). *The syllable in speech production*. New York: Taylor Francis, pp. 379-408.
- SANDLER, W. (2012). The Phonological Organization of Sign Languages. *Language and Linguistics Compass*, v. 6, n° 3, pp.162-182.
- SANTOS, C. J. G. (2010) *Oficina de Pesquisa: tipos de pesquisa*. Disponível em: <http://www.oficinadapesquisa.com.br/apostilas/projeto_rh/_of.tipos_pesquisa.pdf>. Acesso em: 15/01/2018.

- SEGALA, R. R.; QUADROS, R. M. (2015). Tradução intermodal, intersemiótica e interlinguística de textos escritos em Português para a Libras oral. *Cad. Trad.*, Florianópolis, v. 35, n^o especial 2, p. 354-386.
- SENNRICH, R.; HADDOW, B.; BIRCH, A. (2016). Improving neural machine translation models with monolingual data. In: *Annual Meeting Of The Association For Computational Linguistics*, 54, Berlim. *Anais...*, ACL, pp. 86-96.
- SILVA, V. S. S. (2016). *Desambiguação automática de substantivos em corpus do português brasileiro*. Dissertação de Mestrado - USP, São Paulo. 125 f.
- SPECIA, L. (2007). *Uma abordagem híbrida relacional para a desambiguação lexical de sentido na tradução automática*. Tese de Doutorado em Ciências - Ciências de Computação e Matemática Computacional, ICMC, USP - São Carlos.
- STOKOE, W (1960). *Sign language structure*. Silver Spring: Linstok Press, 1978.
- SUTSKEVER, I.; VINYALS, O.; LE, Q. V. (2014). Sequence to sequence learning with neural networks. In: International Conference On Neural Information Processing Systems, 27, Montréal. *Anais...*, Montréal: NIPS, pp. 3104-3112.
- VIEIRA, M. C.; CORRÊA, Y.; CHEIRAN, J. F. P.; SANTAROSA, L. M. C.; BIASUZ, M. C. V. (2014a). Contribuições da Teoria da Aprendizagem Multimídia e da Usabilidade para aprendizagem de Libras e Língua Portuguesa por meio de aplicativos móveis. *RENOTE - Revista Novas Tecnologias na Educação*, v. 12, pp. 1-10.
- VIEIRA, M. C.; CORRÊA, Y.; SANTAROSA, L. M. C.; BIAZUS, M. C. V. (2014b). Além da interação: tradutores automáticos para língua de sinais e a constituição do sujeito surdo. In: *International Society for Augmentative and Alternative Communication ISAAC*, Lisboa, Portugal.
- VIEIRA, M. C.; CORRÊA, Y.; SANTAROSA, L. M. C.; BIAZUS, M. C. V. (2014c). Análise de expressões não-manuais em avatares tradutores de Língua Portuguesa para Libras. In: *XIX Conferência Internacional sobre Informática na Educação - TISE*, Fortaleza. *Nuevas Ideas en Informática Educativa*. v. 10. pp. 172-183.
- WINOGRAD, T. (1972). *Understanding natural language*. NY: Academic Press, 654 p.

Recebido: 15/02/2018

Aceito: 01/03/2018