

Tradução automática de palavras polissêmicas: tecnologias atuais à luz da Linguística Cognitiva

*Machine Translation of
polysemic words: current
technology in light of Cognitive
Linguistics*

Débora Ache BORSATTI (UNISC/IFSul)
deborsatti@gmail.com

Karen Andresa Teixeira SANTORUM (UFRGS/UFSM)
karensantorum13@gmail.com

Alan Ricardo COSTA (UNISC/IFPR)
alan.dan.ricardo@gmail.com

Recebido em: 07 de dez. de 2020.

Aceito em: 01 de jul. de 2021.

BORSATTI, Débora Ache; SANTORUM, Karen Andresa Teixeira; COSTA, Alan Ricardo. Tradução automática de palavras polissêmicas: tecnologias atuais à luz da Linguística Cognitiva. *Entrepalavras*, Fortaleza, v. 11, n. 3, e2173, p. 52-74, set.-dez./2021. DOI: 10.22168/2237-6321-32173.

Resumo: Tendo em vista o aprimoramento constante das tecnologias de tradução automática na atualidade e seu uso por parte de estudantes e aprendizes de línguas adicionais, neste artigo temos por objetivo averiguar como os Tradutores Automáticos (TAs) disponíveis de forma gratuita na internet traduzem palavras polissêmicas desfazendo a ambiguidade de sentidos à luz da abordagem teórica da Linguística Cognitiva (LC). O estudo do funcionamento dos TAs mostra que essas tecnologias digitais estão se tornando mais sofisticadas devido aos sistemas baseados em *corpus* e, mais recentemente, ao sistema neural. Entretanto, tais sistemas ainda apresentam limitações relacionadas às questões linguísticas, dentre elas, a ambiguidade, que pode ser descrita como um dos maiores desafios para um TA hoje. Neste estudo, de caráter qualitativo, analisamos de forma comparativa os resultados de traduções automáticas Português-Inglês de palavras polissêmicas inseridas em frases com diferentes contextos de uso, comparando quatro TAs gratuitos disponíveis na internet: (1) *Google Translate*, (2) *Reverso*, (3) *Collis* e (4)

Wordlingo. Os resultados da investigação indicam a importância da especificação do contexto para a desambiguação das palavras na tradução automática.

Palavras-chave: Tradução automática. Polissemia. Linguística Cognitiva. Semântica lexical. Homonímia.

Abstract: In view of the constant improvement of machine translation technologies today and their use by students and additional language learners, this paper aims to investigate how free online Machine Translators (MTs) work with polysemic words undoing the ambiguity of meanings in the light of the theoretical approach of Cognitive Linguistics (CL). The study of the MTs functioning shows that these digital technologies have become increasingly sophisticated due to corpus-based systems and, more recently, the neural system. However, the MT systems still have limitations related to linguistic issues, among them, ambiguity, which can be described as one of the biggest challenges for MT nowadays. In this qualitative study, we analyzed comparatively the results of Portuguese-English automatic translations of polysemic words inserted in sentences with different contexts of use, comparing four free MTs available on the internet: (1) Google Translate, (2) Reverso, (3) Collis and (4) Wordlingo. The results indicate the importance of specifying the context for word disambiguation in machine translation.

Keywords: Machine translation. Polysemy. Cognitive Linguistics. Lexical semantics. Homonymy.

Introdução

Na atual sociedade em rede, fortemente marcada pelo desenvolvimento acelerado das tecnologias digitais, há uma crescente adesão ao uso de Tradutores Automáticos (TAs), que são “sistemas computadorizados responsáveis pela produção de traduções de uma língua natural para outra com ou sem a assistência humana” (HUTCHINS; SOMERS, 1992, p. 3). Os estudos brasileiros sobre os TAs na seara da Linguística Aplicada (LA) avançaram bastante nos últimos anos (*e.g.* LEFFA, 1993; PAIVA, 2008), e tais tecnologias estão cada vez mais sofisticadas e precisas, graças ao desenvolvimento de novos sistemas no âmbito da Inteligência Artificial ao longo das últimas décadas como o sistema baseado em *corpus* e, mais recentemente, o sistema neural, que têm proporcionado resultados de tradução bem satisfatórios em termos de compreensão (HUTCHINS, 2003; SOMERS, 2003).

A razão para esse aprimoramento dos TAs tem relação com uma tendência atual de desenvolvimento das referidas tecnologias a partir de modelos que tentam se aproximar do processamento da linguagem na mente humana. Em outras palavras, os TAs têm sido desenvolvidos de modo a operar a informação de forma semelhante ao que ocorre na cognição humana.

Entretanto, devido à complexidade que envolve o processamento cognitivo da linguagem natural, há diversos desafios linguísticos para a tradução automática e a inteligência artificial (IA) por trás dela, no que concerne, por exemplo, a tradução de palavras polissêmicas, uma vez que um dos principais desafios para a tradução automática é a ambiguidade, léxica e/ou estrutural (HUTCHINS, 2003).

Palavras com diferentes sentidos/significados¹ (ou polissêmicas) precisam ser contextualizadas para que sejam compreendidas apropriadamente, de modo que uma tradução palavra a palavra não é suficiente, razão pela qual nosso objetivo, neste artigo, é averiguar como os Tradutores Automáticos (TAs) disponíveis de forma gratuita na internet traduzem palavras polissêmicas desfazendo a ambiguidade de sentido à luz dos estudos da Linguística Cognitiva (FILLMORE, 1985; SILVA, 1997; LAKOFF, 1987, 2004; BATORÉO, 2004; GEERAERTS; CUYCKENS, 2007; FERRARI, 2011) destacando as suas inter-relações e dando ênfase para a investigação do significado no âmbito da Semântica Lexical (ULMANN, 1964; GEERAERTS, 1998, 2009; FILLMORE, 1977, 1979, 1982; SILVA, 2010).

Esse objetivo parte de dois pressupostos: (1) o funcionamento dos TAs mostra que essas tecnologias digitais estão se tornando mais sofisticadas devido aos sistemas baseados em *corpus* e, mais recentemente, ao sistema neural; e (2) tais sistemas, contudo, ainda apresentam limitações relacionadas à ambiguidade. Destarte, neste estudo, de caráter qualitativo, analisamos os resultados de traduções automáticas Português-Ingês de palavras polissêmicas inseridas em frases nos diferentes contextos de uso, comparando quatro TAs gratuitos disponíveis na internet: (1) *Google Translate*, (2) *Reverso*, (3) *Collis* e (4) *Wordlingo*.

A linguagem e a construção do significado à luz da linguística cognitiva

Na perspectiva da LC, a linguagem é entendida enquanto meio de conhecimento em conexão com a experiência humana do mundo. Essa concepção trouxe uma mudança paradigmática no âmbito dos estudos linguísticos, contrariando a antecessora teoria gerativa, na qual “a linguagem é o espelho da mente” (CHOMSKY, 1975 *apud* FERRARI, 2011, p. 13), conforme explica Ferrari (2011), a LC defende que a relação entre palavra e mundo é mediada pela cognição. Nesse sentido, o significado não é mais apenas um reflexo direto do mundo, mas uma construção cognitiva por meio da qual a mente experiencia e apreende o mundo.

¹ Utilizaremos ambos os termos neste artigo.

A estrutura formal da linguagem, segundo Geeraerts e Cuyckens (2007), não é autônoma, mas reflexo da organização conceitual, dos princípios de categorização e dos mecanismos de processamento das influências da experiência com o ambiente. Logo, as unidades e as estruturas da linguagem são estudadas não como se fossem entidades autônomas (ao contrário do gerativismo), mas como manifestações de capacidades cognitivas gerais, da organização conceitual, de princípios de categorização, de mecanismos de processamento e da experiência cultural, social e individual. Diante disso, destaca-se a concepção da linguagem humana enquanto um instrumento de organização, processamento e transmissão de informação-*semântico-pragmática*.

Partindo da ideia de que a linguagem se constitui a partir da capacidade cognitiva humana, os principais pressupostos da LC são: (i) os processos de perspectivação que o significado linguístico reflete; (ii) o uso e a experiência como base do significado linguístico; (iii) a natureza dinâmica da gramática; (iv) a categorização linguística construída e organizada na base dos protótipos; (v) a não-autonomia dos sistemas cognitivos, o que leva à construção do sentido em domínios e modelos cognitivos, através dos mecanismos e modelos conceituais (SILVA, 1997; BATORÉO 2004).

Os pressupostos apresentados sugerem que as categorias linguísticas são flexíveis, ou seja, adaptam-se aos diferentes contextos de uso, podendo integrar novas entidades como membros. Em LC, o significado é dinâmico, “construído no conhecimento enciclopédico e configurado em feixes de conhecimento ou domínios, sendo corporizado e encarnado nas necessidades, nos interesses e nas experiências dos indivíduos e nas culturas” (SILVA, 2006, p. 59-60).

A dinamicidade apontada nos revela que estudar linguística a partir do viés cognitivista implica analisar a *semântica pragmática*, ou seja, o significado em uso. O significado é definido como uma estrutura global, constituída por imagens da situação que representa o conteúdo do enunciado presente na mente do falante (BATORÉO, 2004). No âmbito da LC, a linguagem é parte integrante da cognição, fundamentando-se em processos cognitivos, sociointeracionais e culturais, devendo ser estudada no seu uso e no contexto da *conceptualização*, da categorização do processamento mental, da interação e da experiência individual, social e cultural, ou seja, é um conjunto de elementos a ser considerado para a construção do significado das palavras.

Por conseguinte, o significado linguístico e o conhecimento do mundo estão interligados, uma vez que a linguagem é um “meio de interpretar e construir o mundo, de organizar conhecimentos, que refletem as necessidades, os interesses e as experiências dos indivíduos e das suas culturas” (SILVA, 2004, p. 2). O termo *significado*, de acordo com o pensamento de Ullmann (1964), é um dos mais ambíguos e controversos dos estudos linguísticos e, por isso, segundo o autor, é importante que as expressões linguísticas sejam observadas em seu uso real baseado em *corpora*. Diante disso, a contextualização torna-se fundamental nesse processo de análise, especialmente quando se trata de polissemia, como é a proposta deste artigo.

A questão da ambiguidade: polissemia e homonímia

No sentido literal, a origem da palavra polissemia (*poli* = “muitos” e *sema* = “significados”) nos aponta a ideia de que uma determinada palavra ou expressão pode adquirir um novo sentido, além do seu sentido original. A polissemia é estudada através do modelo baseado no uso, que aborda a flexibilidade e variabilidade semânticas (BATORÉO, 2005, p. 237). Nesse paradigma, “a polissemia é o *pivot* da análise semântica” (ULLMANN, 1951, p. 117). Segundo o autor, trata-se de um fenômeno linguístico natural, que sustenta uma característica de economia e flexibilidade para o funcionamento do sistema linguístico, estando diretamente relacionado à frequência das palavras, ou seja, “quanto mais freqüente é uma palavra mais sentidos é possível que tenha” (ULLMANN, 1964, p. 350).

O conceito de polissemia é explicado por Silva como “a associação de dois ou mais sentidos relacionados entre si a uma única forma linguística” (SILVA, 2006, p. 10), isto é, uma palavra ou expressão com vários sentidos. Tendo em vista que definir com clareza os sentidos de uma palavra não é tão simples quanto parece, autores como Kilgarriff (1997) entendem que os sentidos propostos nos dicionários não são suficientes, e que há uma série de limitações para a elaboração de um dicionário, como o formato da impressão, que demanda certa simplificação das definições e impossibilita contemplar todos os exemplos das palavras em uso.

A “polivalência das palavras” é descrita por Ullmann (1964) de duas formas: a polissemia e a homonímia. Com base no autor (ULLMANN, 1964), a polissemia pode levar à ambiguidade em três situações: (1)

no contato entre as línguas, (2) no uso técnico e científico e (3) na fala vulgar, comum. No primeiro caso, a ambiguidade ocorre devido ao empréstimo semântico de uma língua estrangeira; no segundo caso, há uma reutilização de um termo já definido no passado, mas que ganha novas definições, e assim, novos significados. Além disso, quando uma palavra ambígua de uso cotidiano é adotada no contexto técnico e científico, pode gerar confusão e equívocos. Por fim, no terceiro caso (na fala comum), há ambiguidade quando uma palavra apresenta dois ou mais significados em um mesmo contexto.

Ao definir homonímia, Lyons (1977) explica que se trata de um item lexical que carrega, acidentalmente, dois (ou mais) significados distintos e independentes, como no caso de: (1) manga (fruta da mangueira) e (2) manga (parte do vestuário que cobre o braço). Já a polissemia é um único item lexical que carrega vários sentidos diferentes, porém relacionados entre si, como: (3) asa, que pode significar tanto o “membro das aves guarnecido de penas e que serve para voar” como (4) asa “parte saliente de certos utensílios, em geral curva e fechada, que serve para pegar neles”.

Além disso, o autor estabelece dois critérios para diferenciar polissemia de homonímia: a derivação etimológica das palavras e a relação ou a distância entre os significados. No primeiro critério, palavras historicamente derivadas de itens lexicais distintos são consideradas homônimas. Entretanto, nem sempre é fácil identificar com precisão a derivação histórica de certas palavras (LYONS, 1977).

Desse modo, observa-se que o estudo da homonímia e da polissemia não constitui um assunto acabado, sendo a dificuldade de distinção entre ambos os fenômenos uma discussão recorrente entre os pesquisadores do tema. Para Pietroforte e Lopes (2005, p. 131), estudiosos do campo da Semântica Lexical, as palavras polissêmicas “possuem mais de um significado para o mesmo significante”. Os autores exemplificam o conceito através dos possíveis significados da palavra “vela” (objeto para iluminação, peça dos motores, peça de tecido que propulsiona as embarcações, entre outros). O caráter arbitrário do significante é defendido pelos autores a partir da ideia de que o sentido é adquirido nas relações humanas, através da interação nos mais diversos contextos.

Esse fenômeno, segundo Bréal (1992, p. 103), ocorre em todas as línguas das nações civilizadas: “quanto mais um termo acumulou significações, mais se deve supor que ele representa aspectos diversos da atividade intelectual e social”. Para o autor, os diversos sentidos

não se misturam ou se contradizem porque cada um está inserido em um contexto que caracteriza a sua carga semântica. Portanto, um significado só terá sentido em uma determinada situação, dado que os outros significados não existirão (e não se confundem) na mente do interlocutor.

Diante da definição de polissemia, evidenciamos a importância do estudo da língua em uso para a compreensão de uma análise linguística nesse campo. Logo, traduzir esse tipo de fenômeno linguístico é uma tarefa complexa. É preciso utilizar o recurso de Desambiguação Lexical de Sentidos (DLS), que corresponde à tarefa de determinar o sentido mais adequado de uma palavra de acordo com o contexto de uso (sentença, texto, documento etc.), considerando um conjunto finito de possíveis sentidos/significados (AGIRRE; EDMONDS, 2006).

Partindo da compreensão de que a ambiguidade lexical pode ser categorial, no que se refere à categoria gramatical da palavra, ou semântica, conforme explica Specia (2007), a Desambiguação Lexical se apresenta como uma etapa da tradução automática (COLLING; BOSCARIOLI, 2014; CABEZUDO, 2015), sendo uma tarefa computacional que visa a resolução do problema da ambiguidade das palavras (SILVA, 2016). Mihalcea (2006) descreve três heurísticas para a desambiguação: (1) sentido mais frequente, que implica em usar o sentido que ocorre com maior frequência em um *corpus* registrado ou em um repositório de sentidos; (2) sentido por colocação – usar palavras que, frequentemente, co-ocorrem com a palavra a ser desambiguada na busca do sentido exato; e (3) sentido por discurso – usar o sentido de uma palavra em um componente discursivo (sentença, parágrafo ou texto) como o mesmo em todas as ocorrências.

Os seguintes passos são considerados na criação de modelos para a tarefa de DLS: 1) Determinação do conjunto de palavras a desambiguar: todas as palavras de determinada(s) classe(s) gramatical(is), um subconjunto específico de palavras etc. 2) Definição de todos os possíveis sentidos de cada palavra. 3) Criação de um mecanismo para atribuir a cada ocorrência da palavra o sentido mais apropriado, incluindo decisões sobre a metodologia do mecanismo, os tipos de conhecimentos que serão utilizados e as suas fontes, entre outras, e a criação ou adaptação dos recursos necessários. 4) Avaliação desse mecanismo, independentemente da aplicação e/ou no contexto de alguma aplicação específica. Utilizamos estes passos como suporte para a análise deste estudo.

O processo de desambiguação, em se tratando de tradução automática, ocorre no processamento da informação conforme o *input* que é fornecido, ou no caso da Tradução Automática Neural, como as redes são treinadas. Frente a isso, torna-se necessária a compreensão dos diferentes tipos de tradução automática e do funcionamento destes tipos de sistemas para uma análise posterior aprofundada.

Sobre o funcionamento dos tradutores automáticos

Desenvolvidos na década de 1940, os primeiros TAs fizeram parte da história do avanço da ciência da computação como a primeira aplicação não numérica nessa área, conforme relata Nirenburg (1987). O propósito dos TAs naquele período estava relacionado à Guerra Fria, uma vez que os governos americano e inglês investiram na tradução por máquinas a fim de conseguir informações imediatas da inteligência soviética. Os sistemas desenvolvidos para tradução automática nas décadas de 50 e 60, embora fundamentados na linguística computacional e na inteligência artificial, apresentavam resultados aquém das expectativas (MATEUS, 1995), o que levou a um período de descrédito e interrupção de pesquisas em TA.

O panorama mudou na década de 80, com um cenário propício para o desenvolvimento da tradução automática, não apenas devido ao avanço no campo da informática, mas pelo surgimento de teorias linguísticas como a gramática gerativa de Chomsky, bem como da criação da Comunidade Econômica Europeia. Contudo, foi na década de 90, por meio da rede mundial de computadores, que a revolução digital se instalou, transformando radicalmente o modo de comunicação entre as pessoas. Consequentemente, houve um aumento da demanda por traduções que impulsionou as pesquisas no campo da tradução automática (HUTCHINS, 2006).

Neste estudo, abordamos a tradução totalmente automática, ou seja, conforme explica Hutchins (2002), aquela que ocorre quando o sistema traduz todo o texto sem a intervenção humana, que resulta em uma tradução “crua”, também denominada “tradução informativa” ou “tradução para apreensão de sentido”. Um TA pode ser programado para tradução em duas línguas (bilíngue) ou mais línguas (multilíngue).

Os sistemas podem ser agrupados em duas categorias: os sistemas baseados em regras (*rule-based system*) e os sistemas baseados em *corpus* (*corpus-based system*). A primeira categoria (1) contempla

três abordagens principais: (i) a primeira é a tradução direta, cujo funcionamento atua como em um grande dicionário bilíngue (HUTCHINS, 2003), ou seja, o texto é traduzido palavra a palavra, sem uma análise das estruturas sintáticas ou da relação de significados entre as palavras; (ii) a segunda é a abordagem por transferência, que, conforme Hutchins (2003), pode ter programas separados para a transferência lexical e para transferência estrutural, isto é, vai além da tradução palavra por palavra; (iii) e a terceira é a abordagem da interlíngua, desenvolvida na década de 80 com base nos princípios de uma linguística universal, que entende que é possível converter textos da língua de partida em representações sintático-semânticas comuns em diferentes línguas. Esse tipo de sistema apresenta muitas falhas na tradução de frases e textos que contêm palavras polissêmicas, uma vez que não contempla o contexto comunicativo, apenas as estruturas e/ou significados das palavras isoladamente.

A segunda categoria (2), o sistema baseado em *corpus*, se divide em dois modelos: o modelo baseado em estatística e o modelo baseado em exemplos. O modelo estatístico baseia-se fundamentalmente no alinhamento de palavras, expressões e sequências de palavras num *corpus* bilíngue paralelo e no cálculo das probabilidades para que as palavras/expressões em uma determinada frase de uma língua correspondam a uma ou mais palavras da frase equivalente na língua alvo (SOMERS, 2003). O sistema baseado em exemplos, por sua vez, opera através da comparação do *input* com um *corpus* de exemplos já traduzidos, de modo a reproduzir as correspondências mais próximas como modelo de tradução para o texto alvo. Esse modelo busca resolver problemas através de soluções dadas para problemas semelhantes anteriores e isso, de acordo com Somers (1998), equivale à forma de pensamento humano e, por conseguinte, ao modo como a tradução humana é realizada. Assim, as traduções geradas através desse sistema são geralmente mais fluentes e menos literais.

Mais recente, o modelo de Tradução Automática Neural (KALCHBRENNER; BLUNSOM, 2013; LUONG *et al.*, 2015; JEAN *et al.*, 2015) é um sistema de computador que opera como um cérebro, isto é, imitando as redes neurais biológicas. Esse tipo de TA ancora-se no modelo de Aprendizagem de Máquinas (AM), ou *machine learning*, e alimenta-se de dados para aprender progressivamente e conseqüentemente melhorar a sua tradução, de forma semelhante às redes neurais biológicas.

Conforme explica Prati (2006), a aprendizagem é fundamental para o desenvolvimento da Inteligência Artificial por ser o fator determinante do comportamento inteligente de um sistema.

A AM estuda métodos computacionais com o objetivo de melhorar o desempenho na aquisição de novos conhecimentos e novos meios de organizar os nossos conhecimentos. Esse tipo de estudo pode levar a um entendimento mais aprofundado da nossa própria inteligência e do processo de aprendizagem em si.

Apesar de a TA neural contar (assim como o sistema estatístico) com limitações relacionadas a informações sintáticas e semânticas (LUONG *et al.*, 2015), ela conta também com vantagens. Dentre as principais, destacamos o pouco conhecimento linguístico necessário para gerar o tradutor, a otimização conjunta de toda a rede (e não de modelos separados, como ocorre na TA estatística) e o fato de gerar um conhecimento mais compacto do que o conjunto de regras usado na TA baseada em regras.

Nesse sentido, os TAs que utilizam os sistemas híbridos (que aliam e integram diferentes modelos) tendem a apresentar textos mais coesos e coerentes, além de traduções que privilegiam o contexto. No entanto, nos questionamos: até que ponto esses sistemas são eficientes em gerar uma boa tradução de palavras polissêmicas? Buscamos responder a esse questionamento ao longo da próxima seção, por meio dos resultados das análises de frases com palavras polissêmicas traduzidas por distintos TAs.

Metodologia

Em termos metodológicos, desenvolvemos o estudo optando por uma metodologia qualitativa, de cunho interpretativista, que reflete, segundo Leffa (2006), uma tendência da área da Linguística Aplicada no Brasil. Afora isso, com base em Costa *et al.* (2020), entendemos que a pesquisa qualitativa, por seu caráter fenomenológico e interpretativista, pode contribuir significativamente com o campo de ensino-aprendizagem de línguas mediado por tecnologias digitais, como os TAs. Análises semelhantes têm sido aplicadas com o intuito de testar as ferramentas de tradução automática em diferentes campos dos estudos linguísticos (BORSATTI; GABRIEL, 2019; HUMBLÉ, 2019; KARNAL, 2019).

Na presente pesquisa, selecionamos quatro TAs, a saber: (1) *Google Translate*, (2) *Reverso*, (3) *Collise* e (4) *Wordlingo*. A escolha por tais TAs ocorreu principalmente por questões de gratuidade e de disponibilidade na Web, o que foi considerado juntamente com o critério da amplitude

de uso. Após a escolha dos TAs, definimos três palavras polissêmicas da Língua Portuguesa (LP) para serem traduzidas para a Língua Inglesa (LI) e formulamos frases em contextos diferentes de uso. As palavras “banco”, “letra” e “cabo” foram escolhidas a partir de uma pesquisa de palavras polissêmicas por representarem a diversidade de sentidos necessários para análise. Vale ressaltar que a restrição do número de palavras está relacionada à limitação na extensão deste artigo. Antes de apresentarmos a análise dos resultados é importante conhecer cada um dos tradutores e seus sistemas de funcionamento.

O *Google Translate* faz parte da empresa Google Inc. e oferece traduções instantâneas de forma online e gratuita. Permite a tradução de palavras, frases, textos ou páginas da internet (PAIVA, 2008). A princípio, o tradutor do Google era uma versão do tradutor Systran, baseado em regras. Em 2007, a empresa lançou seu próprio sistema de tradução utilizando a abordagem estatística (KOROSEC, 2011). A partir de 2016, a empresa passou a usar o sistema estatístico por um sistema “neural”, ou seja, as traduções são processadas por uma rede de inteligência artificial capaz de traduzir sentenças inteiras de uma só vez, ao invés de palavra por palavra. Nesse método, a ferramenta aprende sozinha a realizar traduções mais precisas, levando em conta o contexto de cada sentença.

De acordo com as informações disponibilizadas na página do *Reverso*², o tradutor opera a partir de um sistema baseado em Inteligência Artificial (IA), chamado Tradução Automática Neural (*Neural Machine Translation*). Essa tecnologia é capaz de aprender com seus recursos linguísticos (textos em ambos os idiomas e seus glossários). O *Reverso* ainda fornece soluções de tradução automática baseada em regras e estatística.

O *Collis* é primordialmente um dicionário com uma história que data de 1824 a partir da primeira edição. O tradutor online, situado na mesma plataforma do dicionário online³, utiliza a tecnologia desenvolvida pela *Microsoft*, que é híbrida, ou seja, utiliza tanto o método estatístico quanto o neural.

Finalmente, o *Wordlingo* é um site que oferece serviços diversos de tradução. Dentre esses serviços estão a tradução humana, a tradução automática paga e a versão gratuita. Os sistemas apresentam diferença

² Disponível em: <https://www.corporate-translation.reverso.com/translation-technologies>. Acesso em: 3 nov. 2019.

³ Disponível em: <https://www.collinsdictionary.com/pt/about>. Acesso em: 20 ago. 2020.

na qualidade dos textos, uma vez que a tradução humana é realizada por profissionais com conhecimento avançado nas línguas em questão. Já a tradução automática paga conta com auxílio de tradutores humanos, uma vez que é abastecida por um sistema de *corpus* monitorado por profissionais. Por fim, a tradução gratuita é 100% automática, ou seja, não apresenta interferência humana, gerando traduções mais propensas a erros. Essa última foi analisada neste estudo. É importante destacar que este TA foi escolhido porque é o único deles que não utiliza o sistema neural e, portanto, poderemos comparar a qualidade com os três demais analisados. É possível notar que o funcionamento dos três primeiros tradutores é semelhante: todos utilizam o sistema híbrido, aproveitando os benefícios que os diferentes modelos oferecem, visando um resultado mais completo.

Análise dos resultados

A primeira palavra escolhida para a análise das traduções foi o substantivo “banco”, que na Língua Portuguesa (LP) pode significar: 1) banco (instituição financeira para operações monetárias); 2) banco (assento com ou sem encosto, geralmente para várias pessoas). Existem outros significados para essa palavra conforme o contexto como, por exemplo, um conjunto, base ou um acúmulo de algo (Ex.: banco de dados, banco de sangue, banco de areia, entre outros). Além disso, existe o verbo “bancar” (ser o responsável pela banca de jogo; tomar ares de, fingir; passar por: bancar o bobo), entretanto, neste estudo, como se trata de um artigo que possui limitações em termos de extensão, optamos por analisar apenas substantivos. Em Língua Inglesa (LI) há diferentes palavras empregadas para cada sentido de “banco” que destacamos acima na língua portuguesa, sendo banco 1) *bank*; 2) *bench* e 3) banco de dados – *database*. Assim, a polissemia em “banco” presente na LP não é correspondente na LI.

A partir da seleção desta palavra, elaboramos frases com os diferentes sentidos/significados, algumas deixando em aberto o contexto, como no exemplo da frase 1, abaixo, e outras especificando melhor a situação em questão. Um tradutor humano se ancora no contexto e no seu amplo conhecimento de uso de ambas as línguas para realizar a tarefa de desambiguação. Vejamos a seguir como o processo de desambiguação ocorre a partir de TAs.

Análises com a palavra “banco”

Quadro 1 – Frases com a palavra “banco”

| FRASE | Google | Reverso | Collins | Wordlingo |
|--|---|---|--|--|
| 1) Depois de muito caminhar, ele avistou um banco. | After a long walk, he saw a bank . | After a long walk, he spotted a bank . | After much walking, he spotted a bank . | After very walking, it sighted a bank . |
| 2) Depois de muito caminhar, ele avistou um banco e parou para descansar um pouco. | After a long walk, he saw a bench and stopped to rest for a while. | After a long walk, he spotted a bench and stopped to rest for a while. | After much walking, he spotted a bench and stopped to rest for a while. | After very walking, it sighted a bank and stopped to rest a little. |
| 3) O banco está quebrado. | The bank is broken. | The bank is broken. | The bank is broken. | The bank is broken. |
| 4) O banco está quebrado, por isso não podemos sentar. | The bank is broken, so we can't sit. | The bench is broken, so we can't sit down. | The seat is broken, so we can't sit down. | The bank is broken, therefore we cannot seat. |
| 5) O banco está sujo. | The bank is dirty. | The bank is dirty. | The bank is dirty. | The bank is dirty. |
| 6) Não podemos sentar aqui porque o banco está sujo | We can't sit here because the bench is dirty | We can't sit here because the bank is dirty. | We can't sit here because the bench is dirty. | We cannot seat here because the bank is dirty. |
| 7) O sistema funciona a partir de um banco de dados. | The system works from a database . | The system works from a database . | The system works from a database . | The system works from a database . |

Fonte: Os autores. Todos os acessos realizados em: 20 ago. 2020.

Percebemos que na frase 1 fica dúbio se nos referimos ao banco de sentar ou à agência bancária, ou seja, o caráter ambíguo da palavra é evidente. No entanto, os TAs traduziram como instituição financeira. É importante notar que a frase apresentada pelo *Wordlingo*, diferentemente dos demais TAs, apresenta erros gramaticais como o uso do advérbio “very” antes de verbo (*walk*), o que não é correto em LI. Além disso, o pronome “he (ele)” é traduzido como “it” (não usado como referência para pessoas). A segunda frase contextualiza a situação em que uma pessoa estava caminhando e parou para descansar, acionando o *frame*⁴ do leitor que será capaz de realizar a inferência de que a palavra

⁴ Em termos gerais, *frames* são blocos de conhecimentos pré-existentes, armazenados na memória e que representam situações estereotipadas. Para Koch e Travaglia (1996), os *frames*

“banco” está referida à condição de assento. Nesse caso, o “leitor” é o sistema do TA, que será acionado a partir de estatística ou pelo sistema neural que fará a combinação das palavras na frase e, portanto, quanto mais elementos contextuais, mais possibilidade de acerto.

O resultado da tradução da frase 2 evidencia a simplicidade do sistema do *Wordlingo* em relação aos demais TAs analisados, uma vez que a palavra *bank* é utilizada apesar de todos os outros elementos linguísticos da frase que contribuem para a contextualização. O exemplo em questão demonstra a diferença de um sistema baseado em regras e os sistemas baseados em *corpus* ou os sistemas neurais, que são muitos mais sofisticados e complexos, se aproximando mais do funcionamento da cognição humana. Ademais, as falhas gramaticais encontradas na primeira frase são novamente apresentadas.

A frase 3 é bastante ambígua, isso porque a palavra “quebrado” além de significar algo partido ou fragmentado, também pode ser utilizada no sentido de financeiramente falido. Essa ambiguidade ocorre em ambas as línguas, LI e LP. Desse modo, a ambiguidade da palavra “quebrado” pode ser a uma explicação plausível para o fato de todos os TAs apresentarem, na referida frase, a tradução de banco relacionada a dinheiro. Diferentemente, na frase 4, “O banco está quebrado, por isso não podemos sentar”, acrescentamos o verbo “sentar”, que mobiliza a ideia da palavra “banco” como um assento. Neste caso, os TAs *Google* e *Wordlingo* não fizeram a desambiguação, continuando com a palavra “*bank*” apresentada na frase 1. Já o *Reverso* desambiguou, resultando na palavra “*bench*” e o *Collins* usou a palavra “*seat*”, que teria o mesmo sentido de assento. Esse exemplo mostra que nem sempre o sistema TA, por mais complexo e desenvolvido que seja, é capaz de realizar o processo de desambiguação lexical a partir de um contexto mais específico.

Outro exemplo em que deixamos a dúvida contextual aparece na frase 5, “O banco está sujo”, que resulta em todos os TAs o uso da

fazem parte do nosso conhecimento de mundo como conjuntos de conhecimentos armazenados na memória sob um “rótulo”, sem que haja qualquer ordenação entre eles. Um exemplo é o *frame* prostituição infantil, cujo tema nos faz recordar palavras como: estrangeiros, carnaval, casas noturnas, pedofilia, etc. Seguindo a teoria da Semântica de Frames, duas noções ancoram o conceito de *frame*: a de perspectiva e a de protótipo. Para Fillmore (1977, 1979, 1982), o falante ativa suas experiências relativas ao contexto com o objetivo de entender. Desse modo, cada cena é ativada na mente por uma palavra do contexto que, neste processo, também focaliza uma parte dela, contemplando uma perspectiva. As cenas estão armazenadas na mente, porém, nas instâncias concretas, elas são representadas de modo particular, logo, o foco está em uma porção da cena, não em sua totalidade. Segundo Fillmore (1977, p. 59), portanto, “os significados são relativizados às cenas”.

palavra “bank” na tradução. Neste caso, podemos pensar na hipótese de que, sem um contexto específico, os TAs fazem a desambiguação de acordo com o sentido mais frequente. Ao inserirmos maior contextualização na frase 6, “Não podemos sentar aqui porque o banco está sujo”, observamos a desambiguação correta por meio da palavra “*bench*” no *Google* e no *Collins*. A palavra “*bank*” foi obtida no *Reverso* e *Wordlingo*. O último sentido explorado para “banco” foi de “banco de dados”, *database* em inglês, foi testado na frase 7. Neste caso, dada a especificidade contextual, evidenciamos a desambiguação acurada em todos os TAs.

Análises com a palavra “letra”

A segunda palavra polissêmica escolhida foi “letra”, que na LP pode significar: 1) caligrafia de uma pessoa; 2) composição musical escrita (letra de música); 3) sinal gráfico do alfabeto. A LI dispõe de três palavras diferentes para esses sentidos: 1) *handwriting* (caligrafia); 2) *lyrics* (letra de música) e 3) *letter* (alfabeto).

Quadro 2 – Frases com a palavra “letra”

| FRASE | Google | Reverso | Collins | Wordlingo |
|---------------------------------------|---|--|--|--|
| 1) Que letra linda! | What a beautiful handwriting! | What a beautiful handwriting! | What a beautiful letter! | How pretty letter! |
| 2) Você tem uma letra linda. | You have beautiful handwriting. | You have a beautiful handwriting. | You have a beautiful handwriting. | You have a pretty letter. |
| 3) Eu adoro a letra desta música. | I love the lyrics of this song. | I love the lyrics to this song. | I love the lyrics to this song. | I adore the letter of this music. |
| 4) Eu adoro esta letra! | I love this letter! | I love this lyrics! | I love that letter! | I adore this letter! |
| 5) Qual a primeira letra do seu nome? | What is the first letter of your name? | What's the first letter of your name? | What's the first letter of your name? | Which the first letter of its name? |
| 6) Ele fez um gol de letra. | He scored a goal. | He hit a home run. | He scored a goal. | It made a letter goal. |

Fonte: Os autores. Todos os acessos realizados em: 20 ago. 2020.

Identificamos que quanto maior a ambiguidade, maior a chance de erros na tradução. A frase 1 representa bem esta proporção, uma vez que ela poderia tanto estar relacionada ao sentido de caligrafia quanto à letra de música. O *Google* e o *Reverso* traduzem

como *handwriting* (caligrafia). Já o *Collins* e o *Wordlingo* traduzem como *letter* (símbolo do alfabeto). Ao realizarmos um simples exercício de tradução inversa da frase, ou seja, trocar para (LI-LP) obtivemos a seguinte frase: “que carta linda!”, uma vez que a palavra *letter* tem mais de um significado em LI, podendo representar tanto uma letra do alfabeto quanto uma carta.

Na frase 2, a inserção de “você tem” especifica um pouco melhor o contexto, dando uma pista mais clara de que se trata da grafia de uma pessoa. Tanto a tradução do *Google* quanto a do *Reverso* e a do *Collins* apresentam uma resposta mais coerente para o contexto, ou seja, *handwriting*. Em contrapartida, a frase traduzida pelo *Wordlingo* apresenta a palavra *letter*, o que gera incoerência, pois não faz sentido alguém ter ou possuir um símbolo do alfabeto. O contexto de música é introduzido na frase 3, gerando um resultado acurado para os três primeiros tradutores. O *Wordlingo* foi exceção, resultando na palavra *letter* novamente.

A falta de contextualização da frase 4 evidencia a ambiguidade da palavra “letra” na LP. Novamente, os três primeiros TAs (*Google*, *Reverso* e *Collins*) traduzem “letra” como *letter* nesse contexto, o que seria pouco provável de ser dito por alguém em relação à uma letra do alfabeto. Os exemplos propostos permitem a observação de que *letter* parece ser o único *input* para a palavra “letra” para o *Wordlingo*, ou seja, esse TA desconhece os outros significados da palavra em LP.

O exemplo da frase 5 é bastante contextualizado, “letra do nome”, contando com um forte indicativo de referência ao símbolo do alfabeto, de modo que a falta de ambiguidade na frase facilita a acurácia dos resultados nos quatro TAs. Diferentemente, a frase 6 apresenta a expressão “gol de letra”, que é restrita ao universo contextual do futebol e significa um gol a partir de um movimento considerado bastante complexo para o jogador. A expressão é conhecida em LI como *rabona* ou *rabona kick*. Devido à especificidade do *corpus*, nenhum dos quatro TAs apresentou a expressão correta. A resposta do *Reverso* se aproximou do contexto futebolístico através da expressão *hit a home run*, que é usada no beisebol e também pode se referir a um “gol de placa” no futebol, o que é diferente de um “gol de letra”. O *Google* e o *Collins* se aproximam do contexto do futebol como se a frase significasse “ele fez um gol”, o que foge do sentido original, pois não se trata de um simples gol, mas um gol especial, de letra. A resposta do *Wordlingo* é “a letter goal”, uma tradução literal, que não faz sentido em LI.

Análises com a palavra “cabo”

A terceira palavra escolhida para a análise foi o substantivo “cabo”. Os possíveis significados aqui abordados foram 1) cargo do exército (*army corporal* em LI); 2) haste de uma ferramenta de trabalho, como vassoura ou martelo (*handle* em LI); 3) fio que conduz eletricidade (*cable* em LI) e 4) característica geográfica de uma porção litorânea de continente, que se estende ou se projeta na direção do mar, em forma de ponta (*cape* em LI).

Quadro 3 – Frases com a palavra “cabo”

| FRASE | Google | Reverso | Collins | Wordlingo |
|--------------------------------------|--|---|--|--|
| 1) Meu sobrinho é cabo. | My nephew is a cable . | My nephew is a corporal . | My nephew is a corporal . | My nephew is handle . |
| 2) Meu sobrinho é cabo do exército. | My nephew is an army corporal . | My nephew is a corporal in the army. | My nephew is an army corporal . | My nephew is handle of the army. |
| 3) O cabo da vassoura está quebrado. | The broomstick is broken. | The broom handle is broken. | The broom handle is broken. | The handle of the broom is broken. |
| 4) Onde está o cabo? | Where is the cable ? | Where is the cable ? | Where is the cable ? | Where it is the handle ? |
| 5) Onde está o cabo João? | Where is cape John? | Where is corporal John? | Where is corporal John? | Where it is the handle João? |
| 6) Onde está o cabo do martelo? | Where is the hammer handle ? | Where is the hammer handle ? | Where's the hammer cable ? | Where it is the handle of the hammer? |
| 7) O Cabo Canaveral fica na Florida. | Cape Canaveral is in Florida | Cape Canaveral is in Florida. | Cape Canaveral is in Florida. | The Canaveral Handle is in the Flowery one. |

Fonte: Os autores. Todos os acessos realizados em: 20 ago. 2020.

Na frase 1, o termo mais apropriado ao contexto em LI seria *corporal* ou *army corporal*. A palavra “sobrinho” indica que se trata de uma pessoa do sexo masculino. Nesse contexto, as possibilidades de *cable*, *handle* e *cape* são descartadas. Observamos que os resultados do *Reverso* e do *Collins* estão corretos, enquanto o resultado do *Google* é a palavra *cable* e o do *Wordlingo* é *handle*. Quando adicionamos “do exército” logo após a palavra “cabo”, na frase 2, o *Google* traduz corretamente, mostrando que a contextualização auxilia na precisão da desambiguação.

O contexto bastante específico da frase 3 em “cabo da vassoura” resulta em respostas acuradas na desambiguação para todos os TAs,

havendo pequenas diferenças de termos como *broom handle* e *broom stick*, ambos de acordo com o contexto de uso. É importante destacar que, embora o *Wordlingo* apresente a palavra adequada ao contexto da frase, a tradução da frase é feita de forma literal, ou seja, palavra por palavra, evidenciando um sistema baseado em regras. Possivelmente um tradutor humano com pouco conhecimento da estrutura da LI traduzisse a frase desta forma, diferentemente de alguém com alto nível de proficiência em LI.

Na frase 4, o leitor pode inferir que se trata de um fio elétrico, mas como falta informação contextual, há ambiguidade. Os três primeiros TAs apresentam a palavra *cable*, que estaria em consonância com o contexto de alguém que está à procura de um objeto. Por outro lado, o resultado gerado pelo *Wordlingo* é a palavra *handle*, como ocorre em todas as outras frases traduzidas com a palavra “cabo”, mostrando que esse é o único *input* para a palavra neste TA.

Adicionamos um nome próprio na frase 5 com o intuito de trazer novamente o sentido do “militar”, porém o *Google* não traduz dessa forma, tendo como resultado o nome do incidente geográfico. Uma hipótese para esta resposta é que o sistema tenha operado com ênfase no início da pergunta, “Where is” (que pode ser traduzida como “onde é” ou “onde está”), sugerindo uma possível correspondência geográfica.

A característica específica de cabo do martelo na frase 6 propiciou a desambiguação em três TAs. Apenas o *Collins* apresentou o sentido de “fio” para “cabo”, ao invés da parte do martelo. Por fim, na frase 7, três TAs traduziram corretamente o incidente geográfico “Cabo Canaveral”, certamente por representar um local específico que provavelmente já foi traduzido antes e, portanto, está disponível nos exemplos. Reiteramos que a tradução do *Wordlingo* é *handle*, comprovando a hipótese de um único *input* para a palavra “cabo”, o que impede a desambiguação. Além disso, o TA traduz o nome do estado norte-americano da Flórida como “Flowery”, ou seja, apresenta uma tradução extremamente literal e incorreta, tornando a frase sem sentido em LI.

Considerações finais

A tradução humana é uma atividade que exige alto nível de proficiência linguística, além de vivência na língua e conhecimento acerca do assunto que está sendo traduzido. A tradução automática, na atual sociedade digital, se utiliza das traduções humanas já realizadas

para, a partir delas, “aprender” e realizar futuras traduções com base nos dados disponíveis. Nesse sentido, os atuais sistemas tecnológicos de Inteligência Artificial, com base em *corpus* e numa perspectiva neural, buscam “imitar” o processamento mental humano para atingir traduções acuradas.

A complexidade subjacente ao processamento da linguagem e das relações sociais que envolvem o consentimento dos diferentes contextos é possivelmente o maior desafio, hoje, para esses sistemas de tradução. Dentre as características mais complexas, em termos de estudos semânticos, destacamos a ambiguidade presente nas palavras. Percebemos através das análises das palavras polissêmicas que, quanto mais informações contextuais o TA recebe, mais acurada será a tradução, isto é, quanto mais específico e menos ambíguo for o texto proposto, maior a chance de um resultado coerente.

Outro ponto importante observado é que os sistemas operacionais divergem em termos de resultados: as traduções do *Reverso*, *Google* e *Collins* (que utilizam sistemas baseados em *corpus* e tradução neural), de modo geral, se assemelham. As traduções do *Wordlingo* (que utiliza o sistema baseado em regras) apresentam mais erros, tanto no aspecto de estrutura sintática quanto léxico-semântico, evidenciando as limitações desse tipo de sistema em termos de *input* (entrada) e processamento da informação.

Ao nos depararmos com a frase “ela viu o banco”, por exemplo, parece difícil identificar se estamos falando do banco no sentido de “assento” ou de “uma agência bancária”. Entretanto, quando lemos “Ela sentou no banco”, uma rede prototipada de conhecimentos ou *frames* é acionada em nossa mente. Sendo assim, podemos argumentar que a perspectiva da LC – cuja premissa é de que a linguagem existe enquanto uma construção social e está para além de um sistema de regras – contribui não só para o entendimento humano do uso da linguagem, mas também para o entendimento do funcionamento dos TAs na atual sociedade digital. Enquanto tecnologias amplamente empregadas, para fins comunicativos e educacionais, principalmente, é importante que os TAs sejam estudados e constantemente analisados em vista de contribuir para o entendimento das inter-relações entre o desenvolvimento da inteligência artificial e o processamento da linguagem humana.

Referências

AGIRRE, E.; EDMONDS, P. "Introduction". In: AGIRRE, E.; EDMONDS, P. (ed.). **Word Sense Disambiguation: Algorithms and Applications**. Heidelberg: Springer, 2006. p. 1-28.

BATORÉO, H. J. **Linguística Portuguesa: Abordagem Cognitiva**. Lisboa: Universidade Aberta, 2004. CD-ROM.

BATORÉO, H. J. Como não 'Pôr o pé em ramo verde' ou do papel da polissemia na construção do sentido. In: RIO-TORTO, M. G.; FIGUEIREDO, O. M.; SILVA, F. (org.). **Estudos em Homenagem ao Professor Doutor Mário Vilela**. Porto: FLUP, 2005. p. 237-251.

BORSATTI, D. A.; GABRIEL R. A tradução automática de textos científicos como suporte pedagógico para o desenvolvimento da compreensão leitora em inglês para propósitos acadêmicos. **Organon**, Porto Alegre, v. 34, n. 66, p. 1-19, 2019.

BRÉAL, M. **Ensaio de Semântica**. Tradução de Aída Ferrás, Eduardo Guimarães, Eleni Jacques Martins e Pedro de Souza. São Paulo: Pontes/Educ, 1992.

CABEZUDO, M. A. S. **Investigação de métodos de desambiguação lexical de sentidos de verbos do português do Brasil**. 2015. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2015.

COLLING, J. P.; BOSCARIOLI, C. Avaliação de tecnologias de tradução Português-Libras visando o uso no ensino de crianças surdas. **RENOTE – Revista Novas Tecnologias na Educação**, Porto Alegre, v. 12, n. 2, p. 1-10, dez. 2014.

COSTA, A. R.; FIALHO, V. R.; BEVILÁQUA, A. F.; OLIVEIRA, E. 10 anos de Jornada de Elaboração de Materiais, Tecnologias e Aprendizagem de Línguas: estado da arte. In: LEFFA, V. J.; FIALHO, V. R.; BEVILÁQUA, A. F.; COSTA, A. R. (org.) **Tecnologias e ensino de línguas: uma década de pesquisa em Linguística Aplicada**. Santa Cruz do Sul: EDUNISC, 2020. v. 1. p. 19-43.

FILLMORE, C. J. The case for case reopened. In: COLE, P.; SADOCK, J. M. (org.). **Syntax and semantics**. New York: Academic Press, 1977.

FILLMORE, C. J. Topics in lexical semantics. In: COLE, P. **Currents Issues in Linguistic Theory**. Indiana University Press, 1979.

FILLMORE, C. J. Frame semantics. In: SEOUL INTERNATIONAL CONFERENCE ON LINGUISTICS. The linguistic society of Korea. **Linguistics in the Morning Calm: selected papers from SICOL**. Korea: Hanshin, 1982.

FILLMORE, C. J. Frames and the semantics of understanding. **Quaderni di Semantica**, [s. l.], v. VI, n. 2, p. 222-254, dez. 1985.

FERRARI, L. **Introdução à linguística cognitiva**. São Paulo: Contexto, 2011.

GEERAERTS, D. Hundred years of lexical semantics. In: VILELA, M.; SILVA, F. (org.). Actas do 1º Encontro Internacional de Linguística Cognitiva. Porto, 1998.

GEERAERTS, D.; CUYCKENS, H. Introducing Cognitive Linguistics. In: GEERAERTS, D.; CUYCKENS, H. **The Oxford Handbook of Cognitive Linguistics**. New York: Oxford University Press, 2007. p. 3-21.

HUMBLÉ, P. Machine translation and poetry: the case of English and Portuguese. **Ilha do Desterro**, Florianópolis, v. 72, n. 2, p. 41-56, maio/ago. 2019.

HUTCHINS, J. Commercial Systems: the state of the art. In: SOMERS, H. L. (ed.). **Computers and translation: a translator's guide**. Amsterdam: John Benjamins, 2002. p. 367.

HUTCHINS, J. Machine Translation. In: MITKOV, R. **The Oxford Handbook of Computational Linguistics**. Oxford: Oxford University Press, 2003.

HUTCHINS, J. Machine Translation: history. In: BROWN, K. **Encyclopedia of Language & Linguistics**. Oxford: Elsevier, 2006. p. 375-383.

HUTCHINS, J.; SOMERS, H. L. An introduction to Machine Translation. London: Academic Press, 1992.

JEAN, S.; CHO, K.; MEMISEVIC, R. ; BENGIO, Y. On using very large target vocabulary for neural machine translation. In: **Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing**. July 26-31, 2015. Beijing: Association for Computational Linguistics, 2015. p. 1-10.

KALCHBRENNER, N.; BLUNSOM, P. Recurrent continuous translation models. In: **Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing**. October 18-21, 2013. Seattle: Association for Computational Linguistics, 2013. p. 1700-1709.

KARNAL, A. R. Um estudo comparativo da desambiguação lexical realizada por tradutor eletrônico e tradutores humanos. **Cadernos de Tradução**, Porto Alegre, n. esp., p. 24-34, 2019.

KILGARRIFF, A. What is Word Sense Disambiguation Good For? In: **Proceedings of the Natural Language Processing Pacific Rim Symposium '97**. Phuket, Thailand: 1997.

KOROSEC, M. K. The Internet, Google Translate and Google Translator Toolkit Nuisance or Necessity in Translator Training? In: **Tralogy I**, Session 3 – Training translators/La formation du traducteur. Paris, 2011. Disponível em: <http://lodel.irevues.inist.fr/tralogy/index.php?id=113>. Acesso em: 25 jul. 2020.

LAKOFF, G. **Women, fire and dangerous things: what categories reveal about the mind**. Chicago: University of Chicago Press, 1987.

LEFFA, V. J. A compreensão de textos gramaticalmente incorretos: implicações para a Tradução Automática. In: **Encontros do GT de Linguística Aplicada**. Florianópolis, Santa Catarina. 1993.

LEFFA, V. J. (org.). **Pesquisa em Linguística Aplicada: temas e métodos**. Pelotas: Educat, 2006.

LUONG, M.; PHAM, H.; MANNING, C. Effective approaches to attention based neural machine translation. *In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisboa: Association for Computational Linguistics, 2015. p. 1412-1421.

LYONS, J. **Linguagem e Lingüística**: uma introdução. Tradução de Wanda Ramos. Lisboa: Editorial Presença/Martins Fontes, 1977.

MATEUS, M. H. M. Tradução automática: um pouco de história. *In: MATEUS, M. H. L.; BRANCO, A. H. (org.). Engenharia da Linguagem*. Lisboa: Edições Colibri, 1995, p. 115-120.

MIHALCEA, R. Knowledge-Based Methods for WSD. *In: AGIRRE, E.; EDMONDS, P. (ed.). Word Sense Disambiguation: Algorithms and Applications*. Heidelberg: Springer, 2006. p. 107-132.

NIRENBURG, S. **Machine Translation – Theoretical and Methodological Issues**. Cambridge: Cambridge University Press, 1987.

PAIVA, V. M. O. Internet e sistemas de busca: ampliando o universo de professores e aprendizes de língua inglesa. *In: MACÍEL, R. F.; ARAUJO, V. A. (org.). Ensino da Língua Inglesa: contribuições da Lingüística Aplicada*. Campo Grande: UNAES, 2008. p. 43-58.

PRATI, R. C. **Novas abordagens em aprendizado de máquinas para a geração de regras, classes desbalanceadoras e ordenação de casos**. 2006. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e Computação, Universidade de São Paulo, São Carlos, 2006.

PIETROFORTE, A. V. S.; LOPES, I. C. A semântica lexical. *In: FIORIN, J. L. (org.). Introdução à Linguística: princípios de análise*. 4. ed. v. 2. São Paulo: Contexto, 2005. p. 111-135.

SILVA, A. S. A Semântica de DEIXAR. **Uma Contribuição para a Abordagem Cognitiva em Semântica Lexical**. 1997. Dissertação (Doutoramento) – Faculdade de Filosofia de Braga, Universidade Católica Portuguesa, Braga, 1997.

SILVA, A. S. Protótipos, imagens e metáforas, ou o experiencialismo da linguagem e do pensamento. *In: DINIS, A.; CURADO, J. M. (org.). Consciência e cognição*. Braga: Publicações Faculdade de Filosofia da UCP, 2004. p. 79-96.

SILVA, A. S. **O mundo dos sentidos em português**: polissemia, semântica e cognição. Coimbra: Almedina, 2006.

SILVA, A. S. Palavras, significados e conceitos: o significado lexical na mente, na cultura e na sociedade. **Cadernos de Letras da UFF**, Niterói, Dossiê Letras e Cognição, n. 41, p. 27- 53, 2010.

SILVA, V. S. S. **Desambiguação automática de substantivos em corpus do português brasileiro**. 2016. Dissertação (Mestrado em Semiótica e Linguística Geral) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2016.

SOMERS, H. Machine translation. *In*: BAKER, M. (ed.). **Routledge Encyclopedia of Translation Studies**. London: Routledge, 1998. p. 136-139.

SOMERS, H. Machine translation: latest developments. *In*: MITKOV, R. **The Oxford Handbook of Computational Linguistics**. Oxford: Oxford University Press, 2003.

SPECIA, L. **Uma abordagem híbrida relacional para a desambiguação lexical de sentido na tradução automática**. 2007. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2007.

ULLMANN, S. **Semântica**: uma introdução à ciência do significado. Tradução de J. A. Osório Mateus. Lisboa: Fundação Calouste Gulbenkian, 1964.

ULLMANN, S. **The Principles of Semantics**. Oxford: Blackwell, 1951.