unesp UNIVERSIDADE ESTADUAL PAULISTA

"JÚLIO DE MESQUITA FILHO"

Faculdade de Ciências e Letras

Campus de Araraquara - SP

MARIA PAULA FIORIM PIRUZELLI

AMBIGUIDADES LINGUÍSTICAS NO INGLÊS E A TRADUÇÃO AUTOMÁTICA INGLÊS-PORTUGUÊS: um estudo exploratório

ARARAQUARA – S.P.

2011

MARIA PAULA FIORIM PIRUZELLI

AS AMBIGUIDADES LINGUÍSTICAS NO INGLÊS E A TRADUÇÃO AUTOMÁTICA INGLÊS-PORTUGUÊS: um estudo exploratório

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras – UNESP/Araraquara, como requisito para a obtenção do título de Mestre em Linguística e Língua Portuguesa.

Linha de pesquisa: Estudos do léxico e Análise Fonológica, Morfossintática, Semântica e Pragmática

Orientador: Bento Carlos Dias da Silva

Bolsa: FAPESP

ARARAQUARA – SP
2011

Piruzelli, Maria Paula Fiorim

Ambiguidades linguísticas no inglês e a tradução automática: um estudo exploratório / Maria Paula Fiorim Piruzelli. – Araraquara
155 f.

Dissertação (Mestrado em Linguística e Língua portuguesa) – Faculdade de Ciências e Letras – Universidade Estadual Paulista, Araraquara, 2011.
Orientador: Bento Carlos Dias da Silva

1. Linguística 2. Processamento Automático de Língua Natural 3 . Ambiguidades linguísticas

MARIA PAULA FIORIM PIRUZELLI

AS AMBIGUIDADES LINGUÍSTICAS NO INGLÊS E A TRADUÇÃO AUTOMÁTICA INGLÊS-PORTUGUÊS: um estudo exploratório

Dissertação de Mestrado, apresentada ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras – UNESP/Araraquara, como requisito para obtenção do título de Mestre em Linguística e Língua Portuguesa.

Linha de pesquisa: Estudos do léxico e Análise Fonológica, Morfossintática, Semântica e Pragmática

Orientador: Bento Carlos Dias da Silva

Bolsa: FAPESP

Data da defesa: 30/06/2011

MEMBROS COMPONENTES DA BANCA EXAMINADORA:

Presidente e Orientador: Prof. Dr. Bento Carlos Dias da Silva

UNESP/FCL/Araraquara.

Membro Titular: Profa. Dra. Ariani Di Felippo

UFSCAR/São Carlos.

Membro Titular: Profa. Dra. Paula Tavares Pinto Paiva

UNESP/FCL/Araraquara.

Local: Universidade Estadual Paulista

Faculdade de Ciências e Letras

UNESP – Campus de Araraquara

AGRADECIMENTOS

Aos meus pais, Ivânia e Valdir, meus primeiros orientadores, por sempre me guiarem para o caminho correto e por me incentivarem e apoiarem também nesta caminhada acadêmica. Com certeza, outros grandes objetivos a conquistar serão ainda estimulados por eles.

Ao meu orientador, Professor Bento, por sua disponibilidade constante, por sua paciência infinda e orientações sempre esclarecedoras, partilhando seu vasto conhecimento a qualquer momento.

Ao meu companheiro, Aguinaldo, que esteve ao meu lado em todos os momentos, sempre paciente.

Aos amigos, em especial Erasmo, Maira e Taísa, a quem recorri nos momentos em que precisei e que nunca me negaram atenção; quando o auxílio era impossível, apoiaram-me e aconselharam-me.

À CAPES, pelo apoio financeiro no período de junho a agosto de 2009 e à FAPESP, que financiou este trabalho entre setembro de 2009 e abril de 2011.

RESUMO

Ao abordar a tradução automática (TA), esta pesquisa insere-se no domínio do Processamento Automático de Língua Natural (PLN), defendendo a hipótese de que se faz necessária a inclusão, nos sistemas de TA, de um módulo de resolução de ambiguidades que ao menos forneça ao usuário as diferentes leituras de frases ambíguas. Norteia-se por estes dois objetivos: (i) propor, de modo sistemático e com exemplificações, uma tipologia dos principais tipos de ambiguidade linguística do inglês; (ii) avaliar, de modo exploratório, o impacto das ambiguidades estudadas em (i) sobre a TA com o auxílio do *Google Tradutor*; (iii) explorar estratégias de resolução de ambiguidades. A metodologia que norteia o estudo distribui a investigação nos domínios linguístico e linguístico-computacional. No primeiro, com foco no inglês, introduz-se a problemática geral das ambiguidades linguísticas que se manifestam na TA de frases/parágrafos de textos em inglês para o português e procede-se à sistematização, com exemplos, dos tipos de ambiguidades linguísticas selecionados na literatura sobre PLN. No segundo domínio, avalia-se, por meio do Google Tradutor, o impacto dos tipos de ambiguidades sistematizados no domínio linguístico sobre as TAs de frases selecionadas do *corpus* paralelo inglês-português selecionado para o estudo e a adequação das TAs, tendo como parâmetros, para a análise da qualidade das TAs, as diferentes leituras potenciais das frases e as traduções realizadas por humanos e selecionadas do *corpus*. As considerações finais resumem o trabalho, apresentam possíveis desdobramentos da investigação e enfatizam o papel decisivo que um módulo de detecção e resolução de ambiguidades pode exercer na qualidade das TAs ao resolvê-las ou apontar, para o usuário, diferentes leituras de uma sentença ambígua no contexto dos tipos de ambiguidade que é capaz de detectar.

Palavras-chave: ambiguidades linguísticas; tradução automática; processamento automático de língua natural.

ABSTRACT

On taking up the study of linguistic ambiguities and machine translation (MT), this thesis unfolds against the backdrop of NLP (Natural Language Processing). It claims that MT systems, no matter whether they are couched in the symbolic or the statistical frameworks, should be equipped with a sort of disambiguation resolution module. In particular, it focuses on: (i) describing and exemplifying the core types of linguistic ambiguities in English in the context of the MT from English into Portuguese; (ii) explore, by means of the use of *Google Translator*, the impact of ambiguities on its Portuguese output from original English sentences; and (iii) sketch ambiguity resolution strategies for selected types. From the methodological point of view, the study is carried over the linguistic and computational-linguistic domains. In the former, it is discussed the different sorts of linguistic ambiguities by means of sample data collected from the NLP literature on ambiguity resolution investigation in English texts. In the latter, sample sentences were first selected in a parallel, English-Portuguese *corpus*, and then tested for their adequacy against the impact of the ambiguity types they illustrate. Sentence potential readings and their human translation counterparts were taken as parameters for the evaluation. The final remarks sum up the study, points out some of its further developments, and stress that the sample ambiguity resolution strategies explored in the study illustrates the key role a MT ambiguitiy detection and resolution module would play to improve the MT quality by either predicting them or by offering the user alternative readings of ambiguous sentences to choose from.

Keywords: linguistic ambiguities; machine translation; natural language processing.

LISTA DE FIGURAS

Figura 1	Uma evidência de polissemia: o verbo *parar*	10
Figura 2	Um *screenshot* da interface de consulta do COMPARA	50
Figura 3	Estruturas sintáticas distintas de *saw the man in the house with a telescope*.	60
Figura 4	Estrutura (i)	61
Figura 5	Estrutura (ii)	62
Figura 6	Estrutura (iii)	63
Figura 7	Estrutura (iv)	64
Figura 8	Estrutura (v)	65
Figura 9	Estrutura sintática que ilustra a noção de domínio local	102
Figura 10	Estruturas sintáticas que ilustram a noção de C-comando	103

LISTA DE QUADROS

Quadro 1	As 27 situações de teste	37
Quadro 2	Tabulação dos resultados do teste com o verbo *look up*	39
Quadro 3	Tabulação dos resultados do teste com o verbo *look for*	41
Quadro 4	Situações de teste que mostram que o verbo *look up* não é reconhecido nestas estruturas	43
Quadro 5	Resumo da comparação entre original e tradução dos verbos *look up* e *look for*	45
Quadro 6	Os tipos e subtipos de ambiguidade no inglês	55
Quadro 7	Síntese: tipos/subtipos, leituras e frases-exemplo com traduções	81
Quadro 8	Confronto entre a THs e a TA	87
Quadro 9	Resumo das estratégias de resolução de ambiguidades discutidas	106

Sumário

Introdução **10**

1. Contextualizando a investigação das ambiguidades linguísticas **14**

1.1 Os objetivos e a metodologia da pesquisa *14*

1.2 Os tradutores humanos e a tradução automática *15*

1.3 Um passeio pela história da TA *17*

1.4 A TA e o PLN *21*

1.5 Os dois paradigmas de estudo do PLN *25*

1.6 Os tipos de sistemas de TA *28*

1.7 A ambiguidade linguística e a TA *32*

2. Fatorando a investigação nos domínios linguístico e linguístico-computacional **34**

2.1 Investigação no domínio linguístico *34*

2.1.1 Apresentando o *Google Tradutor* 34

2.1.2 Apresentando o COMPARA 48

2.2 Explorações no domínio linguístico-computacional *51*

3. Classificando as ambiguidades linguísticas **52**

3.1 As ambiguidades linguísticas no inglês no âmbito da TA *54*

3.2 Uma tipologia das ambiguidades linguísticas no inglês *55*

3.2.1 (A) Ambiguidade lexical 56

3.2.2 (B) Ambiguidades estruturais 58

3.2.3 (B2) Ambiguidade de localização e de preenchimento de lacunas 70

3.2.4 (B3) Ambiguidade analítica 70

3.2.5 (B4) Ambiguidade de escopo da quantificação 76

3.2.6 (C) Ambiguidade anafórica ou referencial 78

3.2.7 (D) Ambiguidade temática 79

3.2.8 (E) Ambiguidade pragmática 80

3.3 Quadro síntese das discussões *80*

4. Discutindo as ambiguidades linguísticas na TA inglês-português **85**

4.1 Confronto entre a tradução humana e a tradução automática *86*

4.2 Estratégias de resolução de ambiguidades linguísticas no inglês *91*

4.3 Explorações do impacto das ambiguidades linguísticas na TA inglês-português *108*

4.3.1 (A) Ambiguidades lexicais 109

4.3.2 (B) Ambiguidades estruturais..... 126

4.3.3 (B4) Ambiguidades de escopo da quantificação 141

4.3.4 (C) Ambiguidades anafóricas/referenciais 142

5. Considerações finais 146

REFERÊNCIAS 150

Introdução

Este estudo teve a sua motivação no trabalho de iniciação científica relacionado com o desenvolvimento de parte da base de verbos da rede WordNet.Br, uma base relacional de itens/expressões lexicais do português brasileiro. Os elementos básicos dessa rede de verbos são os seus nós, linguisticamente preenchidos por *synsets*, isto é, conjuntos de verbos sinônimos em pelo menos um contexto de ocorrência, que lexicalizam um conceito, e as relações de troponímia, causa e acarretamento que se estabelecem entre eles (DIAS-DA-SILVA, OLIVEIRA, MORAES, 2002; DIAS-DA-SILVA, 2010). Nesse trabalho, uma das atividades foi a construção de *synsets* de verbos, tarefa que permitiu observar que grande parte dos itens analisados podia pertencer a mais de um *synset*, não sendo, portanto, verdadeiras unidades lexicais[1] (pois estas são definidas como monossêmicas), fato que sinalizou a propriedade da homonímia/polissemia[2] dos itens lexicais. Essa situação ilustra-se por meio dos *synsets* da Figura 1.

Figura 1 – Uma evidência de polissemia: o verbo *parar*

S='Glosa2038' = "Manobrar um veículo para dentro de uma vaga"
{**encostar:** Alberto, o representante de Benetton em Israel, [encostou seu carro ao lado do «checkpoint».,
estacionar: Ele I]estacionou o carro ao lado do Shopping Paineiras.,
parar: Lá, ele I]parou o carro no estacionamento descoberto do local e o trancou com o controle remoto.,
parquear: Em seguida, I]parqueiam o veículo e deixam-no armazenado, disponível para a venda em toda a América Latina.,}
S='Glosa2046' = "Finalizar um evento"
{**cessar:** De repente ele I]cessou a oração dele.,
deter: Então ele I]deteve o movimento e manteve o ponto de luz imóvel por algum tempo.,
estancar: Com coragem e decisão, I]estancou o processo de privatização das empresas de energia elétrica pelo governo federal.,
interromper: Com a perda do marido, D. Leda [interrompeu as negociações.,
parar: A empresa afirma que apenas um pequeno grupo [parou o trabalho.,}

---

[1] Ruppenhofer et al.(2010) define uma unidade lexical como um item lexical associado a um significado específico e único ("*A lexical unit (LU) is a pairing of a word with a meaning*") (RUPPENHOFER et al., 2010, p. 05). Por exemplo, para o item lexical *casa*, postulam-se diferentes unidades lexicais: `casa` (moradia), `casa` (posição em um jogo de tabuleiro), `casa` (fenda no tecido por onde se passa um botão), etc.

[2] Considerações sobre homonímia/polissemia serão feitas na seção 3.2, p. 49.

Na Figura 1, *parar* é componente de ambos os *synsets*, o que mostra que esse verbo é polissêmico, com pelo menos dois sentidos diferentes: 'manobrar um veículo para dentro de uma vaga' e 'finalizar um evento'. Essa propriedade, por sua vez, acarreta o fenômeno da ambiguidade de formas das línguas naturais, as "ambiguidades linguísticas", o que despertou o interesse no estudo da manifestação desse fenômeno no âmbito da tradução automática (TA), já que a presença da homonímia/polissemia nos dados (morfema, item léxico, sintagma, frase ou parágrafo) de entrada de um sistema de TA, em geral, traz problemas para a qualidade da TA esperada desses dados.

Cumpre, de início, destacar que, por abordar a TA, este estudo insere-se no domínio de estudos do Processamento Automático de Língua Natural (PLN), defendendo que se faz-se necessária a proposição de sistemas de TA que sejam munidos de um módulo de detecção e/ou resolução de ambiguidades que, não sendo capaz de desambiguizar os dados de entrada, ao menos possa fornecer as diferentes leituras desses dados para o usuário.

Nesta dissertação, discusão desse empreendimento, em um primeiro momento, descrevem-se os objetivos da pesquisa, separando aqueles incluídos no domínio linguístico do estudo daqueles que se incluem no domínio linguístico-computacional. Define-se a tarefa da tradução, com o objetivo de apresentá-la ao leitor e argumentar a complexidade desse processo, que envolve a aplicação de conhecimentos de naturezas distintas, mostrando ainda que ela pode ser abordada de dois pontos de vista, o do tradutor humano e o da TA, enfatizando que o foco do estudo é o da TA.

Para localizar o leitor no tempo, apresenta-se uma breve história da TA, comentando que a ideia inicial, equivocada, concebia o processo tradutório como uma simples substituição de símbolos. Nesse panorama, identificam-se os principais fatores que limitaram o desenvolvimento da TA, lembrando que as metas traçadas inicialmente eram pretensiosas demais e subestimavam o processo tradutório, ignorando que nele se incluem os diferentes tipos de conhecimento acima apontados.

Nesse contexto, insere-se o estudo do PLN, porque é com a iniciativa da TA que ele se desenvolveu. Em outras palavras, o estudo do PLN, extrapolando a investigação da TA, passa também a investigar outros tipos de sistemas como os corretores ortográficos e gramaticais, os sumarizadores automáticos, os sistemas de busca e recuperação de informação em bases de textos, entre outros. Enfatiza-se que se trata de um domínio multidisciplinar, que exige um trabalho colaborativo entre linguistas e cientistas da computação, bem como entre profissionais de áreas como Matemática e Filosofia para alcançar os seus objetivos;

colaboração nem sempre bem-sucedida, posto que não são poucos os conflitos de natureza epistemológica e de valores científico-acadêmicos entre linguistas e informatas.

Contextualizada a TA dentro do PLN, caracterizam-se os dois paradigmas (o empiricista/estatístico e o racionalista/simbólico) de investigação do PLN, avaliando-se que o primeiro, embora advogue a construção de sistemas de TA com base eminentemente estatística, seus articuladores, mesmo considerando um "trabalho sujo", reconhecem que é necessário partir de conhecimentos linguísticos estruturados e já construídos.

Na sequência, mapeiam-se os três tipos de sistemas de TA (direto, por transferência e por interlíngua) no âmbito do paradigma racionalista que norteia este estudo.

Estabelecidos os fundamentos teórico-metodológicos, analisam-se os diferentes tipos de ambiguidades linguísticas, um dos maiores obstáculos para a TA. Os dados da análise (frases e parágrafos) são extraídos de *corpus* e da própria literatura estudada sobre TA. A sistematização das ambiguidades, por sua vez, é feita em termos dos diferentes tipos de ambiguidade estudados e da confrontação entre as traduções automáticas de ocorrências desses tipos e as respectivas traduções elaboradas por tradutores humanos.

Feita esta introdução, passa-se ao desenvolvimento do trabalho, que se organiza em cinco seções. A Seção 1 contextualiza a investigação das ambiguidades linguísticas no âmbito do PLN/TA, em relação à tradução humana, e apresenta o contexto histórico, os paradigmas e tipos de sistemas de TA. A seção 2 apresenta a metodologia de pesquisa do trabalho e os seus objetivos, que estão divididos nos domínios linguístico e linguístico-computacional, e descreve os dois recursos centrais que possibilitaram o desenvolvimento da parte empírica do trabalho: o *Google Tradutor* (PAVÃO JUNIOR, 2010), com o seu comportamento idiossincrático, e o *corpus* paralelo COMPARA (FRANKENBERG-GARCIA, SANTOS, 2003; 2002). A Seção 3 aborda, de modo sistemático, o fenômeno das ambiguidades linguísticas e os problemas decorrentes da sua manifestação no processo de TA (inglês-português), descrevendo e exemplificando os tipos e subtipos de ambiguidade linguística. A Seção 4 desenvolve o estudo empírico e exploratório, examinando a manifestação dos tipos e subtipos de ambiguidade na TA de frases do inglês para o português, "calculadas" pelo *Google Tradutor,* sob a luz das leituras possíveis dessas frases e das traduções feitas por tradutores humanos selecionadas no *corpus* COMPARA, e, quando possível, apresentando estratégias de resolução para os tipos de ambiguidade descritos. A Seção 5, apresenta as considerações finais sobre as análises dos tipos de ambiguidade que foram ilustrados com frases do *corpus,* o resultado do confronto entre as TAs geradas em dois

momentos de teste e resume os tipos para os quais não foi possível discutir uma estratégia de resolução, sugerindo possibilidades de trabalhos futuros.

1. Contextualizando a investigação das ambiguidades linguísticas

Esta seção contextualiza a investigação das ambiguidades linguísticas no âmbito do PLN/TA, apresentando, primeiramente, os objetivos da pesquisa, nos domínios linguístico e linguístico computacional, e a metodologia adotada. Em seguida, esta seção contextualiza a TA em relação à tradução humana e aos estudos do PLN, traçando um panorama histórico da TA e descrevendo os paradigmas e tipos de sistema de TA. Por fim, a seção aborda a manifestação das ambiguidades linguísticas nos estudos da TA.

1.1 Os objetivos e a metodologia da pesquisa

Objetivos

Inserindo-se no contexto dos estudos do PLN, esta pesquisa investiga a manifestação do fenômeno das ambiguidades linguísticas no processo de tradução automática do inglês para o português e é norteada por estes três objetivos:

(i) descrever, de modo sistemático e com exemplificações, a tipologia dos principais tipos de ambiguidade linguística do inglês (e do português);

(ii) avaliar de modo exploratório a TA com o auxílio do *Google Tradutor*;

(iii) explorar estratégias de resolução de ambiguidades.

Metodologia

Seguindo a metodologia de estudos do PLN proposta em Dias-da-Silva (1996. 2010), o estudo distribui-se em dois domínios complementares: o linguístico e o linguístico-computacional. No domínio linguístico, (a) descrevem-se e organizam-se, em termos de tipos e subtipos, as ocorrências de ambiguidades linguísticas no inglês; (b) ilustram-se cada um desses tipos e subtipos de ambiguidades com comentaráios e exemplos extraídos da literatura (frases e parágrafos); (c) localizam-se, no *corpus* paralelo (inglês-português) selecionado para o estudo, ocorrências dos mesmos tipos e subtipos já sistematizados em (a) e, quando não se encontram ocorrências nesse *corpus*, propõem-se traduções livres. No domínio linguístico-

computacional, (i) investiga-se o impacto das ambiguidades que se manifestam no inglês sobre o grau da qualidade da TA do inglês para o português. Para isso, submetem-se, ao *Google Tradutor,* todas as frases e parágrafos selecionados em (b) para ilustrar os tipos e subtipos de ambiguidade, comparando-se as TAs do *Google Tradutor* com as frases traduzidas por tradutores humanos extraídas do *corpus* paralelo, e, finalmente, (ii) estudam-se as principais estratégias computacionais de resolução dos diferentes tipos de ambiguidade sistematizados em (a).

1.2 Os tradutores humanos e a tradução automática

Considerando-se que a questão da tradução é central para os objetivos do estudo, é essencial pontuarem-se, de início, características gerais e distintivas da tradução humana e da tradução automática, já que se reconhece que as traduções feitas por tradutores humanos e por tradutores automáticos são processos distintos.

Em linhas muito gerais, pode-se definir a tradução como um processo que realiza a transferência ou a substituição do conteúdo de um texto escrito em uma língua natural para outra língua natural distinta. Contudo, de acordo do Martins e Nunes (2005), é preciso admitir que a tradução não é apenas o transporte do conteúdo lexicogramatical de um texto em uma língua para outra diferente, porque o significado do texto não reside somente na léxicogramática que o constitui. Enfatizam ainda que o significado expresso por um enunciado não está contido nele mesmo, mas depende da interpretação que o leitor pode fazer, recorrendo, para isso, a outros conhecimentos além do conhecimento puramente linguístico. Argumentam, por isso, que, sem esses conhecimentos extras, a frase

> não passa de um amontoado de símbolos isolados, cada um dos quais apontando para múltiplos diferentes significados (como prevê o próprio dicionário), todos à espera de uma decisão (a ser tomada apenas pelo leitor) que os permitirá combinarem-se uns aos outros para que possam vir a constituir um todo (MARTINS, NUNES, 2005, p. 11).

Embora o papel do leitor seja crucial, Hatim e Mason (1990) reconhecem, em consonância com este estudo, que a tradução pode ser abordada a partir de dois pontos de vista: o do **tradutor humano** e o da **tradução automática**, esta realizada ou auxiliada por sistemas de tradução automática. Desnecessário ressaltar que, neste trabalho, o foco é exclusivamente a TA e que se admite que as questões de natureza linguística no âmbito da TA

podem ser analisadas a partir de aspectos eminentemente linguísticos e linguístico-computacionais. Tipicamente, considera-se que o termo TA define os sistemas computacionais desenvolvidos para traduzir textos entre duas ou mais línguas naturais, com ou sem a supervisão humana.

Hutchins (2001) afirma que é preciso atentar para as diferentes necessidades exigidas na TH e na TA, não percebidas durante a década de 80. Sua percepção, argumenta, poderia ter ajudado no desenvolvimento de programas que suprissem as diferentes necessidades da tradução. Distingue ele entre: (i) a TA não-supervisionada pelo tradutor humano, com pós-edição; (ii) a TA não-supervisionada, sem pós-edição, que fornece "versões aproximadas"; e (iii) ferramentas de TA que auxiliam a tarefa de tradução. Daí, não considerar as diferenças de precisão do resultado da tradução e de independência da supervisão contribuiu para a manutenção dos inúmeros equívocos sobre a TA, ao ser esta comparada com a tradução realizada por tradutores humanos.

Desnecessário ainda dizer que os sistemas de TA não têm o objetivo de substituir os tradutores humanos, principalmente diante da limitação da tecnologia existente e da complexidade da linguagem humana, posto que, com a tecnologia desenvolvida até o momento, não possuem independência para operarem sem a supervisão humana. Contudo, eles certamente ajudam os tradutores humanos a aumentar a produtividade, funcionando como uma ferramenta de auxílio ao trabalho dos tradutores humanos, tornando-o menos repetitivo e mais ágil, já que a demanda por tradução de textos de natureza técnica e referencial é maior do que a que os tradutores humanos conseguem atender (ALFARO, 1998).

Como lembram Hatim e Mason (1990), a busca da tradução completamente automática e de boa qualidade deveria ser o ponto de contato entre linguistas e tradutores profissionais, mas o que, de fato, acontece é que as duas áreas se desenvolvem separadamente e, em geral, são conflitantes. Os autores argumentam que as pesquisas iniciais sobre TA deveriam ter levado em conta o processo global da tradução humana. Entretanto, o foco dessas pesquisas restringiu-se na análise gramatical e na resolução de aspectos pontuais da polissemia lexical em uma pequena amostra de frases. Isso ocorreu, dizem os autores, porque havia uma suposição de que era preciso superar contrastes estritamente estruturais entre as línguas. Essa excessiva preocupação com aspectos sintáticos fez com que a semântica e o contexto de uso fossem ignorados e tratados como além das capacidades do processamento automático.

No estágio de desenvolvimento em que se encontram hoje, os sistemas de TA são utilizados, geralmente, nas tarefas para as quais se aceitam traduções aproximadas ou então para as quais os resultados serão pós-editados por um tradutor humano ou ainda para as quais se aceita o uso de uma sub-linguagem limitada (JURAFSKY; MARTIN, 2000).

Como se mostrará a seguir, a tradução feita ou auxiliada por computadores há muito tem sido discutida pelos pesquisadores do PLN, porque, envolvendo a comparação entre línguas, implica a complexa e laboriosa modelagem tanto da competência linguística quanto do comportamento linguístico, este último englobando, além de aspectos puramente linguísticos, aspectos cognitivos, históricos e sócio-culturais. Entretanto, além dos obstáculos inerentes ao tratamento computacional das línguas naturais, devido a essa complexa modelagem do conhecimento e do comportamento linguísticos humanos, Kurzweil (1990) chama atenção para a dificuldade adicional posta pelos diferentes pares de línguas a serem tratados por um sistema de TA, que, individualmente, colocam problemas específicos de tradução.

1.3 Um passeio pela história da TA

Situar a TA nos estudos linguístico-computacionais é recordar um pouco da sua história.

Foi durante a década de 40 que a TA começou a figurar na área das tecnologias linguísticas; a partir de 1946, durante o período de pós-guerra, ou seja, durante o período da guerra-fria entre os Estados Unidos da América e os seus aliados e a antiga União Soviética, a TA recebeu maior atenção dos norte-americanos e ingleses.

Nesse período, o objetivo era desenvolver sistemas que conseguissem decifrar informações soviéticas rapidamente. A ferramenta utilizada nesse processo era uma calculadora científica, capaz de fazer uma tradução de palavra por palavra. Essa iniciativa originou-se com base na hipótese, ingênua e absolutamente equivocada, de que um texto escrito em russo não passava de um texto escrito em inglês elaborado com símbolos diferentes e desconhecidos. Contudo, logo ficou claro que considerar a tradução como um procedimento de decodificação de símbolos desconhecidos não era adequado; as línguas naturais não podem ser tomadas como um sistema artificial, posto que não demonstram uma relação biunívoca entre símbolo e referente (MARTINS; NUNES, 2005). Nesses anos iniciais, acrescenta-se, a

pesquisa em TA também encontrou restrições impostas pelo *hardware* precário, pela falta de linguagens de programação de alto nível e pelo pouco auxílio que os linguistas forneciam aos pesquisadores dessa área, já que nem sintaxe e muito menos a semântica, por exemplo, recebiam a devida importância (HUTCHINS, 2001).

Os objetivos tornaram-se menos modestos na década de 50, quando o objetivo principal passou a ser a mecanização completa da tradução. Em 1952, durante um congresso promovido pelo Instituto de Tecnologia de Massachusetts, novos caminhos foram discutidos e traçados para as pesquisas, o que fez surgir as importantes abordagens com base em conhecimento linguístico. Discutiu-se, nesse congresso, que as abordagens baseadas em conhecimento linguístico deveriam considerar, além de estudos sobre a frequência dos itens lexicais dos textos a serem traduzidos, as equivalências linguísticas entre as línguas em questão e outros aspectos de natureza técnica. Depois de considerados esses aspectos, partia-se, então, para a análise sintática e, por fim, para o efetivo desenvolvimento do sistema de TA (DIAS-DA-SILVA et al*.,* 2007).

De acordo com Dias-da-Silva et al. (2007), outros países europeus também começaram a investir em TA no final dos anos 50, e as pesquisas seguiram constantes até os anos 60, quando se percebeu que o progresso no desenvolvimento dos programas de TA era muito lento e que as ambiciosas previsões feitas pela comunidade de TA da época não haviam sido alcançadas. A principal causa para esse cenário foi a complexidade para tratar computacionalmente elementos formais e teóricos que estavam em construção pela Linguística.

Esses fatos motivaram a elaboração do relatório desanimador do ALPAC[3], em 1966, que fez com que os investimentos norte-americanos em pesquisas sobre TA fossem

---

[3] Em 1964, a pedido das agências de formento do governo dos EUA para a pesquisa em TA, foi montado o *Automatic Language Processing Advisory Committee* (ALPAC) para analisar o que já havia sido alcançado em TA e as perspectivas para o futuro. O relatório produzido por essa agência, influenciado pelos argumentos defendidos pelo filósofo da linguagem Yehoshua Bar-Hillel (pesquisador do Instituto de Tecnologia de Massachusstes), afirmava que a TA era muito mais cara e lenta que a tradução humana e não era precisa. Acrescentou-se ainda que não havia perspectiva da utilização da TA (ALPAC, 1966). Esse relatório é considerado a causa do término da maior parte das pesquisas da TA durante mais de uma década. As agências de fomento não viam motivos para investir em TA, recomendando que haveria mais utilidade investir no desenvolvimento de ferramentas automáticas para o uso de tradutores humanos como, por exemplo, os dicionários legíveis por máquina.

sumariamente cortados (ALFARO, 1998). Deve-se, porém, enfatizar, como observa Alfaro (1998), que esse aparente fracasso da TA foi importante para o desenvolvimento do PLN, da Linguística e também da Inteligência Artificial, porque muitos sistemas que são utilizados atualmente têm suas raízes em problemas observados nesses programas de TA iniciais.

Durante os anos 80, como comentam Dias-da-Silva et al. (2007), a TA foi colocada novamente no foco das investigações, motivadas, entre outros fatores, pela massa de informação que circulava na Internet, pelos avanços feitos nas áreas de Linguística Formal, Inteligência Artificial e Ciências da Computação, pela criação da União Europeia e pelos investimentos vindos das indústrias privadas. Além disso, destaca Alfaro (1998), destaca que os objetivos iniciais das pesquisas sobre TA se tornaram mais realistas e os pesquisadores começaram a perceber que a TA não precisava ser perfeita para ser útil.

Depois de considerar os novos objetivos da TA, agora menos ambiciosos, estratégias, como as de diminuir a complexidade do texto de entrada ou o grau de mecanização do sistema, foram propostas como alternativas para reelaborar a tarefa da TA. Para reduzir a complexidade do texto a ser traduzido, a alternativa foi fazer o sistema de TA operar sobre textos escritos de forma controlada, ou seja, textos elaborados com o uso de uma "sublíngua". A hipótese que está por trás do conceito de sublíngua é que o controle da forma e do conteúdo de um texto é uma ferramenta suficiente para reduzir ou eliminar a ambiguidade presente nas línguas naturais. Essa característica é observada no clássico sistema canadense Météo[4], que até hoje traduz boletins meteorológicos escritos em inglês para o francês de forma bem sucedida.

Martins e Nunes (2005) explicam que há duas formas para controlar uma língua natural de modo que ela seja entendida como uma sublíngua: por meio de regras de redação que o texto a ser traduzido deve obedecer (não provocar elipses e explicitar as relações anafóricas são exemplos dessas regras) e por meio da especificação de etiquetas ou marcações das informações metalinguísticas necessárias. Contudo, os autores observam que utilizar sublínguas é uma alternativa que acarreta o desenvolvimento de sistemas muito limitados, capazes de operar apenas em domínios restritos.

---

[4] O sistema Météo teve sua primeira versão desenvolvida em 1976 pelo grupo TAUM (*Traduction Automatique de l'Université de Montréal*) e, atualmente, ainda é usado. O sistema, que traduz boletins meteorológicos, é bilíngue e unidirecional, traduzindo apenas do inglês para o francês e mescla o método direto e o indireto por transferência (MARTINS; NUNES, 2005).

Embora não tão bem sucedida, a automatização da tradução se tornou uma realidade no final do século XX, à medida que sistemas de TA passaram a ser desenvolvidos e disponibilizados na Internet. Entretanto, esses sistemas não são, na verdade, capazes de traduzir textos de uma língua natural para outra por si mesmos e com adequação plena, posto que a tradução de qualidade só é obtida com a pós-edição de um tradutor humano, assim como se vê argumentado em Martins (2003):

> Nenhum dos sistemas disponíveis, sejam provenientes das iniciativas de mercado, sejam derivados da pesquisa acadêmica, produziram, até hoje, resultados que se pudessem considerar minimamente satisfatórios, tais que pudessem prescindir de edição humana. Pelo contrário: mesmo para línguas, como o inglês, que têm concentrado, pelo poder econômico de sua comunidade de falantes, uma soma considerável de investimentos, não foram observados resultados que permitissem, qual em um teste de Turing, confundir o desempenho do tradutor mecânico com a expectativa de desempenho de um tradutor humano (MARTINS, 2003, p. 1).

Wilks (2009) argumenta que, com os estudos da TA realizados até os dias de hoje, ainda é cedo para se afirmarem fatos de forma categórica e aponta inconsistências sobre o que já se conhece, porque, de um lado, a TA funciona, fato comprovado pela existência de sistemas que traduzem de forma completamente automática, trazendo benefícios para muitos usuários que necessitam recorrer a esse recurso (basta considerarem-se, por exemplo, os resultados de certa forma satisfatórios do *Google Tradutor*); de outro, é evidente a falta de avanços teóricos que possibilitam uma TA de alta qualidade.

Não se desconhece que a tarefa de traduzir é complexa, porque, entre outros fatores, é necessário, como já se observou, compreender o texto sob vários pontos de vista, além de compreender o seu conteúdo intrínseco e de estabelecer relações com outros conteúdos externos a ele. Além disso, ainda é preciso considerar a necessidade de conhecimento robusto das línguas envolvidas, não se descuidando de levar em conta as suas diferenças e semelhanças (SANTOS, 1988, p. 03).

Como consequência dessa complexidade, como observa Martins (2003), o objetivo de desenvolver programas que traduzissem de forma completamente automática, meta original dos desenvolvedores de sistemas de TA, acabou se tornando mais modesto.

Hutchins e Sommers (1992), ao traçarem um panorama histórico da TA, relatam que eles foram reformulados como sistemas de tradução automática auxiliada por humanos

(*human-aided machine translation*), ou seja, aqueles que consideram alguma forma de pré ou pós-edição feita por um tradutor humano, e como tradução humana auxiliada por máquinas (*machine-aided human translation*), quando os sistemas apresentam recursos como dicionários, enciclopédias, corretor ortográfico e gramatical, entre outros.

De acordo com Santos (1999), a tradução de textos técnicos e científicos foi, desde o início, o alvo dos estudos dos pesquisadores, porque, equivocadamente, acreditava-se que as diferenças culturais eram menos rigorosas nos documentos desse tipo e a demanda por tradução na área técnica e científica sempre foi alta, chegando a ultrapassar a capacidade da tradução humana profissional muito frequentemente.

Além disso, para a tradução de textos técnicos e científicos não era necessário nenhum resultado perfeito, o que interessava, de fato, era a rapidez do processo, mesmo que a tradução fornecida fosse "gramaticalmente imperfeita, lexicalmente estranha e estilisticamente grosseira" (HUTCHINS, 2001, p. 8, tradução nossa)[5].

Cumpre, por fim, esclarecer que a tradução técnica não é mais fácil e nem mais simples que a tradução literária. Os documentos técnicos que foram matéria prima da TA eram, na maioria das vezes, produzidos pelas próprias empresas que financiavam os estudos e, por isso, eram material de fácil obtenção e de baixo custo. "Foram, portanto, considerações econômicas, não científicas ou linguísticas, que levaram os textos técnicos a ser a matéria-prima inicial na tradução automática" (SANTOS, 1999, p. 02).

Atualmente, a vasta quantidade de informação disponível na Web e a popularização dos computadores são os fatores que mais impulsionam as pesquisas sobre TA, já que existe uma grande demanda por tradutores automáticos que apresentem, de modo instantâneo, a tradução de textos e páginas da Web (DIAS-DA-SILVA et al., 2007).

1.4 A TA e o PLN

Apesar das dificuldades, desde a década de 40, quando os computadores foram apresentados ao mundo, seu desenvolvimento tem sido constante e sua contribuição para com todos os domínios do conhecimento tem sido evidente; o potencial dessas máquinas para auxiliar o linguista, assim como outros estudiosos, tem sido também evidente:

---

[5] Texto original: *grammatically imperfect, lexically awkward and stylistically crude* (HUTCHINS, p. 8, 2001).

> O computador pode mudar o nível de análise das línguas naturais, assim como o microscópio mudou a biologia. Ele facilita a matematização como ajudou a física. E ele ligou a teoria, estudos empíricos e, talvez, a aplicação prática (ALPAC, 1966, p. 121, tradução nossa). [6]

Todo o desenvolvimento trazido pelos computadores proporcionou o nascimento de uma grande diversidade de desafios, enfatizando sempre o problema central de fazer com que a comunicação entre o usuário e a máquina se torne "mais amigável". Os desafios que surgiram em torno da questão do tratamento computacional das línguas naturais fizeram com que grandes investimentos materiais e humanos fossem aplicados nesse empreendimento, criando, assim, o campo de estudo do PLN (DIAS-DA-SILVA, 2006).

A grande maioria dos usuários dos computadores considera as linguagens de programação, necessárias para a comunicação com a máquina, muito complexas, fato que possibilita afirmar que a comunicação entre homem-máquina é um grande obstáculo para o uso mais difundido dos sistemas de computadores. O ideal seria, como argumenta Kurzweil (1990), que a comunicação com a máquina ocorresse igual à comunicação entre humanos: através de uma língua natural.

Inserida no contexto do PLN, o estudo da TA é multidisciplinar e investiga como desenvolver os sistemas de TA, que têm como requisitos a compreensão de aspectos das línguas naturais e da tradução de uma para a outra, implicando a construção de interfaces em língua natural que venham a auxiliar os usuários da língua e do computador em diferentes pontos do globo.

Como parte dos estudos do PLN, a TA exige também a investigação linguístico-computacional dos processos de produção e de compreensão das línguas naturais para que seja possível modelar a automatização da tradução de textos de uma língua para outra.

Há uma grande diversidade de objetivos dentro do PLN, por abordar questões linguísticas e também questões computacionais. Para que esses objetivos sejam alcançados, o trabalho que una esses dois grandes campos de conhecimento, representados, respectivamente, pela Linguística e pelas Ciências da Computação precisa se desenvolver.

---

[6] Texto original: […] *it can change the level of analysis of natural languages, as the microscope changed biology. It facilitates mathematization as it has aided physics. And it has linked theory, empirical studies, and, perhaps, practical application* (ALPAC, 1966, p. 121).

Mas, apesar da necessidade desse trabalho conjunto ser um fato reconhecido, ele ainda acontece muito timidamente. Além disso, embora haja certo reconhecimento de que a construção de conhecimentos linguísticos e metalinguísticos seja tarefa essencial e indispensável para que uma realização qualitativamente significativa possa ser alcançada no âmbito dos estudos do PLN, os fenômenos linguísticos, por razões diversas, não têm sido descritos com a necessária precisão pelos linguistas para uso dos informatas. Em particular, as pesquisas que se ocupam da TA frequentemente são alvos de críticas que as acusam de não considerar os conhecimentos descobertos e construídos pela Linguística. Isso tudo corrobora que, na prática, a comunicação entre os estudiosos da Linguística e do PLN ainda é deficiente, como já apontara Dias-da-Silva (1996).

Os dados que precisam ser analisados quando se pretende desenvolver um estudo sobre línguas naturais são muito numerosos e, por natureza, complexos. O uso de computadores para auxiliar nesse tipo de estudo significa uma ferramenta para ajudar no controle dos dados e para amenizar sua complexidade.[7] Entretanto, os métodos utilizados para fazer os computadores lidarem com dados linguísticos, tanto na época do relatório do ALPAC quanto nos dias atuais, ainda necessitam de maior desenvolvimento. Para que isso ocorra, é preciso incentivar e apoiar diversos tipos de pesquisas, como apontara o relatório do ALPAC (1966). Em especial, dois tipos de pesquisa merecem maior destaque:

- O desenvolvimento de métodos/ferramentas computacionais para analisar línguas naturais, como as ferramentas que o linguista pode utilizar em suas descobertas e prever generalizações e as ferramentas para verificar essas generalizações;
- O desenvolvimento de métodos/ferramentas que possibilitam a avaliação de teorias linguísticas pelo próprio linguista.

Com efeito, Allen (1995) aponta que os modelos linguísticos computacionalmente tratáveis são úteis a partir de duas perspectivas distintas: a científica e a prática. Considerando o aspecto científico, os computadores permitem que a comunicação linguística humana seja mais bem estudada para se atingir uma melhor compreensão do funcionamento das línguas naturais. Eles fornecem generalizações sobre o comportamento e sobre o processamento

---

[7] É interessante salientar que o computador, quando utilizado como uma ferramenta, tem a característica de possibilitar que novas formas de descrição e de sistematização das nossas próprias capacidades sejam descobertas e testadas, e isso não só no domínio da tradução, bem como em todas as áreas do conhecimento a que ele é aplicado (SANTOS, 1999, p. 04).

linguísticos que podem ser relevantes para o estudo de outras disciplinas como, por exemplo, a psicolinguística. Do ponto de vista prático, como grande parte do conhecimento humano é armazenada sob forma linguística, poder modelar as habilidades linguísticas modifica a forma de utilização dos computadores. Essa revolução permite que os computadores tenham acesso a toda essa informação que os humanos são capazes de armazenar, à medida que conseguirem compreender as línguas naturais (ALLEN, 1995, p. 02).

Santos (1999) observa a multifuncionalidade do computador e mostra que ele pode desempenhar quatro funções diferentes e não excludentes no âmbito da tradução: **facilitador da tradução, auxiliar da tradução, auxiliar do estudo e do ensino da tradução** e **tradutor automático.**

Quando o computador funciona como um facilitador do processo de tradução, ele possibilita que o texto seja editado, que traduções feitas anteriormente sejam armazenadas e consultadas e que diversas fontes de conhecimento sobre as línguas-fonte e alvo sejam acessadas. Como auxiliar no processo da tradução, o computador pode sugerir traduções precisas e estilisticamente apropriadas ao gênero (pensa-se aqui nos textos mais técnicos), por permitir o acesso a vastos bancos terminológicos e a tipos diversos de uso da língua. Como auxiliar no estudo e no ensino da tradução, o computador pode reconhecer regularidades e dificuldades e, dessa forma, contribuir para a construção de bases de dados para consulta posterior. A facilidade de reconhecer problemas também é útil para que o ensino seja aperfeiçoado nos aspectos problemáticos, que podem ser explorados como exercícios durante o processo de aprendizagem. Finalmente, como tradutor automático, o computador gera traduções que podem ajudar o usuário a ter pelo menos uma noção do que um texto em uma língua desconhecida expressa.

De acordo com Hutchins (2001), Bar-Hillel, após auxiliar o ALPAC na análise da TA, escreveu um artigo mostrando as abordagens básicas para a TA que eram utilizadas no período quando nomeado pelo instituto. Nessa época, já era conhecido o fato de que a supervisão do tradutor humano seria necessária para a pré ou pós-edição dos textos, porque uma TA completamente não-supervisionada e de alta qualidade seria impossível. Bar-Hillel criticava o fato de o objetivo das pesquisas sobre TA ser a criação de sistemas completamente automatizados que produzissem traduções iguais às produzidas por tradutores humanos. E, conforme reconhece Nirenburg (1996), retomado avaliações de Bar-Hillel:

> Agora parece relativamente certo… que, com todo o progresso feito no *hardware*, nas técnicas de programação e de compreensão linguística, a qualidade da tradução

> automática completamente não supervisionada, mesmo quando restrita a material científico ou tecnológico, nunca se aproximará daquela feita por tradutores humanos qualificados, e, portanto, só será capaz de competir com a tradução humana sob circunstâncias muito excepcionais. […] Tradutores humanos especialistas usam o conhecimento de experiências, geralmente de modo inconsciente, para resolver ambiguidades sintáticas e semânticas que as máquinas ou terão que deixar sem resolução ou terão que resolver através de alguma regra mecânica que, muito frequentemente, resultará em uma tradução incorreta (apud NIRENBURG, 1996, p. 302, tradução nossa).[8]

Efatiza-se, por fim, com Martins e Nunes (2005, p. 22), que a TA é uma encruzilhada: "nenhuma outra aplicação talvez represente de forma tão cabal os desafios da interação entre modelos matemáticos, linguísticos e cognitivos" (MARTINS; NUNES, 2005, p. 22).

1.5 Os dois paradigmas de estudo do PLN

A primeira distinção a ser feita a respeito dos paradigmas de sistemas de PLN pode ser traçada entre os paradigmas racionalista e empiricista. Mas, antes de tudo, é interessante fazer algumas observações sobre a origem desses dois paradigmas.

De acordo com Manning e Shütze (1999), os estudos linguísticos, em linhas gerais, buscam a explicitação e a caracterização de fatos linguísticos das línguas naturais para a compreensão das estruturas linguísticas por meio das quais os falantes se comunicam. Para isso, uma maneira descritiva é estipular regras de organizção da estrutura das expressões linguísticas de uma língua natural. A estratégia é descrever quais são os enunciados bem-formados em oposição àqueles considerados mal-formados. Entretanto, questiona-se se é possível caracterizar, de forma completa e precisa, os enunciados bem-formados. Por outro

---

[8] Texto original: *It seems now quite certain ... that with all the progress made in hardware, programming techniques and linguistic insight, the quality of fully autonomous mechanical translation, even when restricted to scientific or technological material, will never approach that of qualified human translators and that therefore MT will only under very exceptional circumstances be able to compete with human translation. […] Expert human translators use their background knowledge, mostly subconsciously, in order to resolve syntactical and semantical ambiguities which machines will have either to leave unresolved or resolve by some "mechanical" rule which will every so often result in a wrong translation*." (apud NIRENBURG, 1996, p. 302)

lado, enfatiza-se que não se pode ser radical e não reconhecer que grande parte dos padrões sintáticos de uma língua natural pode ser expressa por meio de regras.

Outro caminho para descrever os fatos linguísticos é investigar quais são os padrões recorrentes na língua, tarefa que pode ser realizada por meio de contagens. Pretende-se demonstrar, com essa afirmação, que os modelos estatísticos da língua podem ser usados por muitas tarefas dentro do PLN (MANNING; SHÜTZE, 1999).

Como informam Manning e Shütze (1999), durante o período entre 1960 e 1985, a Inteligência Artificial e o PLN, entre outras áreas de estudo, eram dominados por uma abordagem racionalista. Tal abordagem caracteriza-se pela crença em que parte do conhecimento existente na mente humana é fixada previamente, talvez por genética. Essa abordagem racionalista baseia-se na aceitação dos argumentos propostos por Chomsky sobre a existência de uma faculdade linguística inata. Já no campo da Inteligência Artificial, as abordagens racionalistas são a base da tentativa de criação de sistemas inteligentes por meio da codificação manual de conhecimentos e mecanismos de raciocínio característicos da mente humana, com o objetivo de simular o material com que o cérebro trabalha.

Por sua vez, as abordagens empiricistas consideram que já existem, no cérebro humano, determinadas habilidades cognitivas. Isso significa dizer que entre as duas abordagens existe apenas uma diferença gradativa. Em ambas as abordagens, assume-se que existe alguma estrutura inicial no cérebro que prefere determinadas organizações. A diferença é que as abordagens empiricistas consideram que o trabalho inicial da mente começa com operações genéricas que permitem associações, reconhecimento de padrões e generalizações, que serão aplicadas às diversas informações sensoriais presentes ao redor do ser humano, possibilitando o aprendizado de uma língua natural (MANNING; SHÜTZE, 1999).

De acordo com Manning e Shütze (1999), a abordagem racionalista baseia-se em julgamentos categóricos sobre as frases, separando-as em dois grupos: frases gramaticais e agramaticais. Por outro lado, a abordagem estatística pretende descrever adequadamente as associações e preferências que ocorrem quando a língua é considerada em sua totalidade do uso real.

Dessas duas perspectivas, depreendem-se os dois tipos de sistemas de TA: os sistemas de TA baseados em regras, ou seja, com base em conhecimento e manipulação de símbolos, daí a denominação “sistemas simbólicos”, que derivam das abordagens racionalistas; e os sistemas de TA baseados em casos/exemplos, ou seja, com base em generalizações de

natureza estatística, também conhecidos como “sistemas conexionistas”, que derivam das abordagens empiricistas (MARTINS, NUNES, 2005).

De acordo com Martins e Nunes (2005), a tradução automática baseada em conhecimento (*Knowledge-Based Machine Translation* ou KBMT) utiliza uma enciclopédia, ou base de conhecimento, além de recursos linguísticos como dicionários e gramáticas. Para a tradução automática baseada em exemplos (*Example-Based Machine Translation* ou EBMT), recorre-se à estatística para realizar traduções com base em outras já existentes e revisadas. A ideia central dessa abordagem é a reutilização de exemplos de tradução pré-existentes para gerar novas traduções. Um *corpus* (ou base de dados de textos já traduzidos) é utilizado para relacionar os novos dados de entrada com dados já existentes, extraindo informações de exemplos adequados que são então recombinados. Esse tipo de abordagem estatística, em geral, não faz uso de conhecimento linguístico, já que a operação de tradução é feita em termos de cálculos estatísticos e matemáticos que estimam parâmetros para modelar as línguas do sistema (SOMERS, 2003).

Wilks (2009) aborda a questão da necessidade de se modelar o conhecimento linguístico para a TA e critica que as abordagens estatísticas não o consideram essencial para esse tipo de tarefa. Porém, depois de muitas experiências, os defensores dessas abordagens são obrigados a reconhecer a necessidade de partir do conhecimento linguístico no processo de desenvolvimento de sistemas de TA, mesmo considerando a possibilidade de aquisição estatística desse conhecimento ou de parte dele.

Martins e Nunes (2005) apontam que os dois paradigmas ainda apresentam problemas. Em primeiro lugar, a base de trabalho dos sistemas de EBMT é a similaridade entre frases previamente traduzidas e aquela submetida à tradução. Essa similaridade, porém, desperta opiniões controversas. Isso acontece porque nem sempre a similaridade estatística entre as frases garante a similaridade de significado (como será apreciado na seção 4.3), já que não é possível afirmar com absoluta certeza que uma mesma estrutura sintática expresse o mesmo significado quando utilizada em variados contextos. Além disso, o desempenho desses sistemas está relacionado ao tamanho do conjunto das traduções previamente traduzidas. Por isso, os sistemas de EBMT têm que lidar com uma lógica contraditória, já que se o *corpus* de traduções prévias é pequeno, os resultados são ruins, e se o *corpus* for muito amplo, os resultados não são precisos.

Já para os sistemas do tipo KBMT, o problema reside na construção, sempre laboriosa e a longo prazo, dos recursos a serem utilizados, ou seja, enciclopédias, ontologias, léxicos,

etc., que devem representar o conhecimento de mundo humano. Além disso, como afirmam Martins e Nunes (2005, p. 14), o conhecimento humano "nem sempre é discreto, preciso e, principalmente, nunca é estático".

1.6 Os tipos de sistemas de TA

Segundo Dias-da-Silva et al. (2007), é preciso considerar dois tipos de informação para classificar um sistema de PLN e, por conseguinte, um sistema de TA: o seu paradigma e o seu método. Segundo os autores, se o paradigma diz respeito ao modo de manipulação e organização dos dados do sistema, o método diz respeito ao modo de organização dos diferentes módulos do sistema. Além disso, o grau de sofisticação dos sistemas é medido de acordo com um ou mais dos três tipos de métodos empregados na operação de tradução. A partir disso, os métodos podem ser agrupados em duas categorias, a tradução **direta** ou a tradução **indireta** e os sistemas podem ser classificados como **diretos**, **de transferência** e **de interlíngua,** variando de acordo com o método que utilizam. (HUTCHINS, 2003; DIAS-DA-SILVA et al., 2007).

Os sistemas que utilizam o método direto traduzem diretamente a partir da língua-fonte; já os sistemas que utilizam o método indireto utilizam algum tipo de representação abstrata intermediária das estruturas linguísticas da língua-fonte, para só depois gerarem o texto na língua-alvo. São incluídos, neste último tipo, os sistemas de transferência e os de interlíngua (SANTOS, 1998; MITKOV, 2003; MARTINS, NUNES, 2005).

O método direto é utilizado por sistemas que são bilíngues e unidirecionais, são os sistemas mais simples e foram típicos durante a primeira geração de sistemas de TA (1950-1970), podendo ser chamados de sistemas de primeira geração. No método direto, o texto fonte é analisado minimamente para poder gerar textos na outra língua, sem recorrer a nenhuma representação intermediária; as análises sintática e semântica são realizadas apenas quando são absolutamente indispensáveis.

Esses sistemas realizam a tradução procurando as equivalências diretas entre os itens lexicais das línguas fonte e alvo e pressupõem que o léxico e a sintaxe da língua-fonte devem ser analisados apenas para resolver alguma possível ambiguidade ou para identificar expressões e para determinar a ordem dos itens lexicais na língua-alvo. Um sistema de

tradução direta é equipado apenas com um grande dicionário e, além disso, um único programa analisa e gera os textos (MITKOV, 2003; SANTOS, 1998).

Diversos motivos contribuem para a inadequação do método direto. Além da possibilidade de um item lexical de uma língua corresponder a mais de um item lexical da outra língua, há também as expressões idiomáticas, que não têm sentido composicional, o que impossibilita a tradução a partir de cada um dos seus elementos. Martins e Nunes (2005) afirmam que os provérbios são um exemplo extremo dessa situação, porque não se pode inferir o sentido de um provérbio a partir da soma dos significados de cada item lexical que o compõe.

Os sistemas de tradução automática que utilizam o método direto baseiam-se na falsa tese que as línguas naturais possuem estruturas sintáticas isomorfas. Contudo, quando se compara o português e o espanhol, línguas consideradas historicamente relacionadas, percebe-se qua há uma distância entre as estruturas sintáticas de cada língua, porque elas se desenvolveram separadamente (MARTINS, NUNES, 2005).

Já nos sistemas de TA indiretos, como explicam Dias-da-Silva et al. (2007), o processo de análise da língua-fonte e de geração da síntese da língua-alvo são processos independentes e só tratam dos problemas da língua com que estão lidando. A tradução indireta se baseia na hipótese de que é preciso considerar o conhecimento linguístico das línguas a serem tratadas e também as diferenças entre elas. Essas informações são expressas por meio de uma representação intermediária.

Dependendo do método que utiliza a TA indireta, a representação intermediária varia: na TA indireta por transferência, a representação intermediária depende das línguas que estão sendo tratadas, sendo preciso recorrer a duas linguagens de representação, uma para a língua-fonte e outra para a língua-alvo. Já na TA indireta por interlíngua a representação intermediária é única e independente das línguas envolvidas (DIAS-DA-SILVA *et al.,* 2007).

Nos sistemas de transferência, a tradução processa-se por meio de regras sintáticas criadas a partir da análise da estrutura sintática da frase da língua-fonte e que produzem a representação sintática da frase da língua-alvo. Esse processamento se dá em três estágios. Durante o primeiro estágio, o texto-fonte é transformado em representações abstratas intermediárias orientadas para a língua-fonte, eliminando-se a ambiguidade; durante o segundo estágio, essas representações são transformadas em representações equivalentes para a língua-alvo; finalmente, no último estágio do processo, o texto na língua-alvo é gerado (DIAS-DA-SILVA *et al.*, 2007).

É comum que os sistemas de TA por transferência realizem apenas uma análise sintática, gerando como representação intermediária uma estrutura arbórea para a língua-fonte que será transformada em uma estrutura de arbórea para a língua-alvo. Entretanto, segundo Dias-da-Silva et al. (2007), para alcançar um resultado de alta qualidade outras análises são necessárias. É por isso que há sistemas que incorporam informações provenientes do nível semântico e que podem ser representadas, como demonstra Dias-da-Silva et al. (2007), por meio de *frames*, ou seja, uma coleção de atributos e valores. Contudo, os autores apontam que esse tipo de análise em sistemas de TA por transferência não é muito frequente e que se recorre a ela para solucionar alguns problemas limitados. De acordo com Mitkov (2003), esses sistemas são, geralmente, equipados com três tipos de dicionários: (a) dicionários da língua-fonte, que trazem as informações morfológica, gramatical e semântica; (b) dicionários semelhantes a esses para a língua-alvo; (c) um dicionário bilíngue, que relaciona as formas básicas das duas línguas tratadas pelo sistema.

As gramáticas utilizadas pelos sistemas de transferência também são distintas para analisar a língua-fonte, para fazer síntese da língua-alvo e para transformar as estruturas da língua-fonte em estruturas equivalentes na língua-alvo. Jurafsky e Martin (2000) apontam que uma das desvantagens do método de transferência é o fato de ser necessária a especificação de regras diferentes para cada par de línguas implantado no sistema, o que, obviamente, dificulta seu uso em ambientes multilíngues. Para citar um exemplo de sistema que utiliza o método indireto por transferência, têm-se o *Eurotra,* que tem como objetivo o desenvolvimento de um ambiente multilíngue que inclua as línguas da Comunidade Europeia (DIAS-DA-SILVA et al*.,* 2007).

Os sistemas de interlíngua, na qual a tradução é vista como um processo de extração do significado do texto da língua-fonte que deve ser expresso no texto da língua-alvo, são uma alternativa para as desvantagens encontradas nos sistemas de TA que utilizam o método indireto por transferência. Esses sistemas são os mais sofisticados e funcionam de maneira diferente. Neles, a tradução é feita em dois estágios: do texto da língua-fonte para uma interlíngua, uma representação abstrata "universal" do significado (pois deve aplicar-se a qualquer língua), e da interlíngua para a língua-alvo (DIAS-DA-SILVA, 2006; HUTCHINS, 1992). A tradução é, portanto, feita com base na possibilidade de transformação de textos em representações conceituais que podem ser codificadas em qualquer língua. O método de interlíngua exige a análise semântica, que não é necessária no método de transferência (este exige apenas a análise sintática), trazendo um trabalho analítico significativamente mais

complexo para a TA. Por outro lado, a geração do texto da língua-alvo pode ser feita diretamente da representação da interlíngua (que é a representação conceitual do conteúdo a ser traduzido), com o uso de regras sintáticas da língua-alvo.

Esse tipo de método traz uma grande economia para construção sistemas de TA, porque o mesmo programa de análise de uma determinada língua-fonte pode servir para mais de uma língua-alvo (JURAFSKY; MARTIN, 2000; MITKOV, 2003).

Vale destacar que os sistemas de transferência e de interlíngua são incluídos no período da segunda geração dos sistemas de TA, que ocorreu durante a segunda metade dos anos 70 até os anos 80. Segundo Mitkov (2003), esses sistemas se baseiam na aplicação de regras – e por isso são conhecidos como **sistemas baseados em regras**, em oposição aos sistemas baseados em estatística, que são **sistemas baseados em *corpus*.**

Santos (1988) menciona uma divisão que atualmente não é mais tão significativa. Trata-se da distinção entre sistemas predominantemente sintáticos e sistemas predominantemente semânticos. Como é necessário primeiro compreender o texto para depois traduzi-lo (fato amplamente reconhecido), essa distinção acaba por se deslocar "para uma opção metodológica de compromisso entre eficiência e qualidade" (SANTOS, 1988, p. 08). Santos propõe, entretanto, que, nesse aspecto, é mais estratégico classificar os sistemas quanto ao grau de conhecimento morfossintático, semântico ou pragmático exigido por eles. Hutchins (1986), porém, defende que a distinção inicial deve ser

> expressa em termos de quem controla quem, ou seja, um **sistema de índole** sintática manipularia marcadores semânticos ajudando a identificação das estruturas, mas possuiria, como unidade básica, por exemplo, a frase (conceito eminentemente sintático). Por outro lado, um **sistema de índole semântica** poderia executar simultaneamente com o varrimento sintático do texto uma análise semântica, ou ter como única representação interna, a partir do texto, uma representação de casos ou na forma da dependência conceitual. (apud SANTOS, 1988, p. 08). (grifos nossos)

É importante apontar que Martins e Nunes (2005) argumentam que o caminho mais interessante para a construção de sistemas de TA talvez seja a união de diferentes métodos e paradigmas. Os autores afirmam, por exemplo, que o sistema Météo, acima mencionado, utiliza os métodos direto e indireto.

Essa defesa do hibridismo confirma que a classificação dos sistemas em métodos e paradigmas com limites nitidamente traçados serve mais para fins didáticos do que para o desenvolvimento real de um sistema de TA.

Para concluir a exposição da tipologia de sistemas de TA, no que diz respeito ao número de línguas, caracterizam-se os sistemas em bilíngues, quando operam sobre um par de línguas, ou em multilíngues, quando operam sobre mais de duas línguas; no que diz respeito à direção da tradução, caracterizam-se em sistemas unidirecionais, quando operam apenas em uma direção (do inglês para o português, por exemplo), ou em sistemas bidirecionais, quando operam nas duas direções. (HUTCHINS, 1992).

1.7 A ambiguidade linguística e a TA

Analisar e interpretar os diversos enunciados de uma língua natural é, como enfatizam Martins e Nunes (2005), a principal dificuldade que os sistemas de TA precisam enfrentar, e isso ocorre porque a TA depende do conteúdo semântico das frases que integram um texto; em outras palavras, o significado do enunciado depende da interpretação que o leitor irá atribuir a ele, o que significa dizer que o significado não pode ser extraído apenas das informações linguísticas aparentes no enunciado, e essa característica pode ser considerada uma das dificuldades centrais com que a TA deve lidar.

Como aponta a literatura da área do PLN e TA, a grande maioria dos itens lexicais mais comuns na língua apresenta mais de um sentido possível e, como consequência dessa característica inerente às línguas naturais, os pesquisadores consideram o tratamento das ambiguidades linguísticas como ponto central nas discussões sobre o aperfeiçoamento da TA (DALE, 2000; IDE & VÉRONIS, 1998; NERENBURG, 1996; SOMERS, 2000).

Dale (2000) argumenta que os conceitos são codificados de modos diversos em cada língua, gerando, por exemplo, ambiguidades no nível lexical, como se observa em relação ao item lexical *leg* do inglês, que pode ser, em francês, traduzido por três unidades lexicais distintas, variando de acordo com quem tem a perna: *jambe*, *patte* e *pied.*

A ambiguidade também pode se manifestar no nível estrutural de análise linguística, assim como fica ilustrado pela frase *Visiting relatives can be boring* (DALE, 2000, p. 333), em que *visiting* pode ser o particípio presente uma forma verbal ou um adjetivo modificador do substantivo *relatives.* Outro exemplo de ambiguidade no nível estrutural está na frase *John*

mentioned the book I sent to Suzan porque a preposição *to* pode indicar diferentes papeis temáticos, nesse caso o Beneficiário da ação ou o Destinatário da mensagem.

A manifestação da ambiguidade é observada em todos os níveis de análise linguística, mas assim como afirmam Martins e Nunes (2005):

> a ambiguidade mais insidiosa é justamente aquela que não se revela na superfície do texto, mas se reveste de suposta univocidade, a nos convencer de que a sentença admite uma única e exclusiva possibilidade de interpretação. Essa monovalência esconde o fato de que a univocidade dos enunciados linguísticos deriva muito mais de fatores contextuais (extralinguísticos, portanto, e em princípio irrepresentáveis para a máquina) do que de fatores propriamente linguísticos (MARTINS; NUNES, 2005).

Com esses breves exemplos apresentados, já se pode vislumbrar como as ambiguidades linguísticas ocorrem no inglês, bem como o desafio que elas representam para o tratamento computacional das línguas naturais, já que compreender a leitura (interpretação) adequada, dentre as possíveis, requer que diferentes tipos de conhecimento linguístico e contextual sejam modelados.

2. Fatorando a investigação nos domínios linguístico e linguístico-computacional

Para o planejamento do estudo dos diferentes tipos de ambiguidade linguística do inglês e avaliação do seu impacto na adequação/qualidade da TA do inglês para o português, adotou-se a metodologia de pesquisa em PLN que fatora o estudo em três domínios que se complementam, sendo eles:

(i) o linguístico, em que se sistematizam os conhecimentos linguísticos necessários para descrever um determinado fenômeno da língua e que serão eleborados no sistema de PLN alvo;

(ii) o linguístico-computacional, em que se formalizam esses conhecimentos;

(iii) o computacional[9], em que o conhecimento formalizado é codificado em programas de computador (DIAS-DA-SILVA, 1996; 2006).

2.1 Investigação no domínio linguístico

Neste domínio, introduz-se, primeiro, a problemática geral do fenômeno das ambiguidades linguísticas, com foco no inglês, que se manifestam na TA de frases e parágrafos de textos em inglês para o português. Em seguida, procede-se à sistematização (catalogação, descrição e exemplificação) dos diferentes tipos de ambiguidade linguística levantados na literatura; esta composta de textos teóricos sobre o tema e provenientes da Linguística Teórica, do PLN e da TA. O quadro 6 sintetiza essa sistematização.

2.1.1 Apresentando o *Google Tradutor*

[9] A atividade do domínio computacional, que seria a implementação computacional das estratégias de detecção de ambiguidades e desambiguação não será realizada neste estudo, porque é própria do programador de computadores.

O *Google Tradutor* foi a ferramenta de TA escolhida para este projeto, disponibilizada na Web no site do *Google*[10], "calcula" uma tradução automática entre cinquenta e sete idiomas. Justifica-se a escolha dessa ferramenta por ela ser de acesso fácil, livre e gratuito e de ser mundialmente utilizada por milhões de usuários.

Este tradutor automático utiliza uma abordagem estatística, gerando traduções com base em reconhecimento de padrões que são encontrados em textos já traduzidos por humanos e armazenados em um grande banco de dados[11]. Isso proporciona ao sistema uma atualização dinâmica: à medida que a base de dados cresce, a probabilidade de uma tradução ser aperfeiçoada também cresce.

Assim como pode ser visto na matéria publicada na *Revista Veja*, em maio de 2010, o banco de dados do *Google* começou a ser construído em 2006 a partir de textos confiáveis disponíveis na internet. Os primeiros foram textos oficiais da ONU traduzidos para seis línguas. Em seguida, textos bilíngues de arquivos públicos também foram acrescentados e, atualmente, a *web* tem sido usada de forma mais abrangente para ampliar o banco de dados (PAVÃO JUNIOR, 2010).

Em todas as suas aplicações, o *Google* realiza trocas com o usuário. Na tradução automática, ao adicionar um texto ou uma página da Web para a tradução, o usuário recebe o resultado de modo imediato e também ajuda a ampliar o banco de dados porque os textos são incluídos nele. Com um banco de dados maior, também é maior a probabilidade de o tradutor automático "calcular" traduções futuras melhores.

Além disso, como informa Pavão Junior (2010), os computadores do *Google* analisam os textos de cada língua separadamente, com o objetivo de reconhecer as construções mais frequentes e, no final do processo, escolhe aquela que tem a probabilidade mais alta de ocorrer. Durante a análise, os textos das duas línguas são segmentados em frases. Os itens lexicais componentes da frase da língua-fonte são alinhados com aqueles presentes na frase da língua-alvo. Por fim, considerando os textos que formam o banco de dados do *Google*, o sistema calcula a probabilidade de ocorrência dessas combinações.

Já é possível perceber, a partir das discussões de Manning e Shütze (1999), a existência de um limite para essa abordagem estatística. Após descrevem o paradigma estatístico do PLN, os próprios autores reconhecem que para **começar** um trabalho estatístico

---

[10] Disponível em: www.translate.google.com.br. Acesso em: 04 maio 2011.

[11] Informação disponível em <http://translate.google.com.br/about/intl/pt-BR_ALL/>. Acesso em: 20 out. 2010.

é preciso lançar mão de determinados recursos linguísticos, alguns deles elencados no trecho a seguir:

> "Como se deve realmente proceder? Bem, antes de mais nada, é preciso lançar mão de **recursos lexicais**: textos legíveis por máquina, dicionários, *thesauri*, e também ferramentas para processá-los" (MANNING; SHÜTZE, 1999, p. 19, tradução nossa[12]).

Esse aspecto é reforçado pelos argumentos de Miles Osborne, observados na reportagem de Pavão Junior (2010), quando aponta que, no estágio atual, as traduções produzidas por sistemas de TA sempre apresentam erros gramaticais e, por isso, estuda-se adicionar regras gramaticais de diferentes línguas no programa, bem como a ampliação do número de textos que fazem parte do banco de dados (o que consequentemente ampliará o léxico do sistema) para que ele possa trabalhar com uma ferramenta a mais, além dos algoritmos.

Fica claro, portanto, que a construção de conhecimentos, nesse caso linguísticos, não pode ser desconsiderada, posto que esses conhecimentos são essenciais para o PLN.

2.1.1.1 O comportamento idiossincrático do Google Tradutor

O *Google Tradutor* apresenta comportamento idiossincrático, traduzindo, por exemplo, a frase *I look up* por *Olhei ao redor*. Essa constatação exigiu que se fizessem testes para verificar esse comportamento e, sobretudo, o grau dessas idiossincrasias, posto que não se tem elementos explícitos e nem documentação para se avaliar como as traduções são elaboradas. Não se sabe como o *Google Tradutor* executa a tradução, além da informação de que ele deve possuir algum conhecimento linguístico e aplicar sofisticados cálculos estatísticos sobre uma considerável massa de textos traduzidos e dispostos em forma de um *corpus* paralelo, português e inglês, conforme as considerações já feitas. Para a realização dos

---

[12] Texto original: "*How does one actually proceed? Well, first of all, one needs to get one's hand on some* ***lexical resources****: machine-readable text, dictionaries, thesauri, and also tools for processing them"* (MANNING; SHÜTZE, 1999, p. 19).

testes, foi submetido ao *Google Tradutor*, um conjunto específico de *phrasal verbs* que apresentam comportamento sintático-semântico específico.

Foram selecionados os verbos *look up* e *look for*[13]. A escolha do verbo *look up* foi determinada por ele ocorrer em um dos exemplos que ilustram um dos tipos de ambiguidade discutidos neste estudo, a saber, o tipo **B.3.1 Detecção de partículas**. Já a escolha do verbo *look for* decorreu do fato de o *Google Tradutor* apresentar resultados semelhantes aos apresentados quando a ele se submete o verbo *look up*. Realizaram-se, então, os testes descritos, a seguir, com o verbo *look up* e, em seguida, com o objetivo de analisar os resultados comparativamente, os mesmos testes foram realizados com o verbo *look* for.

Primeiramente, construiu-se, para cada verbo, um conjunto de construções (sintagmas e frases), formando um universo de 27 situações de teste, descritas no quadro 1:

Quadro 1 – As 27 situações de teste

	Situações de teste
1	verbo
2	verbo-*s*[14]
3	verbo-*ed*
4	*I* + verbo
5	*you* + verbo
6	*he* + verbo-*s*
7	*I* + verbo + objeto
8	*you* + verbo + objeto
9	*he* + verbo-*s*+ objeto
10	*I* + verbo-*ed*

[13] Lista de acepções dos verbos *look up* e *look for*, de acordo com *Longman Dictionary of Contemporary English* (2004) *Look up*: 1 *if a situation is looking up, it is improving = improve, get better.* 2 ***look sth up*** *if you look up information in a book, on a computer, etc, you try to find it there.* 3 ***look sb up*** *to visit someone you know, especially when you are in the place where they live for a different reason.*
Look for: 1 *to ry to find something that you have lost, or someone who is not where they should be = to search for.* 2 ***be looking for sb/sth*** *to be trying to find a particular kind of thing or person.* 3 ***be looking for trouble*** *informal to be behaving in a way that makes it likely that problems or violence will happen.*

[14] Legenda: a marca –*s* representa o morfema verbal terceira pessoa singular do indicativo afirmativo e a marca –*ed* representa o morfema verbal de passado simples ou morfema do particípio passado. É importante ressaltar também que as situações de testes foram desdobramentos dos padrões 0 + V, S+V e S+V+O.

11	*you* + verbo-*ed*
12	*I* + verbo-*ed* + objeto
13	*you* + verbo-*ed*+ objeto
14	*I* + *will* + verbo
15	*you* + *will* + verbo
16	*I* + *will* + verbo + objeto
17	*you* + *will* + verbo + objeto
18	*I* + *have* + verbo-*ed*
19	*you* + *have* + verbo-*ed*
20	*he* + *has* + verbo-*ed*
21	*I* + *have* + verbo-*ed* + objeto
22	*you* + *have* + verbo-*ed* + objeto
23	*he* + *has* + verbo-*ed* + objeto
24	*I* + *had* + verbo-*ed*
25	*you* + *had* + verbo-*ed*
26	*I* + *had* + verbo-*ed* + objeto
27	*you* + *had* + verbo-*ed* + objeto

Fonte: Elaboração própria

Todas as construções do quadro 1 foram testadas com e sem o ponto final, pois se observou que o *Google Tradutor* fornece resultados distintos para cada situação. Esse comportamento inesperado indica que o ponto final fornece uma informação de alta relevância para o comportamento do sistema: o momento de concluir a análise da construção.

Nos quadros 2 e 3 estão tabulados os resultados dos testes. A primeira coluna indica as situações de teste; a segunda coluna contém as construções elaboradas para os testes, com e sem o ponto final; a terceira coluna contém as traduções geradas pelo *Google Tradutor*, no ambiente *Google Translator Toolkit;* a última coluna indica se a tradução gerada automaticamente é ou não a esperada com base nas traduções feitas anteriormente pelo tradutor automático com os mesmos verbos.

Quadro 2 – Tabulação dos resultados do teste com o verbo *look up*

Situação de teste		Construções		*Google Translator Toolkit*		Tradução esperada?	
		+ ponto final	- ponto final	+ ponto final	- ponto final		
1	verbo	*Look up.*	*Look up*	Olhe para cima.	Olhe para cima		
2	verbo-*s*	*Looks up.*	*Looks up*	Olha para cima.	Procura	Sim	Não
3	*ve*rbo-*ed*	*Looked up.*	*Looked up*	Olhou para cima.	Procurou-se	Sim	Não
4	*I* + verbo	*I look up.*	*I look up*	Olhei ao redor.	Olhei ao redor	Não	Não
5	*you* + verbo	*You look up.*	*You look up*	Você olha para cima.	Você olha para cima	Sim	Sim
6	*he* + verbo-*s*	*She looks up.*	*She looks up*	Ela olha para cima.	Ela olha para cima	Sim	Sim
7	*I* + verbo + objeto	*I look up the number.*	*I look up the number*	Eu olho para cima o número.	Eu olho para cima o número	Sim	Sim
8	*you* + verbo + objeto	*You look up the number.*	*You look up the number*	Você olha para cima o número.	Você olha para cima o número	Sim	Sim
9	*he* + verbo-*s*+ objeto	*She looks up the number.*	*She looks up the number*	Ela olha para cima o número.	Ela olha para cima o número	Sim	Sim
10	*I* + verbo-*ed*	*I looked up.*	*I looked up*	Olhei ao redor.	Olhei ao redor	Não	Não
11	*you* + verbo-*ed*	*You looked up.*	*You looked up*	Você olhou para cima.	Você olhou para cima	Sim	Sim
12	*I* + verbo-*ed* + objeto	*I looked up the number.*	*I looked up the number*	Olhei para cima o número.	Eu olhei para cima o número	Sim	Sim
13	*you* + verbo-*ed*+ objeto	*You looked up the number.*	*You looked up the number*	Você olhou para o número.	Você olhou para o número	Não	Não
14	*I* + *will* + verbo	*I will look up.*	*I will look up*	Vou olhar para cima.	Vou olhar para cima	Sim	Sim

15	*you* + *will* + verbo	*You will look up.*	*You will look up*	Você vai olhar para cima.	Você vai olhar para cima	Sim	Sim
16	*I* + *will* + verbo + objeto	*I will look up the number.*	*I will look up the number*	Eu vou procurar o número.	Eu vou procurar o número	Não	Não
17	*you* + *will* + verbo + objeto	*You will look up the number.*	*You will look up the number*	Você vai procurar o número.	Você vai procurar o número	Sim	Sim
18	*I* + *have* + verbo-*ed*	*I have looked up.*	*I have looked up*	Eu olhei para cima.	Eu olhei para cima	Sim	Sim
19	*you* + *have* + verbo-*ed*	*You have looked up.*	*You have looked up*	Você olhou para cima.	Você olhou para cima	Sim	Sim
20	*he* + *has* + verbo-*ed*	*She has looked up.*	*She has looked up.*	Ela olhou para cima.	Ela olhou para cima	Sim	Sim
21	*I* + *have* + verbo-*ed* + objeto	*I have looked up the number.*	*I have looked up the number*	Eu olhei para cima o número.	Eu olhei para cima o número	Sim	Sim
22	*you* + *have* + verbo-*ed* + objeto	*You have looked up the number.*	*You have looked up the number*	Você olhou para o número.	Você olhou para o número	Não	Não
23	*he* + *has* + verbo-*ed* + objeto	*She has looked up the number.*	*She has looked up the number*	Ela olhou para cima o número.	Ela olhou para cima o número	Sim	Sim
24	*I* + *had* + verbo-*ed*	*I had looked up.*	*I had looked up*	Eu tinha olhado para cima.	Eu tinha olhado até	Sim	Não
25	*you* + *had* + verbo-*ed*	*You had looked up.*	*You had looked up*	Você teve olhou para cima.	Você tinha olhado para cima	Não	Sim

26	*I* + *had* + verbo-*ed* + objeto	*I had looked up the number.*	*I had looked up the number*	Eu tinha olhado o número.	Eu tinha olhado o número	Não	Não
27	*you* + *had* + verbo-*ed* + objeto	*You had looked up the number.*	*You had looked up the number*	Você tinha olhado o número.	Você tinha olhado o número	Não	Não

Fonte: elaboração própria

Quadro 3 – Tabulação dos resultados do teste com o verbo *look for*

Situação de teste		Construções		*Google Translator Toolkit*		Tradução Esperada?	
		+ ponto final	- ponto final	+ ponto final	- ponto final		
1	verbo	*Look for.*	*Look for*	Procure.	Procure		
2	verbo-*s*	*Looks for.*	*Looks for*	Procura.	Procura por	Sim	Não
3	verbo-*ed*	*Looked for.*	*Looked for*	Procurado.	Procurou	Não	Sim
4	*I* + verbo	*I look for.*	*I look for*	Eu procuro.	Eu olho para	Sim	Não
5	*you* + verbo	*You look for.*	*You look for*	Você procura.	Você olha para	Sim	Sim
6	*she* + verbo-*s*	*She looks for.*	*She looks for*	Ela procura.	Ela olha para	Sim	Sim
7	*I* + verbo + objeto	*I look for the number.*	*I look for the number*	Eu olho para o número.	Eu olho para o número	Não	Não
8	*you* + verbo + objeto	*You look for the number.*	*You look for the number*	Você olha para o número.	Você olha para o número	Sim	Sim
9	*she* + verbo-*s*+ objeto	*She looks for the number.*	*She looks for the number*	Ela olha para o número.	Ela olha para o número.	Sim	Sim
10	*I* + verbo-*ed*	*I looked for.*	*I looked for*	Eu procurava.	Olhei para	Sim	Sim

11	*you* + verbo-*ed*	*You looked for.*	*You looked for*	Você procurou.	Você olhou para	Sim	Sim
12	*I* + verbo-*ed* + objeto	*I looked for the number.*	*I looked for the number*	Eu olhei para o número.	Eu olhei para o número	Sim	Sim
13	*you* + verbo-*ed*+ objeto	*You looked for the number.*	*You looked for the number*	Você olhou para o número.	Você olhou para o número	Sim	Sim
14	*I* + *will* + verbo	*I will look for.*	*I will look for*	Eu vou procurar.	Vou olhar para	Sim	Sim
15	*you* + *will* + verbo	*You will look for.*	*You will look for*	Você vai procurar.	Você vai procurar	Sim	Não
16	*I* + *will* + verbo + objeto	*I will look for the number.*	*I will look for the number*	Vou olhar para o número.	Vou olhar para o número	Sim	Sim
17	*you* + *will* + verbo + objeto	*You will look for the number.*	*You will look for the number*	Você vai olhar para o número.	Você vai olhar para o número	Sim	Sim
18	*I* + *have* + verbo-*ed*	*I have looked for.*	*I have looked for*	Eu tenho procurado.	Eu olhei para	Sim	Sim
19	*you* + *have* + verbo-*ed*	*You have looked for.*	*You have looked for*	Você tem procurado.	Você tem procurado.	Sim	Não
20	*she* + *has* + verbo-*ed*	*She has looked for.*	*She has looked for.*	Ela tem procurado.	Ela tem procurado	Sim	Sim
21	*I* + *have* + verbo-*ed* + objeto	*I have looked for the number.*	*I have looked for the number*	Eu olhei para o número.	Eu olhei para o número	Sim	Sim
22	*you* + *have* + verbo-*ed* + objeto	*You have looked for the number.*	*You have looked for the number*	Você olhou para o número.	Você olhou para o número	Sim	Sim

23	*she* + *has* + verbo-*ed* + objeto	*She has looked for the number.*	*She has looked for the number*	Ela olhou para o número.	Ela olhou para o número.	Sim	Sim
24	*I* + *had* + verbo-*ed*	*I had looked for.*	*I had looked for*	Eu tinha procurado.	Eu tinha procurado	Sim	Não
25	*you* + *had* + verbo-*ed*	*You had looked for.*	*You had looked for*	Você tinha procurado.	Você tinha procurado	Sim	Sim
26	*I* + *had* + verbo-*ed* + objeto	*I had looked for the number.*	*I had looked for the number*	Eu tinha olhado para o número.	Eu tinha olhado para o número	Sim	Sim
27	*you* + *had* + verbo-*ed* + objeto	*You had looked for the number.*	*You had looked for the number*	Você tinha olhado para o número.	Você tinha olhado para o número	Sim	Sim

Fonte: elaboração própria

A análise dos quadros 2 e 3 mostra que o verbo *look up* não foi reconhecido nas 10 situações de teste com o padrão sintático [sujeito + verbo + objeto], com ou sem ponto final, do quadro 4:

Quadro 4 – Situações de teste que mostram que o verbo *look up* não é reconhecido nestas estruturas

	Situações de teste
7	***I* + verbo + objeto**
8	***you* + verbo + objeto**
9	***he* + verbo-*s* + objeto**
12	***I* + verbo-*ed* + objeto**
13	***you* + verbo-*ed* + objeto**
21	***I* + *have* + verbo-*ed* + objeto**
22	***you* + *have* + verbo-*ed* + objeto**
23	***he* + *has* + verbo-*ed* + objeto**
26	***I* + *had* + verbo-*ed* + objeto**
27	***you* + *had* + verbo-*ed* + objeto**

Fonte: elaboração própria

Já nas 3 situações de teste ***I* + *will* + verbo + objeto (com ou sem ponto)**, ***you* + *will* + verbo + objeto (com ou sem ponto)** e **verbo-*s* (sem ponto),** o verbo *look up* foi adequadamente traduzido. Nas vinte e quatro situações de teste restantes, excetuando-se a situação **verbo-*s*** (com ponto), os resultados não apresentaram um padrão: ou o verbo *look* e a partícula *up* foram literalmente traduzidos por "olhar" e "até" ou "para cima", respectivamente, ou apenas o verbo (*look*) foi traduzido para o verbo "olhar".

Contudo, nas duas primeiras situações do parágrafo anterior, o mesmo não se verifica para o verbo *look for*: a tradução foi literal "olhar para o número". Esse comportamento talvez se explique por meio de cálculos estatísticos baseados na frequência de ocorrência dos dois verbos, observando-se, com o uso do motor de busca do *Google*, os resultados da pesquisa: "*will look up"* (53.000.000 de ocorrências) e "w*ill look for"* (24.900.000 de ocorrências).

Por sua vez, o verbo *look for* foi reconhecido em um número maior de vezes, totalizando 8 situações de teste, com e sem ponto, em que a tradução adequada foi obtida: verbo; verbo-*s*; *verbo*-ed; *you* + *will*+ verbo; *you* + *have* + verbo-*ed*; *she* + *has* + verbo-*ed*; *I* + *had* + verbo-*ed* e *you* + *had* + verbo-*ed*. Em outras 7 situações de teste, a tradução adequada foi gerada apenas a partir da construção seguida do ponto final: *I* + verbo; *you* + verbo; *she* + verbo-*s*; *I* + verbo-*ed*; *you* + verbo-*ed*; *I* + *will* + verbo e *I* + *have* + verbo-*ed* .

Considerando que o verbo *look for* é muito mais frequente, da ordem de 2.730.000.000 ocorrências, que o verbo *look* up, da ordem de 1.680.000.000 ocorrências, mais uma vez são os dados estatísticos que parecem explicar a diferença de comportamento do sistema para os dois verbos.

Outro resultado do teste que parece confirmar a influência desse viés estatístico é o fato de os testes apresentarem um número maior de resultados esperados para o verbo *look for* (da ordem de 44 resultados) do que para o verbo *look up* (da ordem de 34 resultados).

Os resultados exploratórios desses testes já são suficientes para mostrar que o sistema de tradução do *Google* não apresenta resultados consistentes e suas idiossincrasias podem ser atribuídas aos diferentes pesos colocados nas possibilidades de tradução: (1) devido às discrepâncias de frequência de ocorrências das formas e construções linguísticas espalhadas na Internet e (2) devido ao fato de os dados, na Internet, não estarem disponíveis para o tradutor em quantidades constantes, não serem estáveis, mas sim variáveis com o tempo e ao sabor do momento em que a tradução é realizada. Por fim, mesmo contendo um dicionário, o *Google Tradutor* parece preteri-lo a favor da manipulação estatísticas de dados.

O quadro 5 é uma síntese comparativa dos quadros 2 e 3.

Quadro 5 – Resumo da comparação entre original e tradução dos verbos look up e look for

		Construções		*Google Translator Toolkit*		Tradução Esperada?	
		+ ponto	- ponto	+ ponto	- ponto		
1	verbo	*Look for.* *Look up.*	*Look for* *Look up*	*Procure.* *Olhe para cima.*	*Procure* *Olhe para cima.*		
2	verbo-*s*	*Looks for.* *Looks up.*	*Looks for* *Looks up*	Procura. Olha para cima.	Procura por Procura	Sim Sim	Não Não
3	verbo-*ed*	*Looked for.* *Looked up.*	*Looked for* *Looked up*	Procurado. Olhou para cima.	Procurou Procurou-se	Não Sim	Sim Não
4	*I* + verbo	*I look for.* *I look up.*	*I look for* *I look up*	Eu procuro. Olhei ao redor.	Eu olho para Olhei ao redor	Sim Não	Não Não
5	*you* + verbo	*You look for.* *You look up.*	*You look for* *You look up*	Você procura. Você olha para cima.	Você olha para Você olha para cima	Sim Sim	Sim Sim
6	*she* + verbo-*s*	*She looks for.* *She looks up.*	*She looks for* *She looks up*	Ela procura. Ela olha para cima.	Ela olha para Ela olha para cima	Sim Sim	Sim Sim
7	*I* + verbo + objeto	*I look for the number.* *I look up the number.*	*I look for the number* *I look up the number*	Eu olho para o número. Eu olho para cima o número.	Eu olho para o número Eu olho para cima o número	Não Sim	Não Sim
8	*you* + verbo + objeto	*You look for the number.* *You look up the number.*	*You look for the number* *You look up the number*	Você olha para o número. Você olha para cima o número.	Você olha para o número Você olha para cima o número	Sim Sim	Sim Sim

9	*she* + verbo-*s* + *objeto*	*She looks for the number.* *She looks up the number.*	*She looks for the number* *She looks up the number*	Ela olha para o número. Ela olha para cima o número.	Ela olha para o número. Ela olha para cima o número	Sim Sim	Sim Sim
10	*I* + verbo-*ed*	*I looked for.* *I looked up.*	*I looked for* *I looked up*	Eu procurava. Olhei ao redor.	Olhei para Olhei ao redor	Sim Não	Sim Não
11	*you* + verbo-*ed*	*You looked for.* *You looked up.*	*You looked for* *You looked up*	Você procurou. Você olhou para cima.	Você olhou para Você olhou para cima	Sim Sim	Sim Sim
12	*I* + verbo-*ed* + objeto	*I looked for the number.* *I looked up the number.*	*I looked for the number* *I looked up the number.*	Eu olhei para o número. Olhei para cima o número.	Eu olhei para o número Olhei para cima o número.	Sim Sim	Sim Sim
13	*you* + verbo-*ed* + objeto	*You looked for the number.* *You looked up the number.*	*You looked for the number* *You looked up the number*	Você olhou para o número. Você olhou para o número.	Você olhou para o número Você olhou para o número.	Sim Não	Sim Não
14	*I* + *will* + verbo	*I will look for.* *I will look up.*	*I will look for* *I will look up*	Eu vou procurar. Vou olhar para cima.	Vou olhar para Vou olhar para cima	Sim Sim	Sim Sim
15	*you* + will + verbo	*You will look for.* *You will look up.*	*You will look for* *You will look up*	Você vai procurar. Você vai olhar para cima.	Você vai procurar Você vai olhar para cima	Sim Sim	Não Sim
16	*I* + *will* + verbo + objeto	*I will look for the number.* *I will look up the number.*	*I will look for the number* *I will look up the number*	Vou olhar para o número. Eu vou procurar o número.	Vou olhar para o número Eu vou procurar o número	Sim Não	Sim Não

17	*you* + will + verbo + objeto	*You will look for the number.* *You will look up the number.*	*You will look for the number* *You will look up the number*	Você vai olhar para o número. Você vai procurar o número.	Você vai olhar para o número Você vai procurar o número	Sim Sim	Sim Sim
18	*I* + *have* + verbo-*ed*	*I have looked for.* *I have looked up.*	*I have looked for* *I have looked up*	Eu tenho procurado. Eu olhei para cima.	Eu olhei para Eu olhei para cima	Sim Sim	Sim Sim
19	*you* + *have* + verbo-*ed*	*You have looked for.* *You have looked up.*	*You have looked for* *You have looked up*	Você tem procurado. Você olhou para cima.	Você tem procurado. Você olhou para cima	Sim Sim	Não Sim
20	*she* + *has* + verbo-*ed*	*She has looked for.* *She has looked up.*	*She has looked for.* *She has looked up*	Ela tem procurado. Ela olhou para cima.	Ela tem procurado Ela olhou para cima	Sim Sim	Sim Sim
21	*I* + *have* + verbo-*ed* + objeto	*I have looked for the number.* *I have looked up the number.*	*I have looked for the number* *I have looked up the number*	Eu olhei para o número. Eu olhei para cima o número.	Eu olhei para o número Eu olhei para cima o número.	Sim Sim	Sim Sim
22	*you* + *have* + verbo-*ed* + objeto	*You have looked for the number.* *You have looked up the number.*	*You have looked for the number* *You have looked up the number*	Você olhou para o número. Você olhou para o número.	Você olhou para o número Você olhou para o número	Sim Não	Sim Não
23	*she* + *has* + verbo-*ed* + objeto	*She has looked for the number.* *She has looked up the number.*	*She has looked for the number* *She has looked up the number*	Ela olhou para o número. Ela olhou para cima o número.	Ela olhou para o número. Ela olhou para cima o número	Sim Sim	Sim Sim

24	*I* + *had* + verbo-*ed*	*I had looked for.* *I had looked up.*	*I had looked for* *I had looked up*	Eu tinha procurado. Eu tinha olhado para cima.	Eu tinha procurado Eu tinha olhado até	Sim Sim	Não Não
25	*you* + *had* + verbo-*ed*	*You had looked for.* *You had looked up.*	*You had looked for* *You had looked up*	Você tinha procurado. Você teve olhou para cima.	Você tinha procurado Você tinha olhado para cima	Sim Não	Sim Sim
26	*I* + *had* + verbo-*ed* + objeto	*I had looked for the number.* *I had looked up the number.*	*I had looked for the number* *I had looked up the number*	Eu tinha olhado para o número. Eu tinha olhado o número.	Eu tinha olhado para o número Eu tinha olhado o número	Sim Não	Sim Não
27	*you* + *had* + verbo-*ed* + objeto	*You had looked for the number.* *You had looked up the number.*	*You had looked for the number* *You had looked up the number*	Você tinha olhado para o número. Você tinha olhado o número.	Você tinha olhado para o número Você tinha olhado o número	Sim Não	Sim Não

Fonte: elaboração própria

2.1.2 Apresentando o COMPARA

Com o auxílio de um *corpus* paralelo[15] inglês-português, descrito mais adiante, e de traduções livres, também do inglês para o português, exemplificam-se cada um dos tipos e subtipos de ambiguidade.

Duas razões importantes motivam o uso de um *corpus* paralelo. A primeira delas é que, ao conter textos traduzidos, do inglês para o português, por tradutores humanos, além de fornecer os dados para a análise, ele permite a avaliação da qualidade da TA em relação à

[15] Adota-se aqui a definição de *corpus* paralelo tal como a apresentada por Frankenberg-Garcia e Santos (2002, p. 8): "uma coleção bilíngue de textos alinhados com as suas traduções".

qualidade da tradução elaborada por tradutores humanos. Isso porque, de acordo com Tagnin (2004), uma das grandes vantagens da utilização de *corpora* é a possibilidade de observação de uma língua natural em uso. O segundo fator que serviu de motivação para o uso do *corpus* paralelo foi a possibilidade de contextualização dos exemplos discutidos na literatura, utilizando-se textos que estão em uso.

Especificamente, selecionou-se o *corpus* paralelo COMPARA, que é bi-direcional, português-inglês, e reúne textos escritos originalmente em português e em inglês, publicados na África do Sul, Angola, Brasil, Estados Unidos, Moçambique, Portugal e Reino Unido, contando com autores como Aluísio Azevedo, Chico Buarque, Edgar Allan Poe, José de Sena, José Saramago, Machado de Assis, Mary Shelley, Oscar Wild, entre outros. Esses textos são armazenados em uma base de dados e alinhados com as respectivas traduções em inglês e português, permitindo que o *corpus* seja um recurso importante para o estudo da tradução humana, possibilitando, em particular, a comparação entre traduções do português e do inglês com o recuso de pesquisas automáticas (FRANKENBERG-GARCIA; SANTOS, 2002; 2003).[16]

Além do COMPARA, há outro *corpus* paralelo inglês-português no Brasil: o CorTrad[17]. Pode-se ter o acesso *online* gratuitamente ao CorTrad, que é composto de três *subcorpora*: literário (contos australianos), técnico-científico (culinária) e jornalístico (divulgação científica).

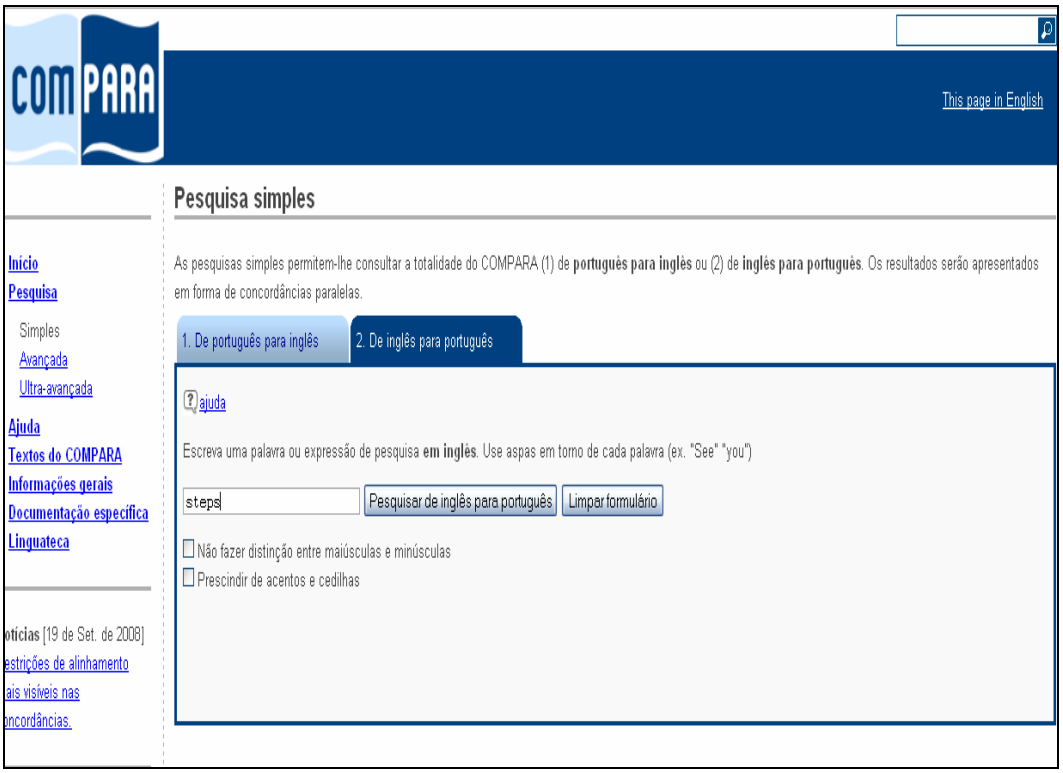
Dentre esses *subcorpora*, somente o literário pode ser consultado na direção inglês-português. Entretanto, ele é composto por apenas 27 contos escritos originalmente em inglês australiano e traduzidos para o português, sendo que esse número restrito de textos fornece poucos resultados para o objetivo desta pesquisa, que exige que se verifique a ocorrência de ambiguidades no processo de tradução de frases do inglês para o português para analisar o grau de adequação das TAs. Assim, optou-se por usar o COMPARA, que é mais robusto.

---

[16] É possível o acesso ao COMPARA *online* gratuitamente. Atualmente, o *corpus* conta com aproximadamente 3 milhões de palavras provenientes de textos de ficção. Entretanto, outros gêneros deverão ser acrescentados. (FRANKENBERG-GARCIA; SANTOS, 2002; 2003).

[17] Disponível em: <http://www.fflch.usp.br/dlm/comet/consulta_cortrad.html>. Acesso em: 25 out. 2010.

Figura 2 – Um *screenshot* da interface de consulta do COMPARA



Fonte: http://www.linguateca.pt/COMPARA/psimples.php

As buscas no *corpus* parelelo COMPARA, por meio da interface da figura 2, foram feitas sempre na direção inglês-português, sendo que essa escolha justifica-se com base nas técnicas de elaboração de dicionários bilíngues, que atestam que os dicionários inglês-português, como o Houaiss (2005), são elaborados por dicionaristas cuja língua nativa é o português, enquanto que os dicionários português-inglês, como Taylor (2003), são produzidos por dicionaristas cuja língua nativa é o inglês.

A escolha dos itens/expressões lexicais que foram alvos de busca no *corpus* foi motivada pela sistematização resultante do estudo dos tipos de ambiguidade. A própria natureza dos tipos de ambiguidade e os outros elementos presentes nos exemplos da literatura forneceram pistas para as diferentes buscas realizadas no *corpus*.[18]

---

[18] Deve-se ainda ressaltar que a utilização das versões em inglês dos textos originalmente escritos em português como dados não compromete a análise, já que o objetivo da pesquisa não é julgar a autenticidade dos textos, mas sim determinar as ambiguidades presentes nas frases do inglês que podem ser causas de má-formação ou inadequação das frases do português que são geradas automaticamente pelo *Google Tradutor*.

2.2 Explorações no domínio linguístico-computacional

Nesse domínio, em primeiro lugar, fundamentando-se na descrição do conhecimento sistematizado no domínio anterior, o puramente linguístico, investigam-se os graus de adequação da qualidade da TA de textos do inglês para o português. Essa investigação é realizada por meio da comparação das traduções feitas por humanos (livres ou extraídas do *corpus* COMPARA) e das frases geradas automaticamente pelo *Google Tradutor*, verificando em que medida a qualidade da TA é influenciada por problemas decorrentes dos diferentes tipos/subtipos de ambiguidades que serão sistematizados no Quadro 6 (p. 55). Para viabilizar a comparação, constrói-se o quadro 8 (p. 87), como uma extensão do Quadro 7 (p. 81), que apresenta as traduções feitas por tradutores humanos das frases selecionadas para ilustração, acrescentando-se a coluna correspondente às TAs. Em segundo lugar, sistematizam-se as principais estratégias computacionais para a implementação de heurísticas de resolução desses tipos de ambiguidade, apontadas na literatura sobre PLN e TA.

3. Classificando as ambiguidades linguísticas

Não se pode, no domínio da TA, descuidar do problema de resolução de ambiguidades, fenômeno amplo e que se manifesta em todos os níveis da análise linguística: lexical, sintático, semântico e pragmático-discursivo (GAZDAR e MELLISH, 1989; HIRST, 1992; DIAS-DA-SILVA, 1996; SPECIA, 2007). Como consequência dessa característica, passa-se à tipologização dos tipos de ambiguidade investigados nesse domínio.

Como explica Cruse (2006), um item lexical é ambíguo quando apresenta mais de um sentido distinto possível. O autor alerta, porém, que o termo "ambíguo" acabou se restringindo para designar somente aqueles itens lexicais que têm mais de um sentido convencionalizado, porque todos os itens lexicais podem ser ambíguos dentro de um ou outro contexto.

No uso da língua, para a compreensão, o contexto é decisivo na determinação de qual das diversas alternativas possíveis de uma leitura ambígua é a adequada. Já para a produção, a ambiguidade não é um problema, posto que, no uso ordinário da língua, o falante faz as escolhas linguísticas com um determinado sentido em mente.

A ambiguidade lexical é o tipo que se manifesta mais frequentemente, já que todos os itens lexicais podem ser ambíguos em algum momento, e é por causa da maior frequência com que esse tipo pode ser verificado que ele foi escolhido para, nesta introdução à tipologização, ilustrar os problemas que podem decorrer das ambiguidades linguísticas.

Allen (1995) define um item lexical como ambíguo quando ele projeta dois ou mais sentidos nitidamente distintos. Não deixa de apontar, porém, a imprecisão dessa definição e a necessidade se proporem testes para determinar quais sentidos são por ele convencionalmente licenciados. Os itens lexicais *kid* e *horse* do inglês ilustram essa preocupação. Como se mostrará com o teste a seguir, *kid* é ambíguo entre os sentidos "criança/filho" e "cabrito", mas *horse*, embora instancie os sentidos "cavalo", "égua" e "potro", apresenta uma subespecificação (vagueza) de sentido acentuada (já que, de certa maneira, como adverte Allen, a vagueza é uma questão de grau), não se tratando propriamente de uma ambiguidade.

Observe-se a seguinte propriedade linguística: há construções na língua que exigem que o mesmo tipo de referente denotado por um item lexical empregado em um determinado lugar da estrutura seja também acionado em outro lugar da estrutura em que esse item foi elidido. Por exemplo, na frase *João comprou três casas e Paulo comprou uma* [ ], o

substantivo elidido na posição marcada entre os colchetes necessariamente denota o mesmo tipo de objeto referido pelo substantivo *casa*.

Considerem-se, agora, estes exemplos do inglês, usados por Allen (1995, p. 232) para exemplificar a aplicação dessa propriedade como teste para mostrar que *kid* ilustra um caso de ambiguidade de sentidos e *horse*, um caso de subespecificação de sentidos, não se tratando de um caso de ambiguidade lexical propriamente dita. Na frase do inglês *Mary has two kids and George has three* [ ], o substantivo elidido na posição marcada entre os colchetes, denota o mesmo tipo de objeto referido por *kids*. Espera-se, portanto, uma destas leituras: *Mary has two kids* (FILHO) *and George has three* (FILHO) ou *Mary has two kids* (CABRITO) *and George has three* (CABRITO).[19] Já na frase *Mary has one horse and George has two* [ ], o substantivo elidido na posição marcada entre os colchetes pode denotar um tipo de objeto diferente daquele referido por *horse*. Estas leituras, dentre outras, são admissíveis: *Mary has one horse* (CAVALO) *and George has two* (CAVALO), *Mary has one horse* (CAVALO) *and George has two* (ÉGUA $\subset$ CAVALO) e *Mary has one horse* (POTRO $\subset$ CAVALO) *and George has two* (CAVALO).

Em suma, *kid* é um item lexical ambíguo (FILHO $\neq$ CABRITO), por apresentar sentidos nítidamente distintos, e *horse*, subespecificado (CAVALO $\supset$ POTRO, ÉGUA, GARANHÃO), por apresentar um sentido vago, que aglutina sentidos mais específicos não necessariamente excludentes.

Allen (1995, p. 232) exemplifica teste análogo para a detecção desse tipo de ambiguidade para os verbos do inglês *run* (ambíguo) e *kiss* (vago em relação ao seu Locativo). Na frase *I ran last year and George did* [ ] *too*, o verbo elidido na posição marcada entre os colchetes necessariamente denota o mesmo tipo de evento referido pelo verbo *run*: *I ran* (CORRER) *last year and George did* (CORRER) *too* ou *I ran* (CANDIDATAR-SE) *last year and George did* (CANDIDATAR-SE) *too*. Já na frase *I kissed Sue and George did* [ ] *too*, o Locativo do verbo elidido na posição marcada entre os colchetes pode ser diferente do Locativo do verbo *kissed*. Assim, estas, dentre outras leituras, são admissíveis: *I kissed Sue* [TESTA $\subset$ LOCAL] *and George did* [NARIZ $\subset$ LOCAL] *too* e *I kissed Sue* [BOCHECHA $\subset$ LOCAL] *and George did* [BOCA $\subset$ LOCAL] *too.*

---

[19] Note que, no português, poderia haver esta frase *Maria tem duas filhas e George dois cabritos*, em que o sentido metafórico de *cabrito* é acionado, ou melhor, "forçado" pelo sentido "filho" de *filho* selecionado no segmento anterior da construção. Assim, é tem-se: *Maria tem duas filhas e George dois cabritos* ("filhos sapecas").

Para os computadores, essas situações são causas de processamento inadequado e a representação do contexto, recurso necessário, é de alto custo para a programação, pois, para isso, as máquinas necessitam recorrer a ferramentas e a bases de dados sofisticadas para modelá-lo e para, a partir dele, selecionar ou até mesmo detectar os sentidos possíveis em contexto.

Assim, considerando-se que a ambiguidade linguística é um fato inerente às línguas naturais, quando se trata de tradução (tanto a elaborada por tradutores humanos quanto a calculada por tradutores automáticos), ela requer ainda maior atenção porque precisa ser observada em pelo menos duas línguas.

3.1 As ambiguidades linguísticas no inglês no âmbito da TA

Ide e Véronis (1998) apontam que as consequências problemáticas da manifestação das ambiguidades linguísticas é o ponto central do artigo de Bar-Hillel, mencionado na Seção 2. Tomando este trecho de um texto do ingles "*Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy*" (NIRENBURG, 1996, p. 302) como exemplo, Bar-Hillel argumenta a impossibilidade de se determinar automaticamente o sentido adequado do item lexical *pen* na frase *The box is in the pen*, reconhecendo que, para o leitor, o conhecimento de mundo sobre *pens* (CANETA) e *playpens* (CHIQUEIRINHO) é o que o faz selecionar o sentidos apropriados ao contexto do trecho. Acrescenta ainda que, se fosse possível desenvolver um sistema que tivesse acesso a esse tipo de conhecimento, seria o mesmo que conceber sistemas de TA equipados com um dicionário e também com uma enciclopédia de conhecimentos universais.

Além dos argumentos postos por Bar-Hillel, encontra-se em Somers (2000) argumentos que apontam o tratamento das ambiguidades linguísticas como o problema central da TA. Para o tradutor humano, resolver as ambiguidades no uso cotidiano da língua é uma tarefa feita sem grandes esforços, porque tanto o co-texto linguístico quanto o contexto situacional determinam auxiliam-no na interpretação apropriada; para a máquina, por outro lado, esses elementos precisam ser modelados, o que representa um empreendimento estremamente complexo e laborioso.

Kurzweil (1990) aponta que as ambiguidades, quando se ignoram os contextos (o co-texto linguístico e o contexto situacional), manifestam-se com uma frequência muito maior do

que a esperada, apontando que um analisador gramatical (um *parser*) pode encontrar mais de um milhão de interpretações sintáticas aceitáveis para uma única frase. Argumenta, assim, que, devido à multiplicidade de ambiguidades, a compreensão de frases (pelo menos) exige, entre outros tipos de conhecimento, não só o conhecimento das relações que se estabelecem entre os itens lexicais que a compõem, como também o conhecimento dos conceitos que eles lexicalizam. Não se pode compreender o sentido da frase sem se ter consciência dos conteúdos que os lexicais que a integram denotam e sem se saber como esses conteúdos se relacionam entre si. Mas, de saída, deve-se também considerar, com Jensen & Binot (1987), que recorrer à sintaxe, ou seja, ao co-texto linguístico, é o passo inicial para a minimização dessa "explosão" de ambiguidades.

Como consequência dessas características, passa-se, na subseção seguinte, à tipologização dos tipos de ambiguidade investigados nesse domínio.

3.2 Uma tipologia das ambiguidades linguísticas no inglês

Com base na literatura estudada (em particular: GAZDAR e MELLISH, 1989; ALLEN, 1995; DIAS-DA-SILVA, 1996; HIRST, 1992; JURAFSKY e MARTIN, 2000; DALE et al. 2000; MITIKOV, 2003), por abordar esse fenômeno de modo com certo grau de profundidade, propõe-se o Quadro 6, que sintetiza os cinco grandes tipos de ambiguidade e seus respectivos subtipos.

Quadro 6 – Os tipos e subtipos de ambiguidade no inglês

TIPOS DE AMBIGUIDADE	
A. LEXICAL	**A1. Ambiguidade devido à homonímia/polissemia** **A2. Ambiguidade categorial** **A3. Ambiguidade de transferência**
B. ESTRUTURAL	**B1. Ambiguidade de fixação de constituintes** B1.1 Ligação de sintagma preposicional a mais de um sintagma nominal ou verbal B1.2 Ligação de oração relativa a mais de um sintagma nominal B1.3 Ligação de sintagma preposicional a adjetivo predicativo B1.4 Ligação de um sintagma preposicional ou advérbio à posição pertencentes ou à oração principal ou à oração encaixada B1.5 Ligação de advérbio como modificador ou de sintagma verbal ou de frase B1.6 Ligação de particípio ao sujeito estrutural da frase ou à frase B1.7 Ligação simultânea de adjunto adverbial a verbos de duas frases distintas **B2. Ambiguidade de localização e de preenchimento de lacunas** **B3. Ambiguidade analítica**

	B3.1 Detecção de partículas B3.2 Diferenciação entre um sintagma preposicional e um sintagma adjetivo resultante de uma operação de alçamento e apagamento do verbo *ser/estar* aplicada ao complemento do verbo B3.3 Diferenciação entre particípio presente e adjetivo B3.4 Diferenciação entre particípio presente e substantivo B3.5 Delimitação da extensão do sintagma nominal B3.6 Diferenciação entre a oração relativa reduzida e o sintagma verbal da oração principal B3.7 Delimitação da estrutura de um sintagma nominal complexo B3.8 Interpretação ambígua de particípios e de orações adjetivas no final de frase B3.9 Diferenciação entre frases clivadas e frases do tipo sujeito-verbo-objeto B3.10 Diferenciação entre particípio passado e um sintagma verbal incompleto, resultando na ambiguidade entre um pergunta e uma ordem B3.11 Delimitação dos diferentes tipos de estrutura formados com esta sequência de elementos: *be ADJ to V* **B4. Ambiguidade de escopo da quantificação**
C. ANAFÓRICA OU REFERENCIAL	
D. TEMÁTICA	
E. PRAGMÁTICA	

Fonte: Elaboração própria

3.2.1 (A) Ambiguidade lexical

No nível lexical, um item léxico polissêmico da língua-fonte codifica mais de um sentido quando traduzido para a língua-alvo. Como consequência, há mais de uma opção de tradução na língua-alvo. Assim, as ambiguidades que se manifestam no nível lexical exigem que uma escolha seja feita entre as possíveis interpretações, porque a escolha não adequada do item léxico resulta em proposições diferentes.

Essa situação pode ser ilustrada com a frase *But it's conditioning, brain-washing: more like a trained* <u>*seal*</u>, em que se deve selecionar um destes itens lexicais do português para o item léxico *seal* do inglês: *selo*, *escudo*, *lacre* e *foca* (SOMERS, 2000, p. 333; SPECIA, 2007, p. 12).

A diversidade cultural é um dos fatores que pode explicar essas diferenças, porque influencia na forma como os conceitos do mundo se realizam sob a forma de itens léxicos.

Nesse nível, as ambiguidades lexicais subdividem-se em três subtipos: **devido à homonímia ou polissemia, categorial e de transferência** (HIRST, 1992; HUTCHINS, 1992; SOMERS, 2000; CRUSE, 2006; SPECIA, 2007).

3.2.1.1 (A1) Ambiguidade devido à homonímia ou polissemia

Embora não seja relevante ser traçada para a TA uma fronteira nítida entre polissemia e homonímia, aceita-se que os itens léxicos são polissêmicos quando os seus possíveis sentidos compartilham algum tipo de relação semântica. Para que os sentidos possam ser considerados pertencentes ao mesmo item lexical, os falantes precisam "senti-los" como relacionados. Algumas das relações que se estabelecem entre itens lexicais polissêmicos são a metáfora, a metonímia e a hiponímia.

Já os itens lexicais são classicamente considerados homônimos quando apresentam sentidos não relacionados. Cruse (2006) aponta que a maior parte dos dicionários tradicionais confere verbetes distintos para os homônimos, diferentemente do que ocorre com os itens lexicais polissêmicos, que são identificados por números dentro do mesmo verbete. Nota-se, por esse critério de relação entre sentidos, que a distinção entre a polissemia e a homonímia é subjetiva e que, se há casos em que a distinção é bastante definida, há outros em que é impossível o estabelecimenton de uma diferença**.**

Almeida (1990) argumenta que, apesar de a polissemia e a homonímia demonstrarem diferenças em suas origens, os dois fenômenos contribuem da mesma forma para a ambiguidade. Para ele, o que realmente é relevante para a investigação léxico-semântica é a especificação dos múltiplos sentidos relacionados a um único item léxico. Concluiu-se, portanto, que, para o tratamento computacional desse tipo de ambiguidade é, de fato, como aconselha Almeida, discernir sentidos e modelar, a partir disso, a biunivocidade entre forma e sentido.

3.2.1.2 (A2) Ambiguidade categorial

Os itens léxicos categorialmente ambíguos são aqueles que pertencem a categorias sintáticas múltiplas, variando de acordo com o contexto de uso, como *canto*, que, além de ser

alvo da ambiguidade devido à homonímia, é também alvo da ambiguidade categorial, porque pode ser substantivo ou verbo na primeira pessoa do presente do indicativo.

Na maioria das vezes, esse tipo de ambiguidade é solucionado pelo analisador gramatical, não representando entraves à TA (HUTCHINS, 1992; SPECIA, 2007). É por isso que Somers (2000, p. 332) argumenta que há manifestações da ambiguidade linguística que podem ser amenizadas ou até mesmo resolvidas com a aplicação de técnicas de análise sintática, embora, naturalmete, haja situações em que se exija a aplicação do contexto ou do conhecimento de mundo para resolvê-las.

3.2.1.3 (A3) Ambiguidade de transferência

Para Hutchins (1992), a ambiguidade categorial e a ambiguidade devido à homonímia ou polissemia são ambiguidades monolíngues, porque causam problemas para a análise da língua-fonte. Já as ambiguidades de transferência são bilíngues, porque se manifestam quando um item lexical da língua-fonte pode ser traduzido por vários itens/expressões léxicas da língua-alvo, manifestando-se, portando, da perspectiva da língua-alvo diante da língua-fonte; para o falante da língua-fonte, o item lexical não é percebido como ambíguo. Essa situação ilustra-se com o item lexical do inglês *wall*, que, ao ser traduzido para o português, exige a escolha entre *parede*, que denota as divisões internas de uma construção, e *muro*, que denota as divisões externas que separam ou circundam construções.

3.2.2 (B) Ambiguidades estruturais

Hutchins (1992) propõe uma importante subclassificação para as ambiguidades estruturais: as **reais**, que são reconhecidas pelo falante, e as **acidentais**, que representam problema para a máquina.

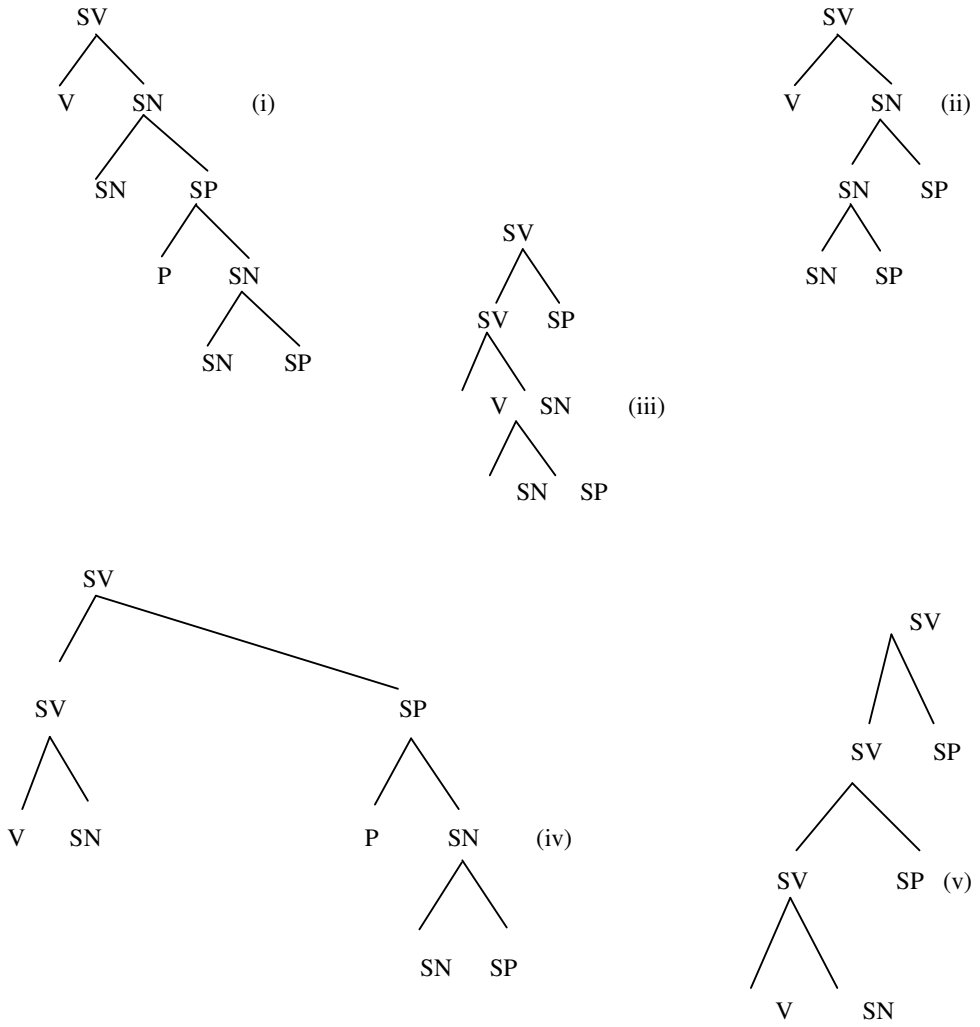
Observa-se ainda que, quando a máquina processa frases ambíguas, fazendo a substituição de itens léxicos ambíguos por outros itens léxicos potencialmente não-ambíguos, novas ambiguidades estruturais podem ocorrer. Essas estruturas, porém, não seriam consideradas ambíguas pelo falante, mas o seriam para a máquina, o que resulta no que se denominam ambiguidades acidentais. Por exemplo, para a frase *The man saw the girl with a*

telescope, mas não para a frase *The man saw the girl with a hat*, o falante encontraria mais de uma possibilidade de interpretação. A máquina, entretanto, trataria as duas frases como ambíguas.

Do ponto de vista sintático, há diferentes possibilidades de agrupamento e de fixação (hierarquização) dos constituintes oracionais. As diferentes alternativas podem resultar em frases, orações ou segmentos de frases ambíguos. A combinação ambígua de itens léxicos, constituintes e fragmentos oracionais é, muito frequentemente, e apontada como a principal causa das **ambiguidades estruturais**.

Considere-se, por exemplo, a frase do inglês *I saw the man in the house with a telescope*. Como mostra a figura 3, cinco interpretações diferentes da frase, devido à ambiguidade estrutural introduzida pelos sintagmas preposicionais (SP) (*in the...* e *with a...*), podem ser inequivocadamente distinguidas pelas cinco estrutras sintáticas distintas.

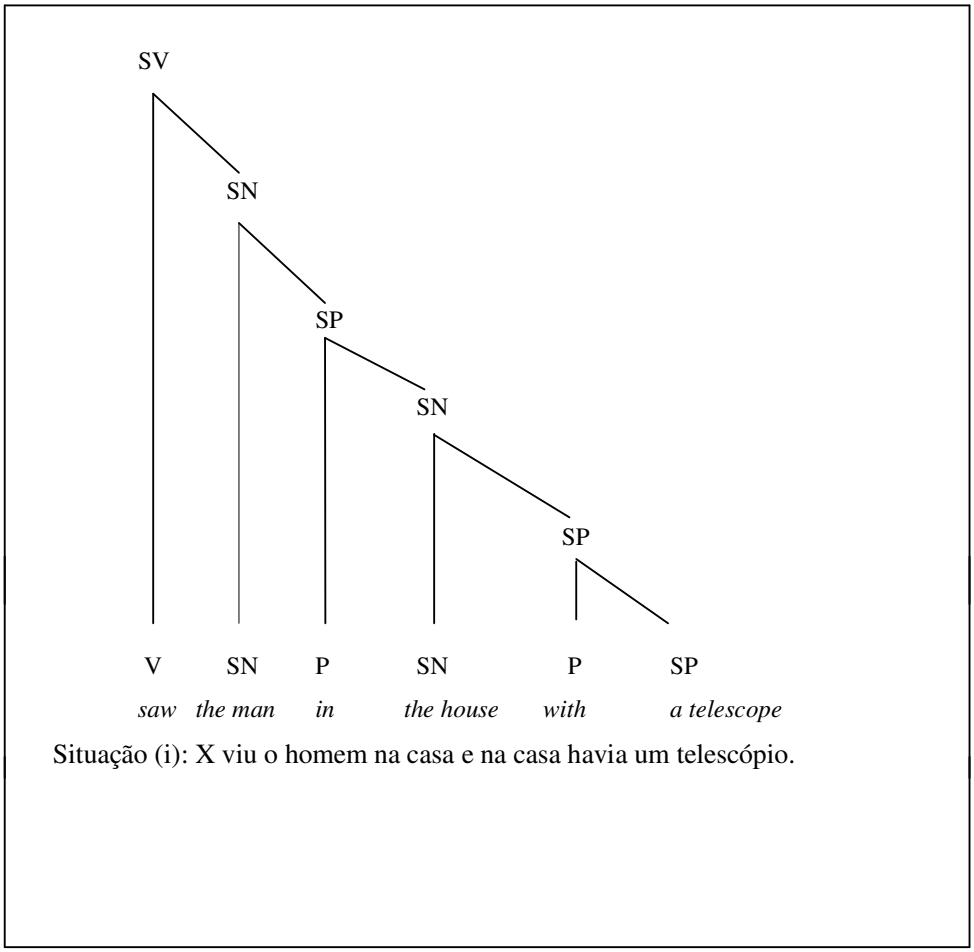
Figura 3 – Estruturas sintáticas distintas de *saw the man in the house with a telescope*.



Fonte: Allen, 1995, p. 176.

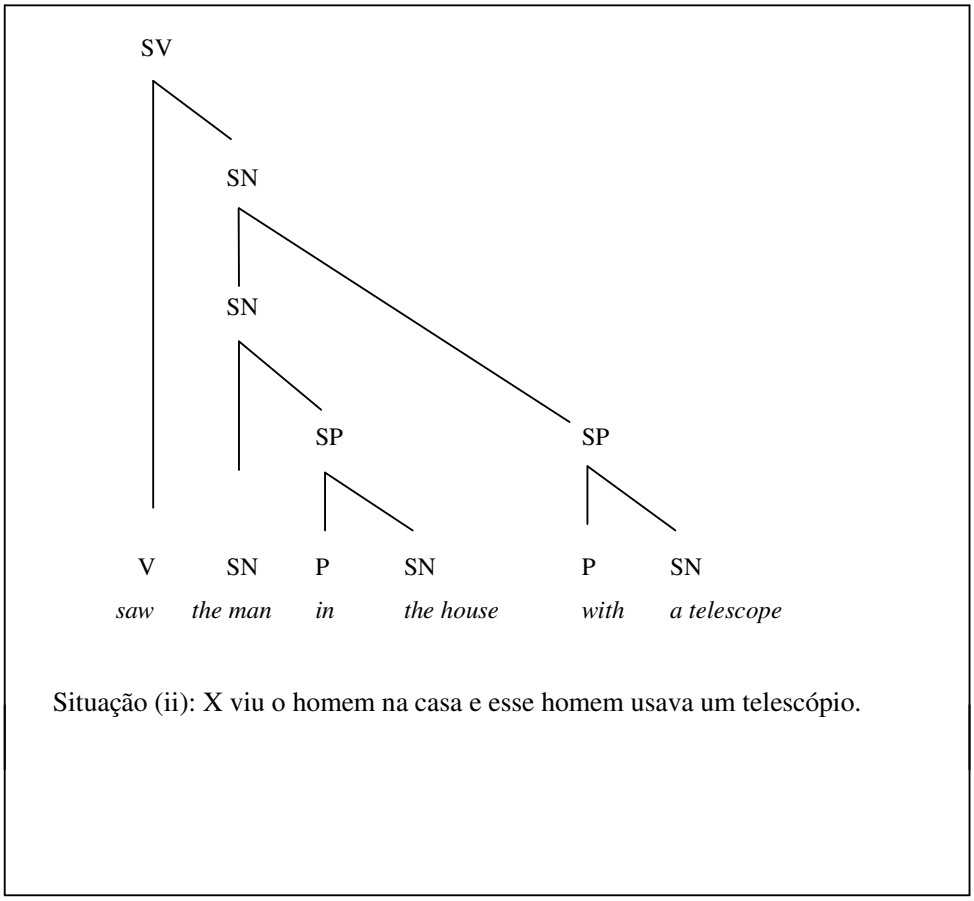
As cinco estruturas da figura 3, estão descritas nas figuras 4, 5, 6, 7 e 8, a seguir

Figura 4 - Estrutura (i)



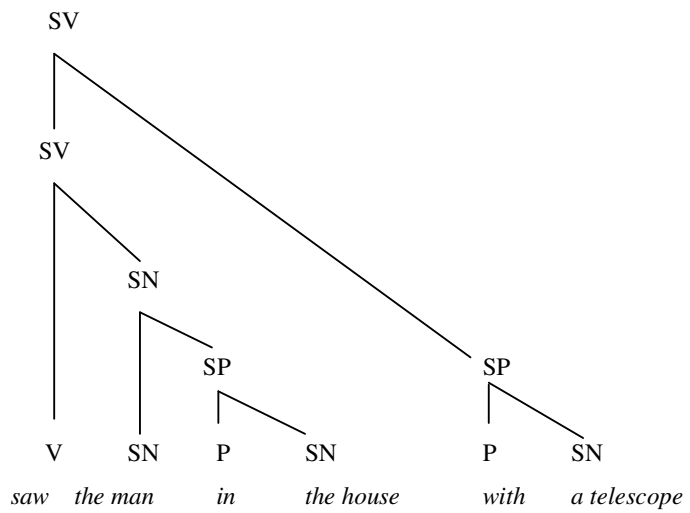
Fonte: Elaboração própria

Figura 5 - Estrutura (ii)



Fonte: Elaboração própria

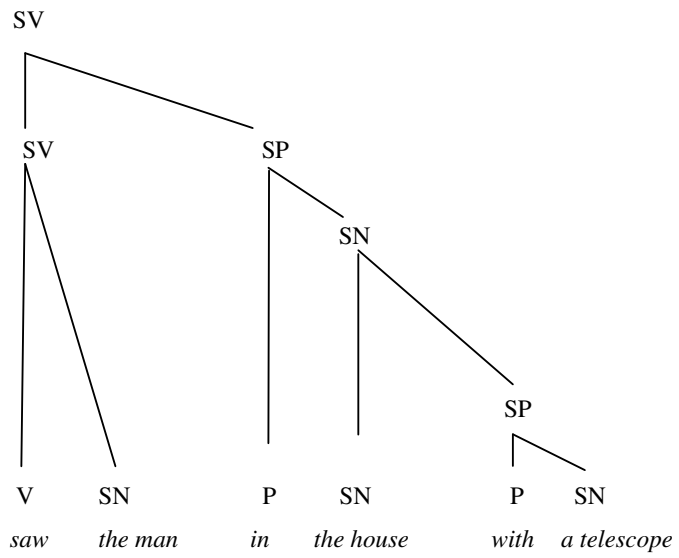
Figura 6 - Estrutura (iii)



Situação (iii): X usou um telescópio e viu o homem na casa.

Fonte: Elaboração própria

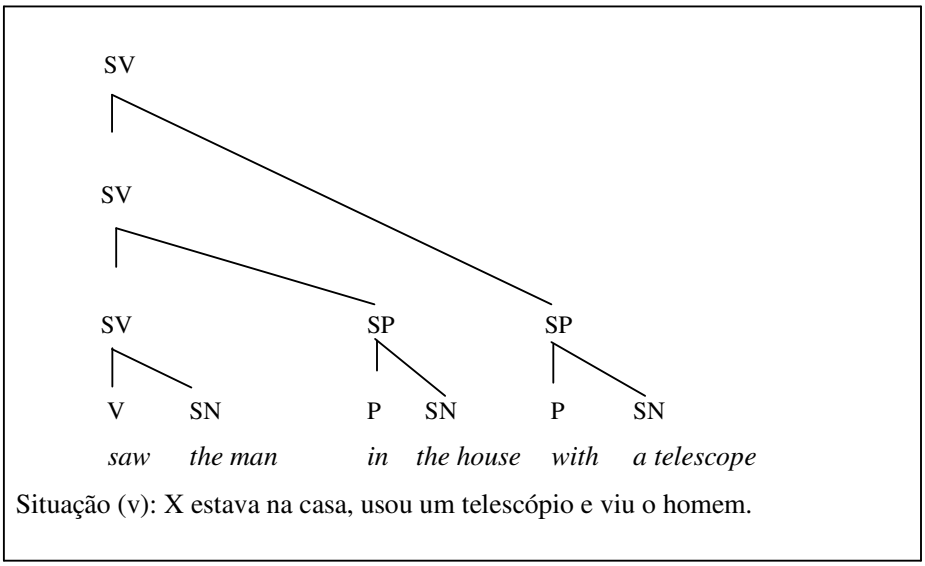
Figura 7 - Estrutura (iv)



Situação (iv): X estava na casa em que havia um telescópio e viu o homem.

Fonte: Elaboração própria

Figura 8 - Estrutura (v)



Fonte: Elaboração própria

Porém, apesar de grande parte das frases permitirem análises sintáticas múltiplas, apenas uma leitura possível permanece, após a semântica e o contexto serem considerados. Hirst (1992, p. 09) ilustra essa afirmação com a frase *Nadia left the university on the wrong bus*, afirmando que, para compreendê-la, é necessário aplicar o conhecimento de mundo de que universidades não andam de ônibus. O autor denomina esse tipo de conhecimento de "viés semântico", que sinaliza a estrutura associada à leitura preferida, em oposição ao "viés sintático", que sinaliza escolhas sintáticas preferenciais. Na frase *The landlord painted all the walls with crack*, o SP *with crack* pode se fixar ao sintagma verbal (SV), gerando as leituras "as paredes estavam sendo pintadas em um estilo 'rachaduras'" ou "as rachaduras foram usadas como instrumento para pintar as paredes", que são semanticamente anômalas. *With crack* também pode se fixar ao sintagma nominal (SN) objeto *all the walls*, podendo ser interpretado como "as paredes que apresentavam rachaduras foram pintadas".

Contudo, diferentemente das ambiguidades lexicais, que sempre exigem que uma escolha seja feita, a ambiguidade estrutural não precisa ser resolvida na língua-alvo em determinadas situações, por não comprometer o sentido da frase. Essa situação pode ser exemplificada com a frase *The man saw the girl with a telescope*, semelhante à ilustrada na figura 3, na qual não se sabe a qual destas situações ela se aplica: a menina está com o telescópio ou o homem está com o telescópio. Essa ambiguidade, que só é solucionada com o

conhecimento do contexto, pode ser transferida para a língua-alvo, sem ser resolvida, porque se trata de fenômeno análogo na língua-alvo.

As ambiguidades estruturais podem ser **locais**, situação em que apenas uma parte da frase admite mais de uma leitura, e **globais**, situação em que a frase toda admite mais de uma interpretação possível (DIAS-DA-SILVA, 1996; HIRST, 1992; SOMERS, 2000). A ambiguidade global é exemplificada com a frase da figura 3. A ambiguidade local, por sua vez, é ilustrada em Dias-da-Silva (1996, p. 103), com a frase *A empresa que comprou a Universal vendeu a Borland,* já que um sistema de TA poderia equivocadamente interpretar que foi a Universal que vendeu a Borland, considerando este trecho *a Universal vendeu a Borland* como uma oração substantiva objetiva direta de *comprou*.

Hirst (1992) argumenta que as frases localmente ambíguas geralmente levam a formação de *garden-path sentences*. Essas frases são ambíguas durante um curto espaço de tempo e causam no leitor a sensação de não ter seguido a análise adequada.[20] Esse tipo de frase é ilustrado por Allen (1995, p. 160), com a frase *The raft floated down the river sank*. Quando o *parser* processa uma frase desse tipo, ele "se confunde" no ponto a partir do qual surge a ambiguidade: ao analisar *sank*, o parser precisa desfazer uma possível análise em que estava considerando frase o segmento *The raft floated down the river*. Se a informação necessária para a desambiguação estiver fora do seu alcance, ele é obrigado a tomar uma decisão arbitrária e, se a escolha for incorreta, ele não resolve a ambiguidade resultante de uma *garden-path sentence* como essa.

Em termos tipológicos, destacam-se quatro tipos de ambiguidades estruturais: **de fixação de constituinte**, **de localização e preenchimento de lacunas**, **analítica** e **de escopo da quantificação** (ALLEN, 1995; HISRT, 1992).

3.2.2.1 B1 Ambiguidade de fixação de constituintes (attachment ambiguities)

As **ambiguidades de fixação de constituintes**, ilustradas anteriormente, manifestam-se quando é possível fixar um novo constituinte em mais de um lugar na estrutura sintática da

---

[20] Por isso que Hirst (1992), concordando com Marcus (1980), argumenta que não se deve, pelo menos em um primeiro momento, buscar a implementação de mecanismos que permitam a um *parser* resolver esse tipo de ambiguidade.

frase. Elas ocorrem quando é preciso acrescentar um modificador à frase. A frase *Ross insisted on washing the dog with the pet shampoo* (HIRST, 1992) exemplifica essa situação: o SP *with the pet shampoo* pode modificar tanto *wash* quanto o SN *the dog*. O que impede que ambas as modificações sejam possíveis são restrições semânticas.

Quando o *Google Tradutor* calcula a tradução dessa frase, ele gera *Ross insistia em lavar o cão com shampoo*, frase que mantém a ambiguidade presente na frase em inglês e é absolutamente interpretável, porque, no português, há fenômeno análogo e a decisão por uma ou outra leitura depende de informações contextuais.

3.2.2.3 (B1.1) Ligação de sintagma preposicional a mais de um sintagma nominal ou verbal

Baldwin, Kordoni e Villavicencio (2009) definem a tarefa de fixação de um SP como aquela em que é preciso especificar qual é o sintagma que governa um determinado SP. Hirst (1992, p. 132) exemplifica esse tipo de ambiguidade com a frase *The door near the stairs with the "Members only" sign had tempted Nadia from the moment she first entered the club*, em que o SP *with the "Members only" sign* pode modificar um dos dois núcleos de SN disponíveis: *stairs* ou *door*. Essa frase, ao ser submetida ao *Google Tradutor*, resulta em *A porta perto da escada com o 'Só' sinal tinha tentado Nadia a partir do momento que ela entrou pela primeira vez o clube*, em que a ambiguidade é prontamente transferida para o português, sem precisar ser mecanicamente resolvida.

A fixação de SPs é um fenômeno muito recorrente e o inglês não é exceção. Como consequência disso, a TA, que exige análise gramatical do inglês pode ser comprometida pela má fixação de SPs (BALDWIN; KORDONI; VILLAVICENCIO, 2009).

3.2.2.4 (B1.2) Ligação de oração relativa a mais de um sintagma nominal

Como exemplo desse tipo ambiguidade estrutural, considere a frase *The door near the stairs that had the "Members only" sign had tempted Nadia from the moment she first entered the club* (HIRST, 1992, p 133). A frase que o *Google Tradutor* calcula é: *A porta perto das escadas que tinha o 'Só' sinal tinha tentado Nadia a partir do momento que ela*

entrou pela primeira vez o clube. É interessante notar que, quando *stairs* é substituído por *windows*, a tradução resultante liga a oração relativa *that had the "Members only" sign* ao SN *the door*, já que o verbo é traduzido pela forma do singular (tinha), o que obriga o sistema fixar uma escolha: para ***The door*** *near* ***the windows*** *that had the "Members only" sign had tempted Nadia from the moment she first entered the club*, tem-se a TA ***A porta*** *perto* ***das janelas*** *que tinha o 'Só' sinal* ***tinha*** *tentado Nadia a partir do momento que ela entrou pela primeira vez o clube*.

Este subtipo de ambiguidade só representa um problema para a TA quando um dos SNs estiver na forma do singular e o outro na forma do plural, fato que obrigará o tradutor automático a tomar uma decisão para poder fazer a concordância verbal, e essa decisão poderá não ser adequada ao contexto.

3.2.2.5 (B1.3) Ligação de sintagma preposicional a adjetivo predicativo

Este exemplo exemplifica este tipo de ambiguidade: *He seemed nice to her* (HIRST, 1992, p. 333) Duas leituras são possíveis: "Ele pareceu agir gentilmente para com ela", quando a ligação é feita ao adjetivo predicativo (*nice*), ou "Na opinião dela, ele pareceu ser uma pessoa gentil", quando a ligação é feita ao SV (*seemed...*). O *Google Tradutor* gera a seguinte frase: *Parecia bom para ela*, em que a ambiguidade é transferida para o português.

3.2.2.6 (B1.4) Ligação de um sintagma preposicional ou advérbio a posições pertencentes ou à oração principal ou à oração encaixada

No exemplo *Ross said that Nadia had taken the cleaning out yesterday* (HIRST, 1992, p. 133), o advérbio *yesterday* pode modificar o verbo da oração principal (*said*), sendo a frase interpretada como "Ross disse ontem que Nadia havia retirado o lixo", ou o SV da oração encaixada (*taken out...*), sendo a interpretação: "Ross disse que ontem Nadia havia retirado o lixo". O *Google Tradutor* transfere a ambiguidade para o português, gerando a frase *Ross disse que a Nadia tinha tomado a limpeza ontem.*

3.2.2.7 (B1.5) Ligação de advérbio como modificador de sintagma verbal ou de frase

Hirst (1992, p. 134) ilustra esse tipo de ambiguidade com a frase *Happily, Nadia cleaned up the mess Ross had left.* Nesse exemplo, *happily* pode modificar a frase, que se aplica à descrição da situação em que o evento foi uma ocorrência feliz, ou modificar o SV, que se aplica à situação em que Nadia estava feliz. A ambiguidade dessa frase é transferida para a frase do português.

O autor enfatiza o fato de que há advérbios, como *francamente* em *Francamente, eu não entendi!*, que modificam o enunciado, expressando um comentário à respeito do que é dito. Esses advérbios, denominados marcadores discursisvos, não se ligam à dimensão frasal e são interpretados na dimensão pragmática. Como são destacados por vírgulas, em geral, não causam problemas para a TA.

3.2.2.8 (B1.6) Ligação de particípio ao sujeito estrutural da frase ou à frase

Para exemplificar esse tipo de ambiguidade, Hirst (1992, p. 134) apresenta as frases *Considering his situation likely to go from bad to worse, he decided to offer his resignation* e *Considering the deficiencies of his education, his career has been extraordinary.* Na primeira frase, o particípio presente (*considering*) modifica o sujeito explícito da oração principal (*he*), "a consideração é dele"; na segunda, modifica o sujeito do enunciado (que profere o enunciado "Considering the deficiencies of his education, his career has been extraordinary"). Ao serem automaticamente traduzidas para o português, as ambiguidade são mantidas: *Considerando sua situação susceptível de ir de mal a pior, ele decidiu oferecer a sua demissão* e *Considerando as deficiências de sua formação, sua carreira tem sido extraordinária*, respectivamente.

3.2.2.9 (B1.7) Ligação simultânea de adjunto adverbial a verbos de duas frases distintas

The lady that you meet now and then came to visit us é o exemplo que Hirst (1992, p. 134) fornece para ilustrar o fato de que o advérbio *now and then* pode modificar o verbo *meet,*

o que leva à interpretação "a senhora que você encontra de vez em quando veio nos visitar", ou o verbo *came*, o que leva à interpretação "a senhora que você encontra veio nos visitar de vez em quando". Na fala, a ambiguidade é resolvida pela entonação; na escrita, poderia ser desfeita com a inserção de vírgulas.

O *Google Tradutor*, por sua vez, traduz *now and then* literalmente (*agora e então*), e, consequentemente, de forma não esperada. Para fazer o teste da TA, substitui-se *now and then* por *sometimes*. Este foi o resultado: *A senhora que você se encontra, por vezes, veio visitar-nos*, que, apesar de o tradutor ter usado vírgulas, ele optou por colocar o advérbio entre elas, o que faz com que a ambiguidade se mantenha na língua-alvo.

3.2.3 (B2) Ambiguidade de localização e de preenchimento de lacunas

Esse tipo de ambiguidade, no inglês denominda *Gap finding and filling ambiguity*, ocorre quando um constituinte deslocado deve ser recolocado no seu local de origem e há mais de uma possibilidade de colocação. Esse tipo é ilustrado pela frase *Those are the boys that the police debated about fighting,* que possui dois lugares em que o pronome relativo referente a *the boys* pode ser recolocado: entre *debated* e *about* ou depois de *fighting* (HIRST, 1992, p. 136). O *Google Tradutor* gera a seguinte tradução: *Esses são os meninos que a polícia debatido sobre a luta*, em que o sistema não fornece tradução adequada para o verbo *debated*; por isso, esse verbo, para este teste, foi substituído pelo verbo *talked*: *Those are the boys that the police talked about fighting.* A nova tradução foi: *Esses são os meninos que a polícia falou em lutar,* em que a ambiguidade foi transferida para o português.

3.2.4 (B3) Ambiguidade analítica

Em oposição às ambiguidades de fixação de constituintes já mencionadas, em que não é a natureza do constituinte a ser fixado que causa o problema, mas o local em que ele deve ser fixado, há as ambiguidades analíticas (*analytical ambiguities*), que colocam em dúvida exatamente a natureza categorial do constituinte**.** Hirst (1992) aponta para o fato de que a língua inglesa admite muitas oportunidades para que esse tipo de ambiguidade ocorra, algumas delas serão descritas a seguir.

3.2.4.1 (B3.1) Detecção de partículas

Hirst (1992, p. 137) apresenta duas frases para ilustrar esse tipo de ambiguidade: *Ross looked up the number* e *Ross looked up the elevator shaft.* Na primeira, *up* é uma partícula integrante do verbo *look* e, na segunda, *up* é uma preposição. As respectivas traduções geradas pelo *Google Tradutor* são: *Ross olhou para cima o número* e *Ross olhou para o poço do elevador.* Ambas as ocorrências de *looked* foram traduzidas por *olhou*, desconsiderando-se a partícula *up* integrante do verbo (*look up=procurar*) ou núcleo de SP (*up=para cima*).

3.2.4.2 (B3.2) Diferenciação entre um sintagma preposicional e um sintagma adjetivo resultante de uma operação de alçamento e apagamento do verbo ser/estar aplicada ao complemento do verbo

Para a exemplificação desse tipo de ambiguidade, considere-se a frase *I want the music box on the table*. Nela, a interpretação esperada é o desejo pela caixa de música que está sobre a mesa ("Quero a caixinha de música que está sobre a mesa") ou de que ela seja colocada sobre a mesa ("Quero que a caixinha de música seja colocada sobre a mesa"). A TA dessa frase resulta em: *Eu quero a caixinha de música sobre a mesa*, situação em que a ambiguidade foi transferida para a frase do português.

3.2.4.3 (B3.3) Diferenciação entre particípio presente e adjetivo

Em primeiro lugar, é importante ressaltar que Hirst (1992) argumenta que nem sempre é possível traçar fronteiras nítidas entre os dois tipos de categorias (particípio presente e adjetivo). O autor ilustra esse tipo de ambiguidade com a frase *Ross and Nadia are singing madrigals* (HISRT, 1992, p. 140). Para esse primeiro exemplo, em que um particípio presente é utilizado, a TA é impecável: *Ross e Nadia estão cantando madrigais*.

Pens and pencils are writing implements é o segundo exemplo considerado por Hirst (1992, p. 140), mostrando o uso de um adjetivo. Para esse exemplo, a TA também é adequada: *Lápis e canetas são para escrever*.

3.2.4.4 (B3.4) Diferenciação entre particípio presente e substantivo

A frase *We discussed running* (HIRST, 1992, p. 143) é apresentada como exemplo desse tipo de ambiguidade. Nessa frase, a ambiguidade está entre a interpretação de *running* como particípio presente (correndo) ou substantivo (corridas). A TA da frase é: *Discutimos em execução,* que não é adequada, já que o sistema interpretou *running* como "executar" (um programa de computador, por exemplo) e construiu o SP com valor adverbial (de modo) *em execução*.

3.2.4.5 (B3.5) Delimitação da extensão do sintagma nominal

Estruturas formadas pela sequência de SNs que podem parecer ser um único SN são a causa desse tipo de ambiguidade. Essa situação exemplifica-se com estes exemplos: *Ross gave the dog some water, and Nadia gave the cat food* e *Ross gave the shampoo, and Nadia gave the cat food* (HIRST, 1992, p. 143). No primeiro exemplo, *the cat* e *food* nucleiam dois SNs distintos. Já no segundo, eles formam um único SN, em que o núcleo é o N *food*, modificado pelo N *cat*, formando o SN (N+N). As respectivas TAs mostram que o sistema considerou as duas situações da mesma forma (N+N): *Ross deu o cão um pouco de água, e Nadia deu a comida de gato* e *Ross deu o shampoo, e Nadia deu a comida de gato*.

Hirst (1992) adverte que as *garden path sentences* também podem ser causadas por esse tipo de ambiguidade, quando o primeiro de dois substantivos consecutivos pode ser analisado como um adjetivo.

A pré-modificação é uma das alternativas para se modificar um SN e pode ser feita por meio de diversas relações entre itens lexicais e gramaticais que o precedam. Além disso, a pré-modificação é um processo que apresenta informações de forma menos explícita do que a pós-modificação porque, em alguns casos, as relações entre os substantivos podem se tornar obscuras (QUIRK et al., 1985, p. 1321). De acordo com Quirk et al. (1985), os principais tipos de pré-modificação são: (a) por adjetivo; (b) por particípio (presente ou passado); (c) por substantivo; (d) por genitivo; (e) por advérbio e (f) por frase. Assim, a pré-modificação por substantivo, assim como ocorre no segundo dos exemplos anteriores, é geralmente uma pós-modificação por meio de um SP, como mostram Quirk et al. (op. cit): *the cat food* pode ser

reformulado como *the food of the cat*, reformulação que pode ocorrer, obviamente, apenas nos casos em que *cat* e *food* formam apenas um SN que é pré-modificado.

3.2.4.6 (B3.6) Diferenciação entre oração relativa reduzida e sintagma verbal da oração principal

Uma oração relativa reduzida pode ser confundida com o SV da oração principal, originando esse tipo de ambiguidade. Hirst (1992) aponta que essa característica gera uma *garden path sentence* clássica e exemplifica com a frase *The horse raced past the barn fell*, que pode ser reescrita como *The horse* [*that was raced past the barn*] *fell* (HIRST, 1992, p. 143). A TA é inaceitável: *O cavalo correu passado, o celeiro caiu*, ao considerar duas frases distintas.

3.2.4.7 (B3.7) Delimitação da estrutura de um sintagma nominal complexo

Esse tipo de ambiguidade deriva do fato de sequências formadas por vários grupos nominais poderem apresentar uma estrutura complexa. Para determinar essa estrutura corretamente, é preciso considerar os conhecimentos pragmático e de mundo, que devem ser aplicados para reconhecer e delimitar as partes que constituem a estrutura complexa (QUIRK et al., 1985).

Airport long term car park courtesy vehicle pickup point (HIRST, 1992, p. 144) é o SN escolhido para a exemplificação, possuindo a seguinte estrutura complexa: [[[airport [[long term] [car park]]] [courtesy vehicle]] [pickup point]]. Considerando a TA *Aeroporto de longo prazo estacionamento cortesia ponto de retirada do veículo*, é possível afirmar que determinar a estrutura de um SN complexo é uma tarefa difícil para o falante. Para o sistema de TA, praticamente implossível. A tradução desse SN do inglês é: *ponto (de parada) do transporte de cortesia do estacionamento do aeroporto.*

3.2.4.8 (B3.8) Interpretação ambígua de particípios e de orações adjetivas no final de frase

Para descrever este tipo de ambiguidade, Hirst (1992) se baseia nos argumentos e exemplos de Quirk, Greenbaum, Leech, Svartvik (1972). São estas as frases: *The manager approached me smiling* e *The manager approached us full of apologies.* Para esses dois exemplos, valem as interpretações: *"The manager approached me and he was smiling*" e "*The manager approached us and he was full of apologies*". Dois tipos de ambiguidade podem ocorrer a partir destas duas situações: (i) tanto o sujeito quanto o objeto da oração principal podem ser o sujeito da oração predicativa, como na frase *We met him leaving the room;* (ii) o particípio presente pode não ser uma oração predicativa, relacionando-se com o objeto da oração nominal ou como uma oração relativa restritiva reduzida ou ainda como um complemento do verbo.

Ilustra-se a segunda situação com as frases *The manager approached the boy smoking a cigar*, que pode ser interpretada como "o gerente, que fumava um charuto, se aproximou do garoto" ou "o garoto que fumava um charuto foi abordado pelo gerente", e *The manager caught the boy smoking a cigar*, podendo ser interpretada como "o gerente pegou o garoto no ato de fumar um charuto" e "o gerente pegou o garoto que fumava um charuto (mas o outro garoto que fumava cachimbo fugiu)". Na primeira frase, é possível a leitura predicativa, mas ela bloqueia *smoking a cigar* como complemento do verbo, já que o verbo *approached* não exige complemento. Contudo, para o segundo exemplo, essa leitura de complemento do verbo é possível, porque *catch* pode se realizar junto a um complemento.

As respectivas TAs para os dois exemplos (primeiro parágrafo desta subseção) são: *O gerente se aproximou de mim sorrindo* e *O gerente se aproximou de nós cheio de desculpas.* No primeiro, a ambiguidade é transferida para o português, cabendo ao usuário escolher o sentido apropriado. Entretanto, no segundo, o tradutor automático é obrigado a fazer a concordância entre o sintagma adjetivo (SAdj) e o substantivo que modifica, característica do português e, por isso, a TA resultante liga *full of apologies* a *the manager*. Ao se substituir o pronome *us* por *me*, por exemplo, a TA, transferindo a ambiguidade para o português, passa a ser adequada. Portanto, tanto *the manager* quanto *me* podem ser o sujeito elidido da oração predicativa sem verbo, em consonância com a explicação de Quirk, Greenbaum, Leech, Svartvik (1972): a *The manager approached me full of apologies*, corresponde a TA *O gerente se aproximou de mim cheio de desculpas*.

Para os dois últimos exemplos (segundo parágrafo desta subseção), as TAs fornecidas são: *O gerente se aproximou do garoto fumando um charuto* e *O gerente pegou o garoto fumando um charuto*, em que a ambiguidade é transferida para a língua-alvo.

Conclui-se que é problemático para o tradutor automático apenas o primeiro tipo de ambiguidade, ou seja, quando o sujeito e o objeto da oração principal podem ser o sujeito da oração predicativa e apenas quando há divergência de número e gênero.

3.2.4.9 (B3.9) Diferenciação entre frases clivadas e frases do tipo sujeito-verbo-objeto

Considere o exemplo *It frightened the child that Ross wanted to visit the lab* (HIRST, 1992, p. 147). A estrutura dessa frase permite duas interpretações: a que corresponde à análise de uma frase clivada, que utiliza o *it* antecipador (o que assustou a criança foi Ross querer visitar o laboratório) e a correspondente à análise da frase de em termos de sujeito-verbo-objeto (alguma entidade que esta sendo referida no discurso assustou aquela criança que Ross queria levar ao laboratório). A TA dessa frase é: *É medo da criança que Ross queria visitar o laboratório*, o que demonstra que esse tipo de construção precisa de tratamento especial pelo sistema.

3.2.4.10 (B3.10) Diferenciação entre particípio passado e um sintagma verbal incompleto, resultando na ambiguidade entre pergunta e ordem

Esse tipo de ambiguidade pode ser exemplificado pelas frases: *Have the crystals dissolved.* e *Have the crystals dissolved?* (Hirst, 1992, p. 147-148), que possuem estrutras sintáticas absolutamente distintas, sinalizadas pelas pontuações distintas. Na primeira, a estrutura é imperativa; na segunda e indicativa. As respectivas TAs são: *Ter os cristais dissolvidos.* e *Ter os cristais dissolvidos?*, ilustrando o fato que o tradutor não reconhece as duas funções introduzidas pelo verbo *have*, como imperativo e como auxiliar do tempo verbal presente perfeito do inglês, apresentando a tradução sem concordância com esse tempo.

3.2.4.11 (B3.11) Delimitação dos diferentes tipos de estrutura formados com esta sequência de elementos: be ADJ to V

Este tipo de ambiguidade está relacionado com a profundidade com que a ambiguidade pode se manifestar. As quatro frases a seguir (HIRST, 1992, p.148), com suas respectivas TAs, ilustram este tipo, porque, apesar de possuírem a mesma estrutura de superfície, elas diferem em um nível mais profundo de análise: *(i) Ross is eager to please* ("**Ross** is eager that **he** pleases somebody*"*), TA: *Ross está ansioso para agradar*; *(ii) Ross is ideal to please* ("Ross is ideal **for someone** to please him*"*), TA: *Ross é ideal de agradar*; *(iii) Ross is easy to please* ("**Pleasing** Ross is easy*"*) TA: *Ross é fácil de agradar*; *(iv) Ross is certain to please* ("That **Ross** will please someone is certain*"*), TA: *Ross é certo para agradar*. Note-se que (ii) e (iii) possuem o mesmo sujeito, *Ross,* para os dois verbos, *be* e *please*, já (i) e (iv) os sujeitos dos dois verbos são diferentes.

Nem sempre se pode determinar a análise gramatical apropriada de frases desse tipo a partir do adjetivo e do verbo apenas, porque, em alguns casos, elas podem coincidir. Considere a frase *The chicken is ready to eat*, que pode ter duas análises sintáticas distintas: uma que considera que o sujeito da oração principal é o mesmo que o sujeito da oração encaixada, o que leva à interpretação "o frango está pronto para comer"; outra que considera que os dois sujeitos, o da oração principal e o da encaixada, são diferentes, levando à leitura "o frango está pronto para ser comido". Nota-se que esse tipo não é problemático para a TA, porque a ambiguidade se transfere para a língua-alvo: *O frango está pronto para comer.*

3.2.5 (B4) Ambiguidade de escopo da quantificação

Este tipo de ambiguidade é discutuido em Allen (1995) e Hutchins (1992) e é exemplificado pela frase *Every boy loves a dog* (ALLEN, 1995, p. 232), em que o escopo de *every* é ambíguo: todos os meninos amam um único cão ou cada menino ama um cão diferente? Segundo Allen (1995, p. 351), esse tipo de ambiguidade se manifesta em frases que contém quantificadores, operadores lógicos, operadores modais e sintagmas adverbiais (SAdv). Além disso, acrescenta que uma frase, contendo dois ou mais quantificadores, tem grande probabilidade de ser ambígua. A seguir, exemplificam-se as diferentes situações.

Conectivo lógico: *Everyone thought that Fido* ***or*** *Fifi would win* (todo mundo pensou que Fido ganharia ou todo mundo pensou que Fifi ganharia ou cada pessoa pensou que ambos, Fido e Fifi, poderiam ganhar, etc.).

Operador verbal: *He will feed the hungriest dog tomorrow* (amanhã, ele vai alimentar o cachorro que estiver mais faminto ou amanhã ou ele vai alimentar amanhã o cachorro que hoje está mais faminto).

Sintagma Adverbial: *A fat dog always loses the race* (o perdedor da corrida é sempre um cachorro gordo qualquer ou um cachorro gordo em particular perde toda corrida).

Os quantificadores podem ser agrupados em, pelo menos, três grandes classes: **definidos**, **existenciais** e **universais**. Os **quantificadores definidos** instauram uma referência definida, situação que se caracteriza pelo uso de determinantes como *the,* podendo se referir a uma única entidade, como em *the man*, ou a um conjunto, *the fat men,* ou como os determinantes possessivos *my*, *their*, etc.: *their ideas* e *John's book*. Os quantificadores **existenciais** ou **indefinidos** introduzem novas entidades ou conjuntos no discurso: o artigo indefinido *a* (*a good book*) e quantificadores existenciais como *some* e *several*. Os **quantificadores universais** como *all, each, every* e *most* do inglês envolvem propriedades de todos ou quase todos os membros de um conjunto. Nessa terceira classe, há ainda os quantificadores que permitem uma interpretação **coletiva** (o conjunto é compreendido como uma unidade) e aqueles que permitem uma leitura **distributiva** (todos os membros do conjunto são considerados individualmente, como em *each man bought a suit*). Allen (1995, p. 353) adverte que há quantificadores que permitem apenas uma única leitura (*each*=somente distributivo); outros são ambíguos entre as duas leituras.

Observem-se estes exemplos: (i) *Each man lifted the piano.;* (ii) *Every man lifted the piano.;* (iii) *All the men lifted the piano.;* (iv) *The men lifted the piano.* A frase (i) (*each*) permite apenas a leitura distributiva, enquanto que há forte preferência para a leitura coletiva na frase (iv) (*the*). Para as frases (ii) (*every*) e (iii) (*all*), tanto a leitura coletiva quanto a distributiva está disponível. A tendência à escolha por uma leitura distributiva ou coletiva influencia as decisões de escopo, pois existem quantificadores que podem estabelecer uma relação de dependência com um SN, mas isso apenas quando são interpretados na leitura distributiva. A frase *Each man lifted a piano* ilustra essa relação de dependência*:* é possível que existam pianos diferentes para cada homem. Diz-se, então, que o SN *a piano* depende do SN *each man*, ou que este tem o escopo maior do que aquele.

A relação de dependência se estabelece apenas quando o SN que tem o escopo maior é distributivo, como *each*, por exemplo. Nota-se que, em *Together, the men lifted a piano*, o SN *the man* é coletivo e, por isso, a dependência não se verifica.

A tarefa de determinar o escopo de quantificadores e operadores não é trivial e envolve informações provenientes de diferentes níveis de análise linguística – nível lexical, sintático e semântico – sendo que o contexto também pode exercer forte influência. Elas, entretanto, por serem transferíveis da língua-fonte para a língua-alvo, não são, em geral, problemas para a TA.

3.2.6 (C) Ambiguidade anafórica ou referencial

A presença de um elemento anafórico, ou seja, um elemento da estrutura da oração que retoma, de forma direta ou indireta, uma entidade anterior que já foi estabelecida previamente no texto (anáfora interfrasal) ou está sendo estabelecida dentro da mesma frase (anáfora intrafrasal) (HALLIDAY & HASAN, 1976; ALLEN, 1995), é também uma fonte de problemas para a TA.

A diferença entre as anáforas interfrasal e intrafrasal fica clara quando se considera a frase *Jack found his hat* (ALLEN, 1995, p. 366). Observa-se que o pronome *his* pode ter como referente o SN *Jack*, configurando uma anáfora intrafrasal, mas é possível também se referir a outra entidade mencionada em uma frase anterior dentro do discurso, o que caracteriza a anáfora interfrasal.

De acordo com Mitkov et al. (1995), é possível considerar a anáfora como um tipo de ambiguidade em que um pronome pode potencialmente ligar-se a mais de um antecedente. Dessa forma, este tipo de ambiguidade manifesta-se quando há diferentes opções de retomada de antecedentes/referentes, assim como fica ilustrado em *But I do play a little of tennis still, wearing a kind of brace on the knee which keeps* ***it*** *more or less rigid*, em que o pronome *it* pode ter como antecedente *brace* ou *knee.*

Para se interpretar um pronome anafórico adequadamente, é preciso identificar os SNs que o antecedem e determinar qual deles é o seu antecedente. Resolver a anáfora é uma tarefa obrigatória para a TA quando a língua-alvo da tradução é uma língua que mostra marcação do gênero dos pronomes – como é o caso do português – e, além disso, há casos, os

predicados associados ao pronome (verbos, substantivos, etc.) podem sofrer variação de acordo com antecedentes diferentes, o que exige a sua correta identificação.

Em Somers (2000, p. 335), encontra-se um exemplo para essa situação. Considere-se a frase: *The monkey ate the banana because it was hungry*. Na TA do inglês para o português, o pronome *it*, que pode retomar *the monkey* ou *the banana*, precisa ter o seu antecedente adequadamente identificado, para que ele seja traduzido com o gênero adequado, já que o português apresenta os pronomes *ele* e *ela* como possibilidades. Considerando que *hungry* é um adjetivo próprio a seres animados, a única retomada possível é a do antecedente *the monkey*. Se, por um lado, existem situações em que a resolução da anáfora é uma obrigatoriedade, por outro lado, em algumas línguas, identificar o antecedente da anáfora é opcional, permitindo que a ambiguidade anafórica seja levada para a língua-alvo sem afetar a qualidade da TA.

3.2.7 (D) Ambiguidade temática

As ambiguidades temáticas são perceptíveis quando há diferentes opções de marcação de papéis temáticos por uma determinada preposição. Esse fato ocorre com a preposição *para* na frase *Maria trouxe um carro para Pedro*, em que *Pedro* pode ser Destinatário ou Beneficiário da ação (DIAS-DA-SILVA, 1996). Esse tipo de ambiguidade é discutido por Hirst (1992), que ilustra esse tipo de ambiguidade com a frase *Nadia tickled Ross with a feather*, indicando que *Nadia* ocupa o papel de Agente, *Ross* o de Paciente, e *feather* ocupa o papel de Instrumento, sendo este último papel temático sinalizado pela preposição *with*. Contudo, *with* pode sinalizar também outros papeis temáticos: Modo e Companhia.

Segundo Badlwin, Kordoni e Villavicencio (2009), os dois aspectos sintáticos mais importantes a serem considerados quando se trata das preposições é a seleção e a valência. Determinadas preposições são subcategorizadas por um regente (o verbo, na maioria das vezes), propriedade denominada de seleção**:** a preposição *with* é, por exemplo, regida ou selecionada pelo verbo *dispense*, em *dispense with introductions*

Por outro lado, as preposições não-selecionadas têm alta importância semântica e não participam da estrutura argumental do regente. Porém, a divisão entre preposições selecionadas e não-selecionadas não é rígida. A sequência *rely on* demonstra essa afirmação,

porque *on* é uma preposição especificada pela estrutura argumental do verbo, porém, preserva sua importância semântica (BADLWIN; KORDONI; VILLAVICENCIO, 2009).

No que diz respeito à valência das preposições, elas podem ser intransitivas (avalentes ou partículas) ou transitivas. Baldwin, Kordoni e Villavicencio (2009) argumentam que as preposições intransitivas podem ocorrer como (i) elemento de uma expressão multi-palavra (*pick it up*); (ii) predicados com cópula (*the doctor is in*) e (iii) modificadores pronominais (*an off day*). Já as preposições transitivas formam SPs com complementos selecionados por elas como, por exemplo, *at home*.

Adverte-se, porém, que os estudos sobre o processamento automático das preposições ainda está engatinhando. Apesar de serem muito frequentes nas línguas[21], elas têm recebido pouca atenção.

3.2.8 (E) Ambiguidade pragmática

Esse tipo de ambiguidade ocorre quando há diferentes opções de associação entre uma expressão linguística e as possíveis funções comunicativas, como se vê exemplificado em Dias-da-Silva (1996, p. 105-106), com a frase do português: *Copie*. Essa frase pode expressar tanto um pedido como uma ordem, dependendo da sua entonação. Observa-se que esse tipo de ambiguidade não é considerado neste estudo, por ser de resolução no contexto.

3.3 Quadro síntese das discussões

Com objetivo de apresentar, de forma sistematizada, os tipos/subtipos de ambiguidades estudados a partir da literatura e sinalizar o contraste entre a frase original, no inglês, e a sua tradução, no português, esta subseção apresenta, no Quadro 7, uma **síntese dos tipos/subtipos** de ambiguidades, associados às **leituras** mais salientes para cada tipo/subtipo, às **frases-exemplo** extraídas ou do COMPARA ou da própria literatura com as suas respectivas **traduções**.

Adotou-se o seguinte critério para a organização do quadro: quando as frases-ocorrência do inglês são extraídas do COMPARA, (sinalizadas com a letra C) as traduções no

---

[21] Com base no *British National Corpus,* quatro dos dez itens lexicais mais frequentes do inglês, são preposições: *of, to, in,* e *for* (BADLWIN; KORDONI; VILLAVICENCIO, 2009).

português também são extraídas desse *corpus*; quando as frases-ocorrência do inglês são extraídas da literatura, as traduções serão livres (sinalizadas por Li); observa-se, também, que o símbolo "*" será utilizado para sinalizar as frases que são agramaticais.

Quadro 7 – Síntese: tipos/subtipos, leituras e frases-exemplo com traduções

TIPOS DE AMBIGUIDADE				EXEMPLOS (extraídos da literatura ou do COMPARA)	LEITURAS	THs (extraídas do COMPARA ou livres.)
A.	**LEXICAL**					
		1. Ambiguidade devido à homonímia ou polissemia		*Hughie tucked the **seal** comfortably under one arm, and put his thumb in.* (C)	Hughie meteu a **foca** debaixo do braço e enfiou o polegar na boca/ Hughie meteu o **selo** debaixo do braço e enfiou o polegar na boca/ Hughie meteu o **lacre** debaixo do braço e enfiou o polegar na boca	*Hughie meteu a foca debaixo do braço e enfiou o polegar na boca. (C)*
		2. Ambiguida-de categorial		*A **fine reward** for trying to preserve academic standards.* (C)	Uma **recompensa pela multa** por tentar manter os padrões acadêmicos/ Uma **bela recompensa** por tentar manter os padrões acadêmicos	*Uma gloriosa recompensa por tentar manter o nível acadêmico. (C)*
		3. Ambiguida-de de transferência		*There was a man in striped pyjamas lying on one of the beds, with his face to the **wall**.* (C)	Um homem de pijama listrado estava deitado numa das camas, com o rosto virado para a **parede**/Um homem de pijama listrado estava deitado numa das camas, com o rosto virado para o **muro**	*Um homem de pijama listrado estava deitado numa das camas, com o rosto virado para a parede. (C)*
B.	**ESTRUTURAL**					
		1. Ambigui-dade de fixação de constituintes				
			1.1 Ligação de sintagma preposicional a mais de um sintagma nominal ou verbal	***I** came back into the **kitchen with the Bible**, rather pleased with myself, and read the verse out to Sally.* (C)	Eu voltei para a **cozinha em que havia uma bíblia**, orgulhoso de mim, e li o versículo em voz alta para Sally/ **Eu** voltei para a cozinha **levando uma bíblia**, orgulhoso de mim, e li o versículo em voz alta para Sally	*Eu voltei para a cozinha com a Bíblia, orgulhoso de mim, e li o versículo em voz alta para Sally. (C)*
			1.2 Ligação de oração relativa a mais de um sintagma nominal	***The door** near **the stairs that had the** "Members only" sign had tempted Nadia from the moment she first entered the*	**A porta com uma placa com os dizeres "Members only" e que fica perto da escada** chamou a atenção de Nadia	*A porta com uma placa com os dizeres "Member only" e que fica perto da escada chamou a atenção de Nadia desde o momento que ela*

			club. (L) (HIRST, 1992, p. 132)	desde o momento que ela entrou no clube/ **Há uma porta perto da escada e nessa escada há uma placa** com os dizeres "Members only" que chamou a atenção de Nadia desde o momento que ela entrou no clube	*entrou no clube. (Li)*
		1.3 Ligação de sintagma preposicional a adjetivo predicativo	*He seemed nice **to her**.* (L) (HIRST, 1992, p. 333)	**Na opinião dela**, ele parecia gentil/Ele parecia se comportar de maneira gentil **para com ela**	*Na opinião dela, ele parecia ser gentil. (Li)*
		1.4 Ligação de um sintagma preposicional ou advérbio à posição pertencentes ou à oração principal ou à oração encaixada	*Ross **said** that Nadia had **taken** the cleaning **out yesterday**.* (L) (HIRST, 1992, p. 133)	O Ross disse que a Nadia tinha **colocado**, **ontem**, o lixo na rua/ **Ontem**, o Ross **disse** que a Nadia tinha colocado o lixo na rua	*O Ross disse que a Nadia tinha colocado ontem o lixo na rua. (Li)*
		1.5 Ligação de advérbio como modificador ou de sintagma verbal ou de frase	***Happily***,**[***Nadia **cleaned up** the mess Ross had left***.***]** (L) (HIRST, 1992, p. 134)	A Nadia **arrumou alegremente** a bagunça que Ross fez/ **Felizmente**, **[**a Nadia arrumou a bagunça que o Ross fez**]**	*Nadia arrumou alegremente a bagunça que Ross fez. (Li)*
		1.6 Ligação de particípio ao sujeito estrutural da frase ou à frase	**Considering** his situation likely to go from bad to worse, **[he** decided to offer his resignation**]** (L) (HIRST, 1992, p. 134)	***Ele, ao considerar*** *que situação dele ia indo de mal para pior, decidiu pedir demissão/* ***Considerando-se (cá entre nós)*** *que a situação dele ia indo de mal para pior, ele decidiu pedir demissão*	***Ele, ao considerar*** *que situação dele ia indo de mal para pior, decidiu pedir demissão. (Li)*
		1.7 Ligação simultânea de adjunto adverbial a verbos de duas frases distintas	*The lady that you meet **now and then** came to visit us.* (L) (HIRST, 1992, p. 134)	A senhora que você encontra **de vez em quando** veio nos visitar/A senhora que você encontra veio, **de vez em quando**, nos visitar	*A senhora que você encontra de vez em quando veio nos visitar. (Li)*
	2. Ambiguidade de localização e de preenchimento de lacunas		*Those are the boys **that the police discussed about fighting**.* (L) (HIRST, 1992, p. 136)	Aqueles são os meninos **com quem a polícia discutiu lutas**/Aqueles são os meninos **contra os quais a polícia discutiu se iam ou não lutar**	*Aqueles são os meninos com quem a polícia discutiu lutas. (Li)*
	3. Ambiguidade analítica				
		3.1 Detecção de partículas	*Ross **looked up** the number.* (L) (HIRST, 1992, p. 137)	Ross **procurou** o número/Ross **visitou** do número	*Ross procurou o número. (Li)*
		3.2 Diferenciação entre um sintagma preposicional e um sintagma adjetivo resultante de uma operação de	*I want the music box **on the table**.* (L) (HIRST, 1992, p. 149)	Eu quero a caixinha de música **que está sobre a mesa**/ Eu quero que a caixinha de música **seja colocada sobre**	*Eu quero a caixinha de música que está sobre a mesa. (Li)*

		alçamento e apagamento do verbo *ser/estar* aplicada ao complemento do verbo		**a mesa**	
		3.3 Diferenciação entre particípio presente e adjetivo	*Ross and Nadia are **singing madrigals**.* (L) (HIRST, 1992, p. 140)	O Ross e a Nadia estão **cantando madrigais/** O Ross e a Nadia são **cantores madrigais**	*Ross e Nadia estão cantando madrigais. (Li)*
		3.4 Diferenciação entre particípio presente e substantivo	*We discussed **running**.* (L) (HIRST, 1992, p. 143)	Nós discutimos **corridas/** Nós discutimos **correndo**	*Nós discutimos corridas. (Li)*
		3.5 Delimitação da extensão do sintagma nominal	*Ross gave the dog some water, and Nadia gave **the cat food**.* (L) (HIRST, 1992, p. 143)	Ross deu um pouco de água ao cachorro e Nadia **deu comida ao gato**/ Ross deu um pouco de água ao cachorro e Nadia **deu comida de gato**	*Ross deu um pouco de água ao cachorro e Nadia deu comida ao gato. (Li)*
		3.6 Diferenciação entre a oração relativa reduzida e o sintagma verbal da oração principal	***The horse raced past** the barn fell.* (L) (HIRST, 1992, p. 143)	***O cavalo passou correndo** pelo celeiro caiu /* **O cavalo que passou correndo** pelo celeiro caiu	*O cavalo que passou correndo pelo celeiro caiu. (Li)*
		3.7 Delimitação da estrutura de um sintagma nominal complexo	*[[[Airport [[long term] [car park]]] [courtesy vehicle]] [pickup point]].* (L) (HIRST, 1992, p. 144)	[[[Ponto de parada] [do carro de cortesia]]] [[do estacionamento permanente] [do aeroporto]]	*Ponto de parada do carro de cortesia do estacionamento permanente do aeroporto. (Li)*
		3.8 Interpretação ambígua de particípios e de orações adjetivas no final de frase	***The manager** approached **me smiling**.* (L) (HIRST, 1992, p. 149)	**O gerente** se aproximou de mim **sorrindo/** O gerente se aproximou de **mim** e **eu estava sorrindo**	*O gerente se aproximou de mim sorrindo. (Li)*
		3.9 Diferenciação entre frases clivadas e frases do tipo sujeito-verbo-objeto	***It** frightened the child **that Ross wanted to visit the lab**.* (L) (HIRST, 1992, p. 147).	***Isso** assustou a criança que Ross queria visitar o laboratório/* **O fato de Ross ter querido visitar o laboratório** assustou a criança	*O fato de Ross ter querido visitar o laboratório assustou a criança (Li)*
		3.10 Diferenciação entre particípio passado e um sintagma verbal incompleto, resultando na ambiguidade entre um pergunta e uma ordem	***Have** the crystals **dissolved**.* (L) (HIRST, 1992, p. 148)	**Dissolva** os cristais (**Ordem**)/Os cristais **dissolveram?** (**Pergunta**)	*Dissolva os cristais (Ordem). (Li)*
		3.11 Delimitação dos diferentes tipos de estrutura formados com esta sequência de elementos: *be ADJ to V*	*The chicken **is ready to eat**.* (L) (HIRST, 1992, p. 148)	O frango **está pronto para comer** (algo)/O frango **está pronto para ser saboreado**	*O frango está pronto para ser saboreado. (Li)*
	4. Ambiguidade de escopo da quantificação		***Every boy loves a dog**.* (L) (ALLEN, 1995, p. 232)	**Existe um cachorro que todos os garotos amam/**	*Todo garoto ama um cachorro. (Li)*

					Todos os garoto amam pelo menos um cachorro	
C.	**AMBIGUIDADE ANAFÓRICA OU REFERENCIAL**			*But I do play a little of tennis still, wearing a kind of **brace** on the **knee** which keeps **it** more or less rigid.* (C)	Mas eu ainda sei jogar tênis, usando um apoio no **joelho** para mantê-**lo** de certa forma firme/ Mas eu ainda sei jogar tênis, usando um **apoio** no joelho para mantê-**lo** de certa forma firme	*Mas ainda jogo um pouco de tênis, usando um tipo de joelheira que o mantém mais ou menos rígido. (C)*
D.	**AMBIGUIDADE TEMÁTICA**			*Nadia tickled Ross **with glee** (MANNER)* . *Nadia tickled Ross **with a feather** (INSTRUMENT).* *(HUTCHINS & SOMERS, 1992, p.7)*	Nadia fez cócegas em Ross **com alegria** (MODO)/ Nadia fez cócegas em Ross **com uma pena** (INSTRUMENTO)	Nadia fez cócegas em Ross com alegria / Nadia fez cócegas em Ross com uma pena *(Li)*
E.	**AMBIGUIDADE PRAGMÁTICA**			*Be quiet, woman.* **(Request/ Order)** (C)	Fique quieta, mulher (**Pedido**)/ Fique quieta, mulher (**Ordem**)	*Fique quieta, mulher. (C)*

Fonte: elaboração própria.

Com base nas informações desse quadro, eveidencia-se que presença de ambiguidades nas frases do inglês exige a aplicação de conhecimentos provenientes de diferentes níveis de análise linguística. Como se mostrará na seção seguinte, ao serem submetidas à TA, as frases-ocorrência apresentaram os mais variados comportamentos, parte dos quais atribuem-se às ambiguidades linguísticas nelas presentes.

4. Discutindo as ambiguidades linguísticas na TA inglês-português

Nesta seção, especificamente, discute-se a manifestação das ambiguidades linguísticas na TA inglês-português, lembrando que Allen (1995) argumenta a favor da necessidade de se fornecer a um sistema de PLN e, em particular, a um sistema de TA, todos os tipos de conhecimento relevantes sobre a língua nele modelada para que ele tenha um desempenho de qualidade.

Além de aspectos do comportamento puramente linguístico, é preciso considerar também o conhecimento de mundo. Basicamente, o sistema deve: (i) reconhecer itens lexicais, (ii) saber as formas possíveis de combinação desses itens lexicais para formar estruturas maiores, e (iii) compreender o significado desses itens e a sua contribuição para o significado de outras estruturas maiores formadas por eles. De acordo com Allen (1995), os seguintes conhecimentos são importantes para um sistema de TA baseado em conhecimento, espelhando os níveis clássicos de análise linguística:

1. Conhecimento da fonética e da fonologia (se se tratar de um sistema de reconhecimento e produção da fala): como os itens lexicais se relacionam com os sons que os realizam.

2. Conhecimento da morfologia: como os itens lexicais são formados com base nos morfemas, as unidades de significado mais básicas.

3. Conhecimento da sintaxe: como os itens lexicais podem ser combinados para formar estruturas maiores, qual a função desempenada por cada item lexical dentro dessas estruturas e quais sintagmas formam outros sintagmas.

4. Conhecimento da semântica: qual o significado dos itens lexicais e como esse significado pode ser combinado para contribuir com o significado de estruturas maiores, independentemente do contexto.

5. Conhecimento da pragmática: como usar os itens lexicais e suas combinaçloes em estrutras maiores em situações diversas e como essas situações podem influenciar a interpretação de frases.

6. Conhecimento do discurso: quais relações podem se estabelecer entre as frases de um texto, considerando que as frases anteriores influenciam na interpretação da próxima. Esse

tipo de conhecimento é particularmente importante na resolução da anáfora, já que ele é necessário para interpretação de pronomes.

7. Conhecimento de mundo: conhecimento geral do mundo, real ou imaginário, necessário para que os usuários da língua se comuniquem com eficiência.

Quando um sistema de TA se depara com dados de entrada ambíguos, ele é forçado a fazer uma escolha. Se a causa da ambiguidade for atribuída a um item lexical, essa escolha tem que ser feita em todas as situações. Quando a origem da ambiguidade se situar no nível da sintaxe, em alguns casos é possível transferir a ambiguidade para a língua-alvo, sem gerar outros problemas de interpretação, por causa da existência de fenômeno análogo também na língua-alvo, fato que representa a não-obrigatoriedade de escolha por parte do sistema.

Allen (1995) ressalta que, na maioria das situações, a estrutura da frase e o seu significado completo podem identificar o sentido adequado de um item lexical. O verbo *run*, por exemplo, apresenta os sentidos "exercitar-se fisicamente" e "operar máquinas". Esses sentidos podem ser diferenciados com base na sintaxe, pois o primeiro se realiza com o verbo intransitivo e o segundo se realiza com o verbo transitivo. Em outras situações, mesmo que a sintaxe não permita distinguir os sentidos, existe apenas uma combinação para aqueles que são possíveis. *Jack ran in the park* (correu) e *Jack ran in the election* (concorreu) (ALLEN, 1995, p. 233) têm as mesmas estruturas sintáticas. O que diferencia os sentidos do verbo *run* são os modificadores, respectivamente, *in the park* e *in the election*.

A presença de ambiguidades em frases do inglês que são submetidas ao processo de TA representa, em maior ou menor grau, dificuldades para o sistema de tradução que, na grande maioria das vezes, é obrigado a fazer uma escolha entre duas ou mais interpretações disponíveis, escolha que para as máquinas não é trivial, assim como é para os humanos.

4.1 Confronto entre a tradução humana e a tradução automática

Apresenta-se, nesta subseção, o Quadro 8, que foi construído a partir do Quadro 7 com o acréscimo de mais um tipo de informação incluída na última coluna: as TAs calculadas pelo *Google Tradutor*, que permitem realizar uma nova comparação entre as frases traduzidas por tradutores humanos (THs) e as TAs calculdas pelo tradutor automático (TA), o *Google Tradutor*. Essa comparação possibilita a apreciação da discrepância entre a TA e a TH, o

padrão de comparação, esclarecendo que o símbolo "*" foi utilizado para marcar as frases que são agramaticais.

Quadro 8 - Confronto entre a THs e a TA

TIPOS DE AMBIGUIDADE				EXEMPLOS (literatura ou COMPARA)	LEITURAS	THs (COMPARA ou livres.)	TAs (*Google Tradutor*™)
A.	**LEXICAL**						
		1. Ambiguidade devido à homonímia ou polissemia		*Hughie tucked the **seal** comfortably under one arm, and put his thumb in.* (C)	Hughie meteu a **foca** debaixo do braço e enfiou o polegar na boca/ Hughie meteu o **selo** debaixo do braço e enfiou o polegar na boca/ Hughie meteu o **lacre** debaixo do braço e enfiou o polegar na boca	*Hughie meteu a foca debaixo do braço e enfiou o polegar na boca. (C)*	** Hughie dobrado o **selo** confortavelmen-te sob um braço, e colocou o dedo dentro.*
		2. Ambiguida-de categorial		*A **fine reward** for trying to preserve academic standards.* (C)	Uma **recompensa pela multa** por tentar manter os padrões acadêmicos/ Uma **bela recompensa** por tentar manter os padrões acadêmicos	*Uma gloriosa recompensa por tentar manter o nível acadêmico. (C)*	**Uma **recompensa multa** para tentar preservar os padrões acadêmicos.*
		3. Ambiguida-de de transferência		*There was a man in striped pyjamas lying on one of the beds, with his face to the **wall**.* (C)	Um homem de pijama listrado estava deitado numa das camas, com o rosto virado para a **parede**/Um homem de pijama listrado estava deitado numa das camas, com o rosto virado para o **muro**	*Um homem de pijama listrado estava deitado numa das camas, com o rosto virado para a parede. (C)*	*Havia um homem de pijama listrado deitado em uma das camas, com o rosto para a **parede**.*
B.	**ESTRUTURAL**						
		1. Ambiguidade de fixação de constituintes					
			1.1 Ligação de sintagma preposicional a mais de um sintagma nominal ou verbal	***I** came back into the **kitchen with the Bible**, rather pleased with myself, and read the verse out to Sally.* (C)	Eu voltei para a **cozinha em que havia uma bíblia**, orgulhoso de mim, e li o versículo em voz alta para	*Eu voltei para a cozinha com a Bíblia, orgulhoso de mim, e li o versículo em voz alta para Sally. (C)*	**Voltei para a **cozinha com a Bíblia**, e não satisfeito comigo mesmo, e ler o versículo em que Sally.*

			Sally/ **Eu** voltei para a cozinha **levando uma bíblia**, orgulhoso de mim, e li o versículo em voz alta para Sally		
	1.2 Ligação de oração relativa a mais de um sintagma nominal	***The door*** *near* ***the stairs that had the*** *"Members only" sign had tempted Nadia from the moment she first entered the club.* (L) (HIRST, 1992, p. 132)	**A porta com uma placa com os dizeres "Members only" e que fica perto da escada** chamou a atenção de Nadia desde o momento que ela entrou no clube/ **Há uma porta perto da escada e nessa escada há uma placa** com os dizeres "Members only" que chamou a atenção de Nadia desde o momento que ela entrou no clube	*A porta com uma placa com os dizeres "Member only" e que fica perto da escada chamou a atenção de Nadia desde o momento que ela entrou no clube. (Li)*	****A porta perto da escada*** *que tinha o "Somente membros" sinal tinha tentado Nadia a partir do momento que ela entrou pela primeira vez o clube.*
	1.3 Ligação de sintagma preposicional a adjetivo predicativo	*He seemed nice* ***to her****.* (L) (HIRST, 1992, p. 333)	**Na opinião dela**, ele parecia gentil/Ele parecia se comportar de maneira gentil **para com ela**	*Na opinião dela, ele parecia ser gentil. (Li)*	***Parecia bom*** *para ela.*
	1.4 Ligação de um sintagma preposicional ou advérbio à posição pertencentes ou à oração principal ou à oração encaixada	*Ross* ***said*** *that Nadia had* ***taken*** *the cleaning* ***out yesterday****.* (L) (HIRST, 1992, p. 133)	O Ross disse que a Nadia tinha **colocado**, **ontem**, o lixo na rua/ **Ontem**, o Ross **disse** que a Nadia tinha colocado o lixo na rua	*O Ross disse que a Nadia tinha colocado ontem o lixo na rua. (Li)*	**Ross* ***disse*** *que tinha tomado a Nadia a limpeza* ***ontem****.*
	1.5 Ligação de advérbio como modificador de sintagma verbal ou de frase	***Happily,*****[***Nadia* ***cleaned up*** *the mess Ross had left.****]*** (L) (HIRST, 1992, p. 134)	A Nadia **arrumou alegremente** a bagunça que Ross fez/ **Felizmente**, **[**a Nadia arrumou a bagunça que o Ross fez**]**	*Nadia arrumou alegremente a bagunça que Ross fez. (Li)*	****Felizmente****, Nadia limparam a bagunça Ross tinha deixado.*
	1.6 Ligação de particípio ao sujeito estrutural da frase ou à frase	**Considering** his situation likely to go from bad to worse, **[he** decided to offer his resignation**]** (L) (HIRST, 1992,	***Ele, ao considerar*** *que situação dele ia indo de mal para pior, decidiu pedir demissão/*	***Ele, ao considerar*** *que situação dele ia indo de mal para pior, decidiu pedir demissão. (Li)*	*Considerando sua situação susceptível de ir de mal a pior, ele decidiu oferecer a sua demissão.*

			p. 134)	***Considerando-se (cá entre nós)*** *que a situação dele ia indo de mal para pior, ele decidiu pedir demissão*		
		1.7 Ligação simultânea de adjunto adverbial a verbos de duas frases distintas	*The lady that you meet* ***now and then*** *came to visit us.* (L) (HIRST, 1992, p. 134)	A senhora que você encontra **de vez em quando** veio nos visitar/A senhora que você encontra veio, **de vez em quando**, nos visitar	*A senhora que você encontra de vez em quando veio nos visitar. (Li)*	**A senhora que você se encontra* ***agora e, em seguida****, veio nos visitar.*
	2. Ambiguidade de localização e de preenchimento de lacunas		*Those are the boys* ***that the police discussed about fighting****.* (L) (HIRST, 1992, p. 136)	Aqueles são os meninos **com quem a polícia discutiu lutas**/Aqueles são os meninos **contra os quais a polícia discutiu se iam ou não lutar**	*Aqueles são os meninos com quem a polícia discutiu lutas. (Li)*	**Esses são os meninos* ***que a polícia discutido sobre a luta****.*
	3. Ambiguidade analítica					
		3.1 Detecção de partículas	*Ross* ***looked up*** *the number.* (L) (HIRST, 1992, p. 137)	Ross **procurou** o número/Ross **visitou** do número	*Ross procurou o número. (Li)*	**Ross* ***olhou para cima*** *o número.*
		3.2 Diferenciação entre um sintagma preposicional e um sintagma adjetivo resultante de uma operação de alçamento e apagamento do verbo *ser/estar* aplicada ao complemento do verbo	*I want the music box* ***on the table****.* (L) (HIRST, 1992, p. 149)	Eu quero a caixinha de música **que está sobre a mesa**/ Eu quero que a caixinha de música **seja colocada sobre a mesa**	*Eu quero a caixinha de música que está sobre a mesa. (Li)*	*Eu quero a caixinha de música* ***sobre a mesa****.*
		3.3 Diferenciação entre particípio presente e adjetivo	*Ross and Nadia are* ***singing madrigals****.* (L) (HIRST, 1992, p. 140)	O Ross e a Nadia estão **cantando madrigais**/ O Ross e a Nadia são **cantores madrigais**	*Ross e Nadia estão cantando madrigais. (Li)*	*Ross e Nadia estão* ***cantando madrigais****.*
		3.4 Diferenciação entre particípio presente e substantivo	*We discussed* ***running****.* (L) (HIRST, 1992, p. 143)	Nós discutimos **corridas**/ Nós discutimos **correndo**	*Nós discutimos corridas. (Li)*	*Discutimos* ***em execução****.*
		3.5 Delimitação da extensão do sintagma nominal	*Ross gave the dog some water, and Nadia gave* ***the cat food****.* (L) (HIRST, 1992, p. 143)	Ross deu um pouco de água ao cachorro e Nadia **deu comida ao gato**/ *Ross deu um pouco de água ao cachorro e Nadia **deu comida de gato**	*Ross deu um pouco de água ao cachorro e Nadia deu comida ao gato. (Li)*	**Ross deu o cão um pouco de água, e Nadia* ***deu a comida de gato****.*

		3.6 Diferenciação entre a oração relativa reduzida e o sintagma verbal da oração principal	***The horse raced past*** *the barn fell.* (L) (HIRST, 1992, p. 143)	***O cavalo passou correndo** pelo celeiro caiu / **O cavalo que passou correndo** pelo celeiro caiu	*O cavalo que passou correndo pelo celeiro caiu. (Li)*	**O cavalo correu passado, o celeiro caiu.*
		3.7 Delimitação da estrutura de um sintagma nominal complexo	*[[[Airport [[long term] [car park]]] [courtesy vehicle]] [pickup point]].* (L) (HIRST, 1992, p. 144)	[[[Ponto de parada] [do carro de cortesia]]] [[do estacionamento permanente] [do aeroporto]]	*Ponto de parada do carro de cortesia do estacionamento permanente do aeroporto. (Li)*	**Aeroporto de longo prazo estacionamento cortesia ponto de retirada do veículo.*
		3.8 Interpretação ambígua de particípios e de orações adjetivas no final de frase	***The manager*** *approached* ***me smiling****.* (L) (HIRST, 1992, p. 149)	**O gerente** se aproximou de mim **sorrindo**/ O gerente se aproximou de **mim** e **eu estava sorrindo**	*O gerente se aproximou de mim sorrindo. (Li)*	***O gerente*** *se aproximou de mim* ***sorrindo****.*
		3.9 Diferenciação entre frases clivadas e frases do tipo sujeito-verbo-objeto	***It*** *frightened the child* ***that Ross wanted to visit the lab****.* (L) (HIRST, 1992, p. 147).	***Isso** assustou a criança que Ross queria visitar o laboratório/ **O fato de Ross ter querido visitar o laboratório** assustou a criança	*O fato de Ross ter querido visitar o laboratório assustou a criança (Li)*	**Isto assustou a criança que Ross queria visitar o laboratório.*
		3.10 Diferenciação entre particípio passado e um sintagma verbal incompleto, resultando na ambiguidade entre um pergunta e uma ordem	***Have*** *the crystals* ***dissolved****.* (L) (HIRST, 1992, p. 148)	**Dissolva** os cristais (**Ordem**)/Os cristais **dissolveram?** (**Pergunta**)	*Dissolva os cristais (Ordem). (Li)*	**Já os cristais dissolvidos.*
		3.11 Delimitação dos diferentes tipos de estrutura formados com esta sequência de elementos: *be ADJ to V*	*The chicken* ***is ready to eat****.* (L) (HIRST, 1992, p. 148)	O frango **está pronto para comer** (algo)/O frango **está pronto para ser saboreado**	*O frango está pronto para ser saboreado. (Li)*	*O frango* ***está pronto para comer.***
	4. Ambiguidade de escopo da quantificação		***Every boy loves a dog****.* (L) (ALLEN, 1995, p. 232)	**Existe um cachorro que todos os garotos amam/ Todos os garoto amam pelo menos um cachorro**	*Todo garoto ama um cachorro. (Li)*	***Todo garoto ama um cachorro****.*
C.	**AMBIGUIDADE ANAFÓRICA OU REFERENCIAL**		*But I do play a little of tennis still, wearing a kind of* ***brace*** *on the* ***knee*** *which keeps* ***it*** *more or less rigid.* (C)	Mas eu ainda sei jogar tênis, usando um apoio no **joelho** para mantê-**lo** de certa forma firme/ Mas eu ainda sei jogar tênis, usando um **apoio** no joelho para	*Mas ainda jogo um pouco de tênis, usando um tipo de joelheira que o mantém mais ou menos rígido. (C)*	**Mas eu jogar um pouco de tênis ainda, vestindo uma espécie de* ***cinta no joelho*** *que a mantém mais ou menos rígida.*

				mantê-**lo** de certa forma firme		
D.	**AMBIGUIDADE TEMÁTICA**		*Nadia tickled Ross **with glee** (MANNER). Nadia tickled Ross **with a feather** (INSTRUMENT).* *(HUTCHINS & SOMERS, 1992, p.7)*	Nadia fez cócegas em Ross **com alegria** (MODO)/ Nadia fez cócegas em Ross **com uma pena** (INSTRUMENTO)	Nadia fez cócegas em Ross com alegria / Nadia fez cócegas em Ross com uma pena *(Li)*	**Nadia cócegas Ross com alegria./ *Nadia Ross cócegas com uma pena.*
E.	**AMBIGUIDADE PRAGMÁTICA**		*Be quiet, woman.* **(Request/ Order)** (C)	Fique quieta, mulher (**Pedido**)/ Fique quieta, mulher (**Ordem**)	*Fique quieta, mulher. (C)*	**Ser mulher, quieta.*

Fonte: elaboração própria

A observação das THs e TAs do Quadro 8 evidencia que a TA é, na maioria das vezes, afetada pela presença dos diferentes tipos/subtipos de ambiguidade linguísticas, gerando, em alguns casos, frases completamente inadequadas para o português. Com o objetivo de resolver ou amenizar o problema, a literatura apresenta várias estratégias de resolução de ambiguidades. Parte delas é o alvo de atenção da subseção seguinte.

4.2 Estratégias de resolução de ambiguidades linguísticas no inglês

Por causa do aumento do número de dados legíveis por máquina disponíveis nos últimos anos e das técnicas estatísticas que podem ser aplicadas para identificar e utilizar as informações retiradas desses dados, as propostas de resolução de ambiguidades linguísticas também aumentaram. De acordo com Stevenson e Wilks (2003), a tarefa de "desambiguação lexical de sentido" (DSL, no inglês: WSD, *word sense disambiguation*), é tema de interesse dos pesquisadores desde o começo dos estudos sobre a TA.

Os autores afirmam que a DSL é considerada uma tarefa "intermediária", por oposição às tarefas finais, porque ela é necessária, ou pelo menos traz benefícios, para o desempenho de muitas outras tarefas de PLN/TA:

> é improvável que qualquer pessoa, que não os linguistas, estaria interessada apenas em seus resultados. As tarefas finais produzem resultados para uso daqueles que não têm um interesse

> específico na língua e geralmente utilizam as tarefas intermediárias[22] (STEVENSON, WILKS, 2003, p. 260)

A TA é considerada uma tarefa final e, para citar outras, tem-se a recuperação de informação (*information retrieval)* e o processamento de fala (*speech recognition*), por exemplo. Aliás, como se vê afirmado em Ide e Véronis (1998), os primeiros trabalhos sobre DSL foram desenvolvidos dentro do contexto da TA. Os mesmos autores apontam que a tarefa de DSL é descrita como "*AI-complete"*, o que significa dizer que é um problema que poderá ser solucionado apenas quando todos os outros problemas da Inteligência Artificial também tiverem alcançado uma solução.

Resumindo: a DSL é a tarefa que identifica o significado dos itens lexicais quando inseridos em um contexto, ou seja, a DSL faz a associação de um determinado item lexical de um texto com uma definição (o seu sentido) dentre as várias que podem ser potencialmente atribuídas a ele (IDE; VÉRONIS, 1998; STEVENSON, WILKS, 2003; SPECIA, 2007). Essa tarefa requer duas etapas: (i) a determinação de todos os sentidos diferentes relevantes para cada item léxico dentro do texto (para isso, é preciso modelar no computador como os significados diferentes que podem ser atribuídos àquele item lexical se manifestam) e (ii) a escolha de um meio de atribuir um sentido apropriado a cada ocorrência de um item léxico.

Para realizar a primeira etapa, conta-se, em geral, com acervos de sentido já definidos, como os sentidos de um dicionário ou informações retiradas de um *thesaurus*. A segunda etapa é realizada com base em informações provenientes do contexto da palavra ambígua e de outras fontes de conhecimento, como recursos lexicais e enciclopédicos, e também de fontes de conhecimento feitas manualmente (IDE e VÉRONIS, 1998, p. 03).

Muitas outras alternativas foram sugeridas para solucionar as ambiguidades linguísticas. Hirst (1992), por exemplo, apresenta um conjunto de propostas para a resolução de ambiguidades que se incluem no tipo **A** A**mbiguidades lexicais**. A primeira delas é a utilização de *scripts,* ou seja, uma representação formal e esquemática de um evento ou situação, na qual existe apenas um sentido apropriado para um item lexical ambíguo. Essa estratégia deve enriquecer o componente conceitual de um sistema de TA, evitando que ele

---

[22] Texto original: *Although important, WSD is an "intermediate" task in language processing: like part-of-speech tagging or syntatic analysis, it is unlikely that anyone other than linguistis would be interested in its results for their own sake. "Final" tasks produce results of use to those without a specific interest in language and often make use of "intermediate" tasks* (STEVENSON; WILKS, 2003, p. 260).

produza resultados ambíguos. Se a frase *Hughie tucked the* ***seal*** *comfortably under one arm, and put his thumb in*[23], for processada por um sistema que contem o *script* dos "animais" ativado, ao substantivo *seal* será provavelmente atribuído o sentido "animal mamífero que vive no mar", mesmo que, em outros contextos, seja possível atribuir a ele outros sentidos diferentes. A desvantagem em se utilizar essa proposta reside em determinar quais *scripts* devem estar ativos durante o processo de TA. E, se o sistema contar com diversos *scripts* ativos, o problema será, então, decidir qual dos sentidos escolher.

Hirst (1992) aponta que conhecimentos provenientes da sintaxe, como já se mencionou, para aperfeiçoar o componente gramatical de um sistema de TA com informações sobre restrições selecionais e subcategorização dos predicados, configuram outro recurso para auxiliar na resolução de ambiguidades lexicais.

Hirst (1992) enfatiza ainda a necessidade de modelagem computacional do contexto. Essa necessidade decorre da possibilidade de os itens léxicos próximos ao item léxico ambíguo fornecerem pistas para o sistema, ou seja, é um forte indício para a desambiguação, se um dos possíveis sentidos de um item léxico corresponder semanticamente ao sentido de outro próximo a ele.

Resumindo: as seguintes estratégias são previstas como necessárias para a resolução da ambiguidade no nível lexical:

(i) Reconhecimento do contexto e co-texto,

(ii) Associações semânticas entre os itens lexicais,

(iii) Informação sobre a sintaxe das formas,

(iv) Informação sobre as restrições selecionais dos itens léxicos e

(v) Inferências lógicas e pragmáticas.

Wilks (2009) reforça essa argumentação e acrescenta que os diferentes tipos de conhecimento necessários para um sistema de TA dependem do método por ele utilizado, e confirma que a maioria dos sistemas utiliza, comumente, as seguintes informações: morfológicas, regras gramaticais e informações provenientes de léxicos. No caso do inglês, por ser uma língua de flexão simples, a morfologia não é tão necessária como o é para uma língua rica em flexões, como o português.

No que diz respeito às ambiguidades dos tipos **B1 Ambiguidade de fixação de constituintes** e **B1.3 Ligação de sintagma preposicional a adjetivo predicativo**, inúmeras

---

[23] Frase extraída do *corpus* da pesquisa e analisada na subseção 4.3.

pesquisas foram feitas sobre como os falantes decidem sobre o local adequado para fixar um novo constituinte durante a fase de análise sintática. Como resultado desses estudos, dois princípios gerais foram postulados: o princípio do ***Minimal Attachment*** ("fixação mínima") e o princípio da ***Right Association*** ("associação à direita") ou ***Late Closure*** ("posposição do fechamento estrutural") (ALLEN, 1994; HIRST, 1992)**.**

O princípio do *Minimal Attachment* é o mais geral e afirma a existência de uma preferência para a estruturação sintática que cria o menor número possível de nós na árvore sintática. A frase *The man kept the dog in the house* (ALLEN, 1995, p. 160) exemplifica o princípio. Normalmente, essa frase é interpretada com o SP *in the house* modificando o verbo *kept*, o que, consequentemente, produz uma árvore sintática com um número menor de nós. O princípio da *Right Association* estipula que um novo constituinte deve ser interpretado como parte do constituinte que está sendo construído e não deve ser fixado em nenhum outro constituinte pertencente a um nível superior na hierarquia da árvore sintática. O princípio da *Right Association* é ilustrado por Allen (1995) com a frase *George said that Henry left in his car,* que pode ter duas interpretações sintaticamente aceitáveis – "George disse que Henry saiu dirigindo o carro dele" e "George disse, dentro do carro, que Henry saiu" – sendo que a interpretação preferida é a primeira. Na interpretação preferida, o SP fixa-se ao SV que lhe é imediatamente anterior (*left*). Na outra interpretação, o SP fixa-se ao SV mais alto na árvore sintática (*said*).

Na frase *The man kept the dog in the house*, entretanto, esses dois princípios entram em conflito, porque o princípio da *Right Association* estipula que o SP deve ser fixado ao SN *the dog* e o princípio do *Minimal Attachment* estipula a fixação desse SP ao SV *kept*. Como consequência, Allen (1995) conclui que haverá situações em que as preferências lexicais serão desejáveis em detrimento de preferências baseadas nesses princípios. Por exemplo, quando um verbo exibe uma subcategorização que exige um SP, como o verbo *put,* que subcategoriza um SP com *in, on* ou *by,* então, nas situações em que houver necessidade de escolher um local para a fixação de um SP, o SV terá preferência. Em outros casos, é o próprio SP que demonstra preferência por se fixar dentro do SV. Se essas duas situações não se manifestarem, então os princípios gerais devem ser obedecidos.

A hipótese de que informações lexicais poderiam resolver parte das ambiguidades linguísticas já era a base da teoria proposta por Ford et al. (1982), que propuseram uma teoria baseada nas preferências lexicais demonstradas pelos itens léxicos. Essa teoria propõe que informações sobre preferências do momento em que constituintes devem ser fechados são

fornecidas por preferências de caso ou expectativas sobre o verbo; ou seja, cada verbo no léxico mostra marcas de casos que devem ser preenchidos. O exemplo fornecido por esses autores para ilustrar essas preferências é o verbo *discuss,* que exige o preenchimento de dois papeis: o papel semântico de Agente e o de Paciente. Logo que o verbo tenha o preenchimento de seus papeis completos, a estrutura sob análise pode ser fechada.

A teoria de Ford et al. (1982) tem sua base no verbo da frase e por isso aposta na sua correta identificação, exigindo um analisador gramatical preciso; isso implica dizer que ela não pode ser aplicada com sucesso em casos em que a ambiguidade lexical impossibilita a determinação adequada do verbo. Uma das dificuldades na utilização desta teoria, de acordo com Hirst (1992, p. 157) é que ela assume que um verbo possui um conjunto de casos preferidos e acaba por não considerar que existem verbos que não demonstram nenhuma preferência.

Kilgarriff (1997, p. 212) também defende a hipótese de que informações lexicais podem resolver grande parte das ambiguidades sintáticas, sem que os itens lexicais precisem ser desambiguados em relação ao sentido, fato já discutido em Ford et al. (1982) e em Allen (1995). Para exemplificar essa argumentação, considere as frases-exemplo extraídas de Kilgarriff (1997, p. 4), (i) *I love baking cakes with friends* e (ii) *I love baking cakes with butter icing*, já que a frase *I came back into the kitchen with the Bible, rather pleased with myself, and read the verse out to Sally,* demonstrada no Quadro 7 e que ilustra o tipo de ambiguidade **B1.1 Ligação de sintagma preposicional a mais de um sintagma nominal ou verbal**, não pode ter a ambiguidade resolvida por esse método.

Para resolver a ambiguidade de fixação do SP (*with...)*, considera-se a semântica do substantivo núcleo do SN final *(friends* ou *butter icing*). Em (i), o núcleo do SN é humano e, por isso, o SP deve ser fixado ao verbo *baking*; em (ii), o núcleo do SN é um tipo de ingrediente de bolo, a cobertura, e, consequentemente, a sua fixação deve ser feita a *cakes*. Assim, nem *friends* nem *icing* é ambíguo entre humano e ingrediente de bolo e, por isso, a DSL não é necessária.

Jensen e Binot (1987) reafirmam a importância das informações lexicais que podem ser extraídas dos dicionários. Os autores argumentam que a grande maioria das abordagens sugeridas para solucionar problemas do PLN/TA se fundamenta na codificação manual das informações. Em consequência disso, uma das críticas que pode ser feita a essas abordagens é que, quanto mais se deseja que um programa “saiba”, mais é necessário fornecer conhecimento de forma manual, e essa é uma tarefa que exige inúmeros recursos humanos e

tempo. De acordo com esses autores, não é preciso construir manualmente um repositório de conhecimento, porque existem volumosas bases de dados que armazenam informações sobre as línguas naturais previamente codificadas nos dicionários padrões e enciclopédias, ou seja, as definições dos sentidos dos itens lexicais. Representaria um grande avanço se os computadores fossem capazes de acessar todo esse conhecimento e aplicá-lo na resolução de alguns problemas, como a ambiguidade.

Muitas informações úteis para a desambiguação podem ser encontradas nos dicionários padrões que hoje podem ser acessados *on-line.* Jensen e Binot (1987) afirmam que é possível fazer com que esse conhecimento se torne disponível para acesso por sistemas de PLN com o objetivo de solucionar problemas relativos à ambiguidade. Com o uso de técnicas que possibilitam esse acesso, duas vantagens são apontadas pelos autores: o tempo gasto para codificar manualmente informações sintáticas e semânticas seria reduzido consideravelmente e essas informações seriam reunidas de uma forma mais completa.

Essa abordagem não deixa de receber críticas também porque, por serem obras feitas pelos homens, os dicionários e enciclopédias "são arbitrários e inconsistentes" e, além disso, "apenas ter a informação não garante o uso" (JENSEN e BINOT, 1987, p. 251).

O principal objetivo de Jensen e Binot (1987) é a sugestão de técnicas para processar as definições trazidas pelos dicionários padrões *on-line* para que se possa extrair delas informações semânticas úteis para resolver ambiguidades que se manifestam no momento de fixar SPs. Os autores apresentam, então, outra possibilidade de resolução da ambiguidade de fixação de constituinte, mais especificamente a do tipo **B1.1 Ligação de sintagma preposicional a mais de um sintagma nominal ou verbal**. As frases que são usadas para exemplificação são todas extraídas de Jensen e Binot (1987) porque a frase que ilustra esse tipo, apresentada no Quadro 7, é insuficiente para demonstrar todas as heurísticas propostas pelo trabalho. De acordo com os autores, consultar um dicionário *on-line* para extrair informações relevantes para o processo de desambiguação é uma opção para solucionar esse tipo de ambiguidade. Segundo estudos deles, foram encontradas, nos dicionários, as mesmas informações que já haviam sido fornecidas anteriormente por *scripts*, *frames* e outras ferramentas manualmente construídas. As seguintes informações, relevantes para a tarefa de desambiguação, são encontradas em um dicionário:

- as próprias definições;
- frases e sintagmas que servem como exemplos;
- sinônimos;

- notas sobre o uso e outros comentários.

No trabalho de Jensen e Binot (1987), as regras são desenvolvidas para guiarem a descoberta de resultados aproximados, ou seja, são desenvolvidas como **heurísticas** necessárias para processo de resolução da ambiguidade gerada pela fixação de SPs com a preposição *with*, bem como as relações que *with* sinaliza. Assim como se pode observar em Jensen e Binot (1987), a preposição *with* pode sinalizar Locativo, Tempo, Ação ou Instrumento, Cooperação, Oposição, Posse ou Característica, Separação ou Alienação, Associação, entre outros sentidos. Os autores focam no desenvolvimento de heurísticas para as relações de Instrumento e Parte-Todo, que diz respeito ao que é denominado "posse inalienável", o que significa dizer posse por natureza e não acidental.

As três frases a seguir ilustram essas relações, nestas três situações em que o SP (*with...*) pode modificar o SN *a fish* ou o V *ate*: (i) *I ate a fish* ***with a fork****; (ii) I ate a fish* ***with bones*** e (iii) *I ate a fish* ***with my fingers*** (JENSEN; BINOT, 1987, p. 253). Para processar a frase (i), o sistema de PLN/TA precisa analisar a plausibilidade de duas construções, que correspondem às duas possibilidades de ligação do SP: (a) *eat with fork* e (b) *fish with fork*. Depois de analisadas as construções, o sistema deve ordenar as soluções que encontrou e escolher aquela que tem uma classificação melhor. O esquema de classificação utilizado pelas heurísticas oferece uma pontuação maior para a construção para a qual é mais fácil validar os itens lexicais componentes e suas relações de acordo com as definições do dicionário utilizado.

O processo de validação relatado, a seguir, exemplifica a situação. Considera-se, para esse processo, a seguinte definição de *fork: Fork 1. an implement with two or more prongs used for taking up (as in eating), pitching or digging* (JENSEN; BINOT, 1987, p. 254). Procura-se *fork* em um dicionário. Se, na sua definição, for localizado o sintagma *eat with a fork*, então (a) seria validada muito facilmente. Os autores apontam, contudo, que a possibilidade de essa sequência exata não ocorrer é muito alta. Dessa forma, deve-se observar que, na definição para *fork* apresentada anteriormente, o sintagma *used for taking up* está presente, e, além disso, o dicionário define *eating* como um tipo de *taking*. Se essa relação puder ser verificada, valida-se (a).

Essas relações observadas entre as definições dos itens lexicais são estabelecidas com base na (1) identificação de padrões funcionais equivalentes que podem fazer parte das definições, por exemplo: *used for* e *with* instrumental são equivalentes; na (2) relação entre

itens lexicais presentes nas definições (geralmente itens lexicais centrais: os núcleos dos sintagmas ou então sinônimos).

O que torna possível essa tarefa de desambiguação é o fato de que os dicionários utilizam itens lexicais e/ou sintagmas específicos para compor as suas definições, que formam **padrões** reconhecíveis**.** Para as relações de Instrumento e Parte-Todo, Jensen e Binot (1987) apontam que alguns dos padrões são: Instrumento: *for, used for, used to, a means for*, etc.; Parte-Todo: *part of, arises from, end of, member of*, etc.

Tomando como exemplo a frase *I ate a fish with a fork*, o sistema deve analisar primeiramente a sequência *eat with fork*, ou seja, a fixação do SP no SV. Por meio da análise gramatical da definição de *fork* (o complemento do SP) fornecida por um dicionário e apresentada anteriormente, descobre-se um padrão Instrumento, pelo menos: *used for*. Esse padrão fornece três termos como objeto, são eles: *taking up*, *pitching* e *digging*, que representam, cada um, uma tentativa de ligação com *eat.* Fazendo a análise gramatical das definições, o sistema extrai informações úteis como o seu núcleo e, assim, consegue concluir que *eat* é um hipônimo direto de *take* (de acordo com o dicionário utilizado). Essas conclusões permitem que a ligação de *with a fork* com *eat* seja estabelecida com uma grande probabilidade. O sistema também analisa *fish with fork,* mas nenhum padrão do tipo Parte-Todo pode ser encontrado nas definições do complemento *fork*, tornando essa sequência menos provável.

Uma possível solução que pode ser aplicada na resolução dos problemas causados pelo tipo **B3.4 Diferenciação entre particípio presente e substantivo** encontra-se na argumentação de Quirk et al. (1985). Os autores esclarecem que, quando o item lexical da forma *verb+ing* sob análise é um substantivo deverbal, ou seja, um substantivo derivado de um verbo, ele mostra características próprias dos substantivos e, por isso, pode ocorrer em ambientes sintáticos compostos por determinantes e adjetivos.

Considerando a frase extraída do *corpus* COMPARA para ilustrar esse tipo (apresentada na subseção 4.3), *It would be better to miss* ***swimming*** *and go straight down to the coast road or look for my friends in their boarding houses,* é possível acrescentar o artigo definido *the* e um adjetivo, por exemplo, resultando na seguinte frase: *It would be better to miss* ***the nice swimming*** *and go straight down to the coast road or look for my friends in their boarding houses.* Contudo, se o item lexical for um particípio presente, ele apresenta características verbais, ocorrendo com advérbios e a partícula *not*. Como no exemplo considerado para ilustração *swimming* é um substantivo deverbal, a frase **It would be better*

to miss ***not swimming*** *and go straight down to the coast road or look for my friends in their boarding houses*, formada pelo acréscimo da particula *not*, é agramatical.

Considerando ainda as argumentações de Quirk et al. (1985), no que diz respeito ao tipo **B3.5 Delimitação da extensão do sintagma nominal,** é possível decidir a leitura adequada, ou seja, se a estrutura é formada por apenas um SN pré-modificado ou por dois SNs distintos, por meio de uma reformulação do trecho ambíguo. Considerando que a pré-modificação por substantivo é, geralmente, uma pós-modificação por SP, se a reformulação da estrutura ambígua utilizando a preposição *of,* como no exemplo apresentado na seção 3.2 (p. 62) (*the cat food* → *the food of the cat*), for possível e resultar em uma frase gramatical e semanticamente adequada, é um indicativo de que a estrutura sob análise é composta por apenas um SN.

No que concerne a resolução do tipo **C. Ambiguidade anafórica ou referencial**, é fato reconhecido que, para se interpretar um pronome anafórico adequadamente, é preciso identificar os SNs que o precedem e determinar qual deles é o seu antecedente. A tradução acrescenta, ainda, um aspecto que torna essa tarefa mais complexa, porque, além de identificar qual é o sintagma antecedente de um pronome anafórico, é preciso também recodificar a referência na forma de uma expressão co-referencial da língua-alvo (MITKOV, 2003). Para resolver a ambiguidade anafórica, Mitkov et al. (1995) descrevem uma versão do modelo de resolução de anáfora implementada no sistema CAT2 (SHARP, 1988), que considera algumas restrições e preferências sintáticas e semânticas para solucionar o problema representado pela referência anafórica. Nesse modelo, para eliminar os possíveis antecedentes que não são adequados, recorre-se a várias restrições sintáticas, porém, ressalta-se que a mais óbvia é a concordância de número, pessoa e gênero do pronome com seu antecedente (MITKOV et al., 1995; SAGGION; CARVALHO, 1994).

No que diz respeito às preferências sintáticas, o modelo utiliza duas: (i) preferência de paralelismo sintático – preferem-se antecedentes com a mesma função sintática que o pronome; (ii) preferência de topicalização – em primeiro lugar, procura-se estruturas topicalizadas como prováveis referentes anafóricos. No caso das restrições semânticas, dois tipos de restrição devem ser satisfeitas: a de papel temático e a de redes semânticas. A semântica de papeis temáticos restringe quais os papéis que podem preencher os casos: se um pronome anafórico preenche as restrições, é necessário que o antecedente também o faça. As redes semânticas, por sua vez, mostram relações entre conceitos e também entre conceitos e

seus atributos. Já para as preferências semânticas, escolhem-se os antecedentes que desempenham o mesmo papel semântico que o pronome anafórico.

Durante a resolução da anáfora, é preciso respeitar algumas restrições sintáticas existentes sobre como SNs podem ser co-referentes. Para ilustrar algumas dessas restrições, Allen (1995, p. 366) considera as seguintes frases (para a exemplificação, consideraram-se as frases apresentadas pelo próprio autor, já que apenas a frase presente no Quadro 7, que ilustra o tipo **C. Ambiguidade anafórica ou referencial**, é insuficiente para exemplificar todos os tipos de co-referência):

(1) **When Jack^{j} arrived at the party, she^{j} was drunk.*

(2) *When Jill^{j} arrived at the party, she^{j} was drunk.*

(3) **He^{j} said Jack^{j} wants to leave.*

(4) *Jack^{j} said he^{j} wants to leave.*

(5) **Jill^{j} saw her^{j} in the mirror.*

(6) *Jill^{j} saw herself^{j} in the mirror.*

(7) **Jill^{j} thought that Jack saw herself^{j}.*

(8) *Jill^{j} thought that Jack saw her^{j}.*

(9) **Jack^{j} thought the tired man^{j} was dying.*

(10) *Jack^{j} thought he^{j} was dying.*

De acordo com Allen (1995), a primeira restrição diz respeito às frases (1) e (2): é preciso que o SN antecedente e o pronome anafórico concordem em gênero, número e pessoa, para que eles possam ter o mesmo referente. É por essa razão que a primeira frase não é gramatical, já que *Jack* é um nome tipicamente masculino e *she* um pronome feminino. As frases (3) e (4) são comumente explicadas por uma restrição muito geral e que, por isso, acaba não sendo uma opção adequada: a ordem do antecedente e do pronome deve ser sempre aquela em que o antecedente vem antes do pronome anafórico e, por isso, a frase (4) não está bem formada porque o pronome aparece primeiro.

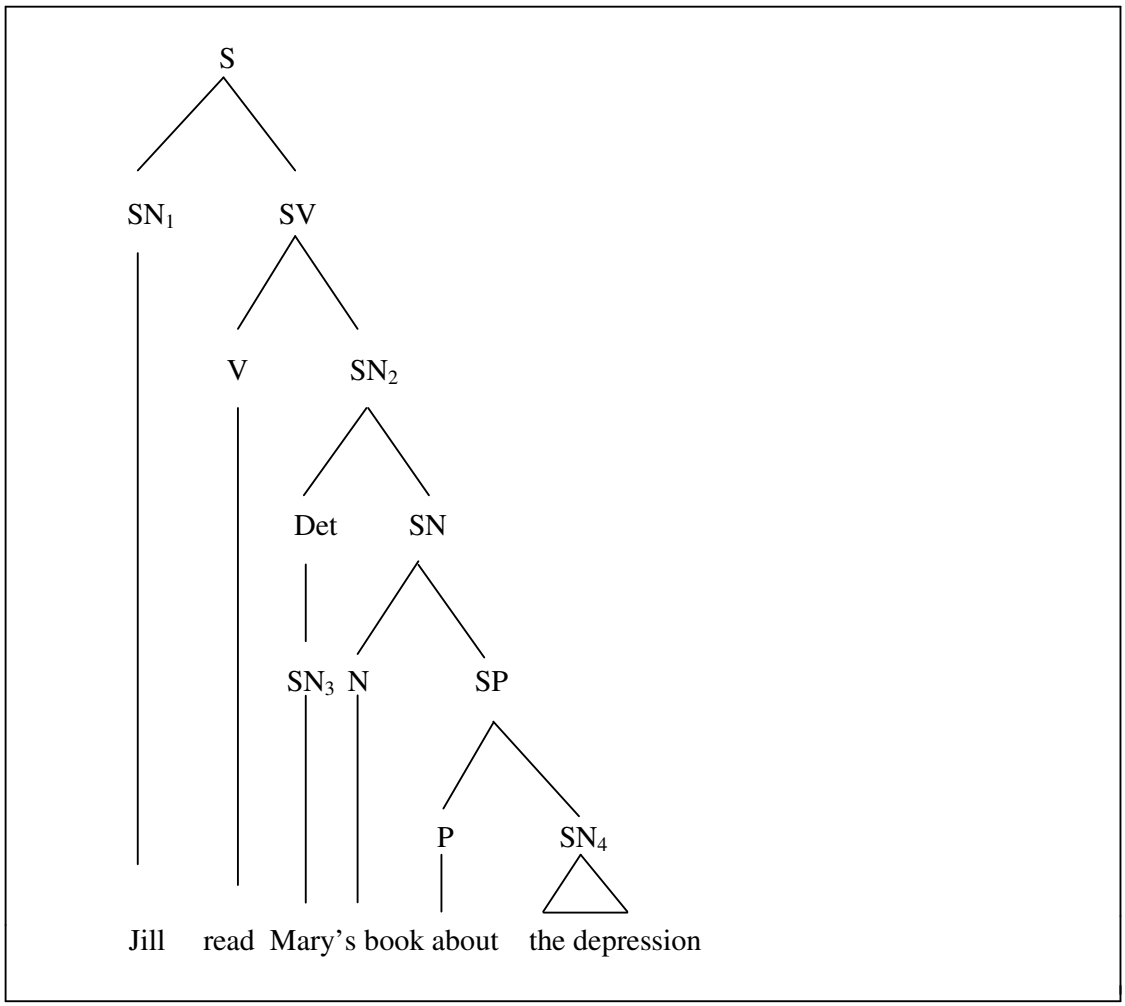
O exemplo, *In **its** dreams, the **goldfish** lives in the ocean*, retirado de Allen (1995, p. 366), mostra a presença do pronome precedendo o antecedente sem que a frase seja agramatical, isso demonstra que essa restrição não é válida para todos os casos. As restrições que dizem respeito ao uso do pronome reflexivo são ilustradas nas frases (5) – (8).

Para esses casos, um bom ponto de partida é a noção de "domínio local", a saber: o domínio local de um constituinte C é definido como sendo "o conjunto de constituintes

contidos no SN ou frase mais próximos que contenha C[24]" (ALLEN, 1995, p. 354, tradução nossa). Considerando-se a figura 9, o domínio local de SN1 é definido pelo constituinte S (constituinte dominante) e contem também SN2. SN2, por sua vez, define outro domínio local, que contem SN3 e SN4.

[24] Texto original: *"the set of constituents contained in the closest S or NP that contains C"* (ALLEN, 1995, p. 354).

Figura 9 - Estrutura sintática que ilustra a noção de domínio local



Fonte: adaptado de Allen, 1995, p. 354.

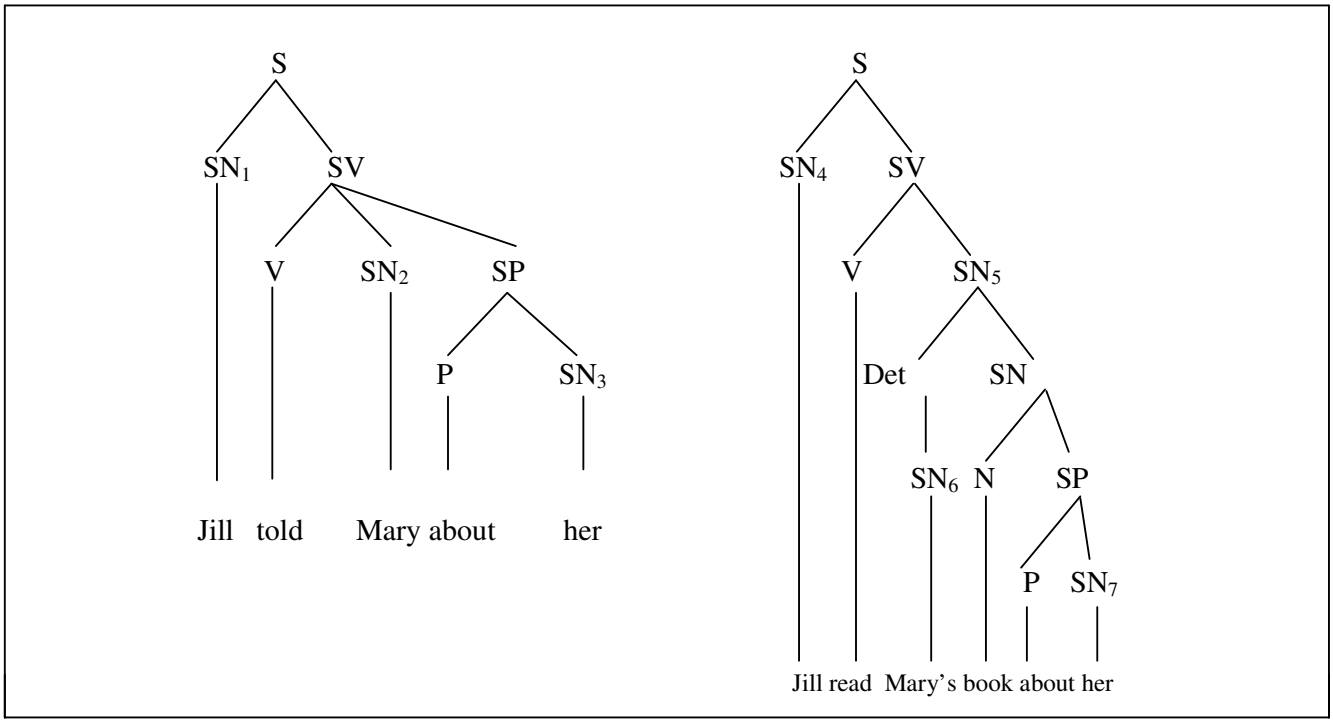
Com base nessa noção de domínios, Allen (1995) argumenta que, para que um pronome reflexivo seja co-referencial com um SN, ambos precisam fazer parte do mesmo domínio local. O inverso vale para os pronomes não-reflexivos, que, para serem co-referentes com um SN, devem pertencer a domínios locais distintos. A agramaticalidade da frase (5) consiste no fato de *Jill* e *her* pertencerem ao mesmo domínio local sendo *her* um pronome não-reflexivo. Para que fossem co-referenciais, o domínio local de *Jill* precisaria ser diferente do domínio local de *her.* Já na frase (7), por causa da presença do pronome reflexivo, para que fosse possível *Jill* e *herself* serem co-referenciais, eles não poderiam estar em domínios locais diferentes.

Com a análise das frases (9) e (10), exemplifica-se a existência de situações em que se exige o uso específico de pronomes. De início, seria possível afirmar a restrição que SNs não pronominais pertencentes à mesma frase não podem ser co-referenciais. Porém, essa afirmação é muito geral e também é facilmente questionada. A frase *After Jill had been questioned for hours, Sue took the tired witness out to lunch* demonstra a fragilidade dessa restrição, já que o SN não pronominal *the tired witness* é co-referencial com *Jill.*

Outra noção linguística análoga, que também poderá ser útil nessa tarefa, é a de C-comando, proveniente da gramática chomskiana. De acordo com Allen (1995), um constituinte C c-comanda outro constituinte X se, e somente se, (i) C não dominar X e (ii) O primeiro nó da ramificação que dominar C também dominar X.

A figura 10 ilustra as relações de C-comando.

Figura 10 - Estruturas sintáticas que ilustram a noção de C-comando



Fonte: Adaptado de Allen, 1995, p. 368.

Como se observa na figura 10, na árvore sintática à esquerda, correspondente à frase *Jill told Mary about her*, o SN1 (*Jill)* é dominado por S e, dessa forma, *Jill* c-comanda o SN2 (*Mary)* e SN3 (*her)*. O SN2 é dominado pelo VP, que c-comanda O SN3. A árvore sintática à direita representa a frase *Jill read Mary's book about her* e nela observa-se que SN4 (*Jill*) é

dominado por S e c-comanda o SN5 (*Mary's book*), o SN6 (*Mary*) e O SN7 (*her*). Por sua vez, o SV domina o SN6, que c-comanda o SN7.

Com a noção de C-comando, torna-se possível re-elaborar a restrição sobre a co-referencialidade dos pronomes reflexivos, que passa a ser a seguinte: 1. A referência de um pronome reflexivo deve ser a de um SN dentro do mesmo domínio local e que o c-comande. 2. A referência de um pronome não reflexivo não pode ser a de um SN que o c-comande dentro do mesmo domínio local.

Para uma melhor ilustração das restrições sobre os pronomes reflexivos, considere as frases a seguir, retiradas de Allen (1995, p. 368): (i) *Jill told Mary about her* e (ii) *Jill read Mary's book about her.* Na frase (i), não há possibilidade de o pronome *her* ter como referente *Mary* e nem *Jill*, isso porque o pronome é c-comandado por eles e faz parte do nó S que os contém. Se a frase fosse modificada para *Jill told Mary about herself*, que inclui um pronome reflexivo, então seriam possíveis duas alternativas de referente para ele, *Jill* ou *Mary*. Em outra modificação da frase, *Jill told herself about Mary*, a única opção de referente para o pronome reflexivo é *Mary*, porque *herself* não o c-comanda. Na frase (ii), *Jill* pode ser o referente do pronome *her* porque, apesar de *Jill* c-comandar *her*, não pertence ao mesmo domínio local do SN7 (*her).*

Uma explicação satisfatória ainda é necessária para as frases (3), (4), (9) e (10) e, para isso, uma restrição sobre SNs que não são pronominais será utilizada. A restrição é formulada tal qual se segue: "um SN não-pronominal não pode ser co-referente com um SN que o c-comande"[25] (ALLEN, 1995, p. 368, tradução nossa). Isso explica a agramaticalidade da frase (3), já que *He* c-comanda *Jack*. Também a frase (9), *Jack* c-comanda *the tired man,* o que justifica a frase não estar bem formada. Ressalta-se que, na frase *After Jill had been questioned for hours, Sue took the tired witness for lunch*, mencionada acima como exemplo da fragilidade da restrição sobre a co-referencialidade de SNs não pronominais pertencentes à mesma frase, a relação de c-comando entre os dois SNs não pronominais (*Jill* e *the tired witness*) não se verifica, permitindo que eles sejam co-referenciais.

As técnicas mais simples de se identificar a concordância de número, pessoa e gênero para a resolução de anáforas podem ser complementadas por outras heurísticas. A forma mais garantida de resolução das ambiguidades anafóricas é fazer com que o tradutor automático compreenda melhor o texto em análise. Isso é possível quando o sistema é alimentado com

---

[25] Texto original: *a nonpronominal NP cannot co-refer with an NP that C-commands it* (ALLEN, 1995, p. 368).

informações linguísticas mais refinadas, incorporadas a bases de conhecimento ontológico e com acesso à modelagem do contexto.

Para interpretar estruturas com quantificadores que podem originar ambiguidade do tipo **B4 Ambiguidade de escopo da quantificação**, deve-se considerar restrições e preferências, tais como as expostas em Allen (1995). Para analisar o fenômeno do escopo, usa-se também a noção de "domínio local". Se dois constituintes pertencem ao mesmo domínio local, diz-se que são "horizontalmente relacionados", situação que se observa em relação à SN1 e SN2, na figura 9. Para serem "verticalmente relacionados", é preciso que um deles seja o constituinte dominante do domínio local do outro. Allen (1995) entende que essa terminologia pode ser aplicada da mesma forma aos quantificadores. Assim, dois quantificadores serão considerados horizontalmente relacionados se os constituintes aos quais eles fazem partem também forem horizontalmente relacionados. A aplicação dessa terminologia aos quantificadores permite que o fenômeno do escopo seja, então, dividido em "escopo horizontal" e "escopo vertical".

Esse autor argumenta que a técnica mais comum para se tratar o escopo horizontal é a comparação de cada par de operadores e depois a verificação da possibilidade de se estabelecer um padrão em que um operador tenha o escopo maior do que o do outro. A existência de um padrão revela a "força" de um quantificador: aquele que é mais forte, em muitos casos, ultrapassa o alcance do mais fraco. O autor argumenta que a hierarquia de força de quantificadores a seguir foi proposta para solucionar essa questão: *each* > *every* > *all, some, several, a*.

Entretanto, por não considerar as relações estruturais que se estabelecem entre os quantificadores, essas preferências não se confirmam em todas as situações, ou seja, as preferências de escopo são influenciadas pela posição do quantificador na frase. Isso significa dizer que, por exemplo, pode haver variação nas preferências dependendo do quantificador que está exercendo a função sintática de sujeito.

Dessa forma, determinou-se a seguinte hierarquia das preferências gerais para o escopo com base na posição dos quantificadores:[26] *constituintes antepostos* > *sujeitos* > *adjuntos adverbiais pospostos* > *objetos diretos/indiretos.*

As argumentações de Badlwin, Kordoni e Villavicencio (2009) fornecem pistas para a resolução do tipo **D. Ambiguidade temática.** Segundo os autores, quando se trata de

---

[26] Allen (1995, p. 355) aponta que alguns quantificadores em particular não aceitam essas preferências.

preposições, deve-se considerar dois aspectos, já descritos na p.71: a seleção e a valência. Isso porque as preposições intransitivas tendem a não ser ambíguas quanto ao papel temático que introduzem. Um inventário dos papeis temáticos possíveis para as preposições mais frequentes também seria fonte de informações importantes para que os problemas com esse tipo de ambiguidade fossem resolvidos ou amenizados.

Porque as estratégias de resolução dos diferentes tipos de ambiguidades descritos apresentam limitações, existe ainda a possibilidade oferecida pela técnica da interação: o tradutor humano é incluído no processo da TA para auxiliar o sistema com a pré-edição ou com a pós-edição.

Outra opção é limitar o domínio dos textos de entrada e fornecer ao sistema a maior quantidade de conhecimentos sobre o domínio em questão.

Para finalizar esta seção, o Quadro 9 a seguir resume as estratégias de resolução de ambiguidades, que foram discutidas na seção, em função dos tipos e subtipos já sistematizados. Na primeira coluna, sinaliza-se o nível linguístico em que as ambiguidades se manifestam; na segunda, apresentam-se os tipos de ambiguidade, sendo que os subtipos estão especificados na terceira coluna e, por fim, na última coluna, reúnem-se as estratégias de resolução discutidas.

Quadro 9 - Resumo das estratégias de resolução de ambiguidades discutidas

TIPOS DE AMBIGUIDADE				ESTRATÉGIA(S) DE RESOLUÇÃO
A.	**LEXICAL**			
		1. Ambiguidade devido à homonímia ou polissemia		Informação semântica (associação semântica entre os significados dos itens lexicais), sintática (restrições selecionais) e sobre o co-texto. Utilização de *scripts*.
		2. Ambiguida-de categorial		Informação sintática e sobre o co-texto. Utilização de *scripts*.
		3. Ambiguida-de de transferência		Informação semântica (associação semântica entre os significados dos itens lexicais) e sobre o co-texto.
B.	**ESTRUTURAL**			
		1. Ambigui-dade de fixação de constituintes		
			1.1 Ligação de um	Informação sintática.

			sintagma preposicional a mais de um sintagma nominal ou verbal	Princípios gerais para a fixação de novos constituintes e preferências lexicais.
			1.2 Ligação de oração relativa a mais de um sintagma nominal	Princípios gerais para fixação de novos constituintes.
			1.3 Ligação de sintagma preposicional a um adjetivo predicativo	Princípios gerais para fixação de novos constituintes.
		3. Ambiguidade analítica		
			3.4 Diferenciação entre particípio presente e substantivo	Informações sintáticas sobre itens lexicais que podem ocorrer em ambientes sintáticos compostos de substantivos e verbos.
			3.5 Delimitação da extensão do sintagma nominal	Informações sintáticas sobre pré-modificação/pós-modificação. Possibilidade de reformulação do trecho ambíguo como um sintagma pós-modificado por um SP.
		4. Ambiguidade do escopo da quantificação		Informação sintática e semântica (hierarquia dos quantificadores).
C.	**ANAFÓRICA OU REFERENCIAL**			Restrições e preferências sintáticas (concordância de gênero e número; paralelismo sintático) e semânticas.
D.	**TEMÁTICA**			Informações sobre seleção lexical e valência das preposições.

Fonte: Elaboração própria

4.3 Explorações do impacto das ambiguidades linguísticas na TA inglês-português

Nesta subseção, desenvolve-se a parte **exploratória** do trabalho, analisando-se as TAs das frases selecionadas no *corpus* COMPARA em função dos tipos de ambiguidade descritos na subseção 3.2[27]. Ressalta-se que nem todas as frases apresentadas nesta subseção fazem parte do Quadro 8 (p. 87).

Os objetivos aqui são: (i) avaliar se a adequação da TA, do inglês para o português, foi afetada pela presença desses tipos de ambiguidade, (ii) sugerir elementos que corrijam ou melhorem a adequação da TA e (iii) confrontar as TAs das frases extraídas do *corpus* que foram utilizadas para a exemplificação dos tipos e subtipos de ambiguidade na subseção 3.2 em dois momentos: março-setembro de 2010[28] e março de 2011. O confronto das TAs foi motivado pelo fato de o *Google Tradutor* apresentar comportamento idiossincrático (apresentar traduções distintas para a mesma frase em momentos diferentes e com diferentes graus de adequação), por talvez ser constantemente atualizado (ver seção 2.1.1).

O texto da discussão segue o seguinte formato:

- **Nome do tipo de ambiguidade**, precedida por sua letra e numeração identificadoras;
- **Descrição da TA** gerada pelo *Google Tradutor* (para as ambiguidades lexicais) e **descrição de uma instância** do tipo de ambiguidade (para os demais tipos de ambiguidade);
- **Frase do inglês** extraída do *corpus* COMPARA e adaptada para a discussão[29];

---

[27] Ressalta-se que nem todas as frases apresentadas nesta subseção fazem parte do Quadro 8 (p. 77). Quando possível, serão apresentados, nesta subseção, outros exemplos diferentes daqueles que compõem o quadro, já que para elaborá-lo escolharam-se apenas uma frase para ilustração.

[28] A facilidade e a consequente rapidez da reavaliação das frases novamente submetidas ao *Google Tradutor* em março de 2011 explicam-se pelo fato de não ser mais necessária a pesquisa no *corpus* COMPARA, etapa essencial no momento da extração das frases do inglês e do português.

[29] Quando a frase originalmente extraída do *corpus* contém elementos (itens léxicos, expressões e sintagmas) que não são relevantes para o teste de desempenho do tradutor no que diz respeito ao tipo de ambiguidade em discussão, ela foi reformulada, delas se extraindo esses elementos com o objetivo de melhor isolar os elementos que podem ser causas do desempenho inadequado observado.

- **Leituras** possíveis (isto é, interpretações possíveis) da parte da frase relevante para a análise;
- **Traduções** (tradutor humano,TH, e tradução automática, TA);
- **Avaliação**;
- **Confronto** (das TAs).

4.3.1 (A) Ambiguidades lexicais

Retomam-se, nesta subseção, os três tipos de ambiguidades lexicais sistematizados no Quadro 6, subseção 3.2.

4.3.1.1 (A1) Ambiguidade por polissemia/homonímia

Descrição da TA: bank

O dicionário do *Google Tradutor* registra, para o substantivo *bank*, as seguintes opções de tradução: *banco*, *beirada*, *aterro*, *casa bancária*, *dique*, *margem*, *boca de mina*, *rampa*, *ladeira* e *ribanceira*. Diante dessa variedade, o problema de tradução é a escolha do item léxico apropriado, escolha que poder ser orientada com o auxílio de preferências contextuais imediatas e cálculos estatísticos.

Frase do inglês

*1) He walked up the **bank** of the little brook, head bent, eyes fixed upon the mud and the water*[30].

Leituras

(A) Ele caminhou ao longo da **margem** do riacho.
(B) Ele caminhou ao longo do **banco** do riacho.
(C) Ele caminhou ao longo da **casa bancária** do riacho.

Traduções

---

[30] Frase extraída do *corpus: He began to walk slowly up the rough **bank** of the little brook, head bent, eyes fixed upon the mud and the water.*

TH: Começou a caminhar lentamente ao longo da **margem** do riacho, cabisbaixo, com os olhos fixos na lama e na água.

TA: Ele começou a caminhar lentamente até a **margem** do pequeno riacho, cabeça baixa, olhos fixos na lama e na água.

Frase do inglês

*2) On the other **bank** you only hear the stream, the water absorbs the music*[31].

Leituras

(A) Na outra **margem** ouve-se apenas o riacho, a água absorve as músicas.

(B) No outro **banco** ouve-se apenas o riacho, a água absorve as músicas.

(C) Na outra **casa bancária** ouve-se apenas o riacho, a água absorve as músicas.

Traduções

TH: ...da outra **margem** ouve-se apenas o riacho, a água absorvendo as músicas.

TA: Na outra **margem**, apenas ouvirá o córrego, a água absorve a música.

Avaliação

Nas frases (1) e (2), os substantivos vizinhos (*brook, bridge, stream, mud* e *water*) fornecem o co-texto mínimo a partir do qual elementos centrais do sentido desses substantivos direcionam a interpretação da ambiguidade de *bank,* por meio de associações semânticas.

Confronto

A TA da frase *He began to walk slowly up the rough **bank** of the little brook, head bent, eyes fixed upon the mud and the water*, no primeiro momento do teste, é inadequada: *Ele começou a caminhar lentamente até o **banco** em bruto do pequeno riacho, cabeça baixa, olhos fixos na lama e na água*. Esse fato se explica devido à presença do adjetivo *rough*, pré-modificador de *bank*; quando se elimina esse adjetivo a TA passa a ser adequada.

A TA da frase *I cross the bridge and on the other **bank** you only hear the stream, the water absorbs the music* é *Atravesso a ponte e no outro **banco** você só ouvir o riacho, a água*

---

[31] Frase extraída do *corpus: I cross the bridge and on the other **bank** you only hear the stream, the water absorbs the music.*

absorve a música. Para essa frase, o problema é a presença do substantivo *bridge*, que, quando substituído por *street*, por exemplo, torna a TA adequada.

Descrição da TA: paper

O dicionário do *Google Tradutor* registra, para *paper*, os seguintes substantivos: *papel*, *papel moeda*, *folha de papel*, *jornal periódico*, *letra de câmbio*, *nota de banco* e *promissória*. As opções *trabalho*, *documento* e *artigo* não são apresentadas quando apenas o dicionário do sistema é consultado. Nota-se, porém, observando-se os resultados, que o sistema usa essas opções para gerar a TA da frase *Rodney gives a sickly grin: «No, I think I'll stay in and look over my **paper** for tomorrow*, cuja TA é *Rodney dá um sorriso doentio: Não, eu acho que vou ficar e olhar por cima do meu **trabalho** para amanhã* e a TH*, Rodney brinda-o com um sorriso amarelo: -- Não, acho que vou ficar e dar uma última revisão à minha **palestra*** (sic) *de amanhã*.

Frase do inglês

*3) The newspaper boy is here, or there is no **paper** today because of a strike*[32].

Leituras

(A) ... talvez hoje não haja **jornal** por causa de uma greve qualquer.

(B) ... talvez hoje não haja **papel** por causa de uma greve qualquer.

(C) ... talvez hoje não haja **artigo científico** por causa de uma greve qualquer.

Traduções

TH: ... talvez hoje não haja **jornais** por causa de uma greve qualquer.

TA: O menino jornal está aqui, ou não há o **jornal** de hoje devido a uma greve.

Frase do inglês

*4) The stewardess offered him the morning **paper***[33].

Leituras

(A) A comissária oferece-lhe o **jornal** da manhã

---

[32] Frase extraída do *corpus*: *The newspaper boy is late, or perhaps there is no **paper** today because of a strike.*

[33] Frase extraída do *corpus: A stewardess offers him the morning **paper**.*

(B) A comissária oferece-lhe a **papel** da manhã.

(C) A comissária oferece-lhe o **artigo científico** da manhã

Traduções

TH: A hospedeira (sic) oferece-lhe o **jornal** da manhã.

TA: A comissária de bordo ofereceu-lhe o **jornal** da manhã.

Frase do inglês

*5) She reads about that in the **paper** every day*[34].

Leituras

(A) Ela lê sobre aquilo no **jornal** todos os dias.

(B) Ela lê sobre aquilo na **pepel** todos os dias.

(C) Ela lê sobre aquilo no **artigo científico** todos os dias.

Traduções

TH: Todos os dias se lêem coisas dessas nos **jornais**.

TA: Ela lê sobre isso no **jornal** todos os dias.

Avaliação

Nas frases (3), (4) e (5), o sistema gerou TAs adequadas. Nas frases (3) e (4), mesmo sem se saber se o sistema utilizou informações contextuais, as expressões *newspaper boy* e *morning paper*, respectivamente, são pistas contextuais importantes para a escolha de *jornal* e não *papel*. Já, na frase (5), no entanto, não há elementos no co-texto que direcionem a interpretação, já que *jornal*, *papel* e *artigo* são complementos licenciados por *ler*. É preciso considerar um co-texto mais amplo ou até a situação.

Confronto

Observa-se, que, no primeiro momento, as TAs das frases (4) e (5), sem adaptações, e da frase (3) não foram adequadas, conforme ilustram estas respectivas TAs: *A aeromoça oferece-lhe o **papel** de manhã*, *Você pode ler sobre esse tipo de coisa no **papel** todos os dias*, e *O menino do jornal é tarde, ou talvez não há **papel** hoje por causa de uma greve*. Esse fato

---

[34] Frase extraída do *corpus*: *You read about that sort of thing in the **paper** every day.*

mostra que ou informações contextuais foram acrescidas ao sistema ou novas estatísticas privilegiaram as escolhas ou ainda as adaptações feitas simplificam a análise pelo sistema e permitem-lhe elaborar os cálculos estatísticos com maior chance de acerto.

Descrição da TA: seal

O dicionário do *Google Tradutor* registra, para o substantivo *seal,* presente na frase (6), registra as seguintes opções no português: *selo, sinal, sinete, lacre, cunho, chancela, carimbo, marca, foca*, e *pele* ou *couro de foca*.

Frase do inglês

6) He put the ***seal*** *under one arm, and put his thumb in*[35].

Leituras

(A) Hughie colocou a **foca** debaixo do braço e colocou o polegar na boca.
(B) Hughie colocou o **selo** debaixo do braço e colocou o polegar na boca.
(C) Hughie colocou o **lacre** debaixo do braço e colocou o polegar na boca.

Traduções

TH: Hughie meteu a **foca** debaixo do braço e enfiou o polegar na boca.
TA: Ele colocou o **selo** embaixo do braço, e colocar o polegar dentro.

Avaliação

Na frase (6), mesmo com a substituição de *tucked* por *put* e *Huggie* por *He*, o sistema manteve a TA *selo*. Essa ambiguidade poderia ser solucionada com a aplicação de conhecimento de mundo, que permitiria decidir por *seal* (=foca) no co-texto de *put his thumb in*.

Confronto

Não se constatou tradução distinta.

Descrição da TA: nail

---

[35] Frase extraída do *corpus*: *Hughie tucked the seal comfortably under one arm, and put his thumb in.*

Para o substantivo *nail*, o dicionário do *Google Tradutor* registra: *unha, prego, tacha, garra* e *cravo*. Aqui, como já se comentou, o problema é a escolha do substantivo adequado no contexto.

Frase do inglês

*7) He hung it on the **nail** where the previous occupant of the hut had hung her hoe*[36].

Leituras

(A) Ele pendurou-a no **prego** onde a ocupante anterior da cabana havia pendurado a sua enxada.

(B) Ele pendurou-a no **unha** onde a ocupante anterior da cabana havia pendurado a sua enxada.

(C) Ele pendurou-a no **cravo** onde a ocupante anterior da cabana havia pendurado a sua enxada.

Traduções

TH: ... pendurou-a no **preg**o onde a anterior ocupante da cabana havia arrumado (sic) a sua enxada.

TA: Pendurou-a no **prego**, onde o ocupante anterior da cabana tinha pendurado seus enxada.

Avaliação

O segmento *Handed it back*, separado da frase seguinte *he hung...* pelo ponto e vírgula, desviou o sistema para a TA inadequada. A simples substituição do ponto e vírgula pelo ponto final resolve a TA: *Devolveu. pendurou-o no prego que o ocupante anterior da cabana tinha pendurado sua enxada*. Ao se substituir *hung* por *painted*, a TA passa a: *Devolveu. Pintou-a sobre a unha, onde o ocupante anterior da cabana tinha pendurado sua enxada*. Embora apresentando comportamento idiossincrático, pode-se supor que o sistema "calcula" informação co-textual.

Confronto

---

[36] Frase extraída do *corpus*: *Handed it back; he hung it on the **nail** where the previous occupant of the hut had hung her hoe.*

No primeiro momento, a TA foi inadequada porque não se consideraram as possibilidades testadas na avaliação acima.

Descrição da TA: bath

As frases (8) e (9) incluem o substantivo *bath*, para o qual o sistema registra as seguintes opções de tradução: *banho, banheira* e *banheiro*.

Frase do inglês

*8) Robert knelt by the **bath**.*[37]

Leituras

(A) Robert ajoelhou-se perto da **banheira.**

(B) Robert ajoelhou-se perto do **banho.**

(C) Robert ajoelhou-se perto do **banheiro.**

Traduções

TH: Robert ... ajoelhou-se junto da **banheira**.

TA: Robert se ajoelhou ao lado da **banheira**.

Frase do inglês

*9) She sat in the **bath**[38].*

Leituras

(A) Ela sentou-se na **banheira**.

(B) Ela sentou-se no **banho**.

(C) Ela sentou-se no **banheiro**.

Traduções

TH: Ela sentou-se na **banheira...**

TA: Sentou-se na **banheira**.

Avaliação

Nas frases (8) e (9), os verbos *knelt* e *sat* e suas regências, as preposições *by* e *in*, licenciam um Complemento Locativo, informação aparentemente considerada pelo sistema e relevante para o resultado adequado da TA.

Confronto

---

[37] Frase extraída do *corpus*: *Robert put his toothbrush down and then he knelt by the bath.*

[38] Frase extraída do *corpus*: *She sat in the **bath** soaping her neck.*

A frase (8) manteve a adequação da TA. Já a TA da frase (9) verificou-se adequada neste segundo momento como decorrência da simplificação, que eliminou a possibilidade de o sistema analisar *in the bath* como o Adjunto Adverbial de Lugar sinalizado nesta paráfrase entre os colchetes: *[No banho], ela se sentou*.[39]

4.3.1.2 (A2) Ambiguidade categorial

Descrição da TA: comb

O dicionário do *Google Tradutor* registra, para o verbo *comb*, dentre outras, estas opções de tradução: *pentear*, *vasculhar*, *alisar*, *rastelar* e *rebentar*. Fato que faz com que o sistema calcule escolhas.

Frase do inglês

10) If after my bath I don't comb it right away[40].

Leituras

(A) Se, depois do banho, eu não o **penteio** logo.
(B) *Se, depois do banho, eu não o **pente** logo.
(C) Se, depois do banho, eu não o **aliso** logo.

Traduções

TH: E, quando (sic) depois do banho, não me (sic) **penteio** logo...
TA: Se depois do banho eu não **penteá-lo** imediatamente

Avaliação

A TA da frase (10) é adequada, mostrando que o sistema reconhece evidências sintáticas, como o pronome *I* e o auxiliar negativo *don't,* que exigem a presença de um verbo como elemento que os segue, fato que ressalta que informações provenientes da sintaxe são

---

[39] TA inadequada, com o Adjunto Adverbial de Lugar sinalizado entre os colchetes: *Ela sentou-se [no banho] ensaboar o pescoço*.

[40] Frase extraída do *corpus: And if after my bath I don't* ***comb*** *it out right away, or if I let one day go by without applying oil...*

essenciais para o desempenho adequado dos analisadores gramaticais, solucionando parte das ambiguidades linguísticas.

Confronto

No primeiro momento, a TA da frase *And if after my bath I don't* ***comb*** *it out right away, or if I let one day go by without applying oil* foi indadequada: *E se depois do banho eu não* ***pente****-lo imediatamente, ou se eu deixar um dia passar sem aplicação de óleo*, resultado que revela comportamento idiossincrático do sistema.

Descrição da TA: fine

O item lexical *fine* pertence a diferentes categorias sintáticas: substantivo, verbo, adjetivo e advérbio, fato que exige escolhas categoriais pelo sistema. As opções de tradução fornecidas pelo dicionário do *Google Tradutor* são: como substantivo, *multa, penalidade, fim,* etc.; como verbo, *multar, refinar,* etc.; como adjetivo, *fino, bom, excelente, belo, agradável, bonito,* etc.

Frase do inglês

11) A ***fine*** *reward for trying to preserve academic standards.*

Leituras

(A) Uma **bela** recompensa por tentar manter o nível acadêmico.
(B) *Uma **multa** recompensa por tentar manter o nível acadêmico.
(C) Uma **boa** recompensa por tentar manter o nível acadêmico.

Traduções

TH: Uma **gloriosa** recompensa por tentar manter o nível acadêmico.
TA: Uma recompensa **muito bem** para tentar preservar os padrões acadêmicos.

Frase do inglês

12) It was a ***fine****, soft rain, and the air was warm and balmy*[41].

Leitura

---

[41] Frase extraída do *corpus*: *It was, however, a* ***fine****, soft rain, and the air was warm and balmy.*

(A) Era uma chuva **boa**, suave, e o ar estava morno e perfumado.

(B) Era uma chuva **multa**, suave, e o ar estava morno e perfumado.

(C) Era uma **bela** chuva, suave, e o ar estava morno e perfumado.

Traduções

TH: Era, no entanto, uma chuva **boa** (sic), suave, e o ar estava morno e perfumado.

TA: Foi uma chuva **fina** e macia, e o ar estava quente e ameno.

Avaliação

Na TA das frases (11) e (12), *fine* é adjetivo nas funções atributiva e predicativa, respectivamente. Em (11), a TA é inadequada, porque o sistema não associa o adjetivo atributivo *fine* como o valor de um atributo de *reward* (para *car*, por exemplo, esse valor ("belo" do atributo "qualidade") está previsto, conforme demonstra esta TA: *a fine car* e *um belo carro*, mas *a fine reward*, **uma recompensa multa*). Na função predicativa, o adjetivo fine não causa problema para a TA: *The reward is fine* e *A recompensa é boa*. Mas note-se a avaliação abaixo, em "Confronto".

Confronto

As TAs inadequadas obtidas no primeiro momento, em que o sistema optou por *multa* nas TAs das frases (11), *Uma recompensa **multa** para tentar preservar os padrões acadêmicos*, e (12), *Foi, no entanto, uma **multa**, a chuva suave, e o ar estava quente e ameno*, parecem ser causadas pela análise equivocada tanto de *reward* quanto da estrutura sintática do segmento *It was, however, a fine, soft rain*: no primeiro caso, *reward* é analisado como verbo; no segundo caso, ao invés desta análise, *It was, however,* [*a fine, soft rain*], em que *fine* é adjetivo modificador de *rain*, o sistema seleciona esta [*It was, however, a fine*]*, soft rain*, em que *fine* é um substantivo na função predicativa.

Descrição da TA: ring

O item léxico *ring* pode pertencer às categorias verbal ou nominal. O dicionário do *Google Tradutor* registra, para o verbo *ring*, as seguintes opções de tradução: *badalar*, *cercar*, *circundar* e *rodear*. Não registra: *tocar*, *soar* e *telefonar*. Surpreendentemente, observa-se que essas opções ausente são usadas nas TAs do sistema: para a frase *I hear the phone **ring** once,*

twice, five times, the phone in an old person's house, o sistema gera a TA: *Eu ouvi o telefone* ***tocar*** *uma, duas, cinco vezes o telefone na casa de uma pessoa velha.*

Frase do inglês

13) I ***ring*** *the Customs House*[42].

Leituras

(A) **Ligo** para a Alfândega.

(B) ***Anel** para a Alfândega.

(C) **Cerco** para a Alfândega.

Traduções

TH: **Ligo** para a Alfândega...

TA: Eu **toco** a Alfândega.

Frase do inglês

14) Shall I ***ring*** *Mary and ask her*[43]*?*

Leituras

(A) **Telefono** à Maria para lhe perguntar?

(B) ***Anel** à Maria para lhe perguntar?

(C) ***Cerco** à Maria para lhe perguntar?

Traduções

TH: **Telefono** à Phyllis Cameron para lhe perguntar?

TA: Devo **anel** a Maria e pergunta a ela?

Frase do inglês

15) The bell for the break ***would ring***[44].

---

[42] Frase extraída do *corpus: I* ***ring*** *the Customs House, a man answers in a gruff voice, I ask for her.*

[43] Frase extraída do *corpus*: *Shall I* ***ring*** *Phyllis Cameron and ask her?*

[44] Frase original extraída do *corpus: The bell for the break would* ***ring*** *and we would all simultaneously slam shut our books on Christian doctrine.*

Leituras

(A) O sino do intervalo **tocava.**

(B) *O sino do recreio **anel.**

(C) O sino do recreio **cercava.**

Traduções

TH: A sineta do recreio **tocava**...

TA: O sino para o intervalo **tocaria.**

Avaliação

A TA da frase (13) e (14), evidencia que a ambiguidade originada por *ring* não foi obstáculo para o sistema, já que a TA, embora sintaticamente imprecisa, não é semanticamente totalmente inadequada, porque se tem, em português esta opção: *Eu **toco**/ligo/telefono p'ra você/FCL/Rodoviária mais tarde*.

Fato interessante é constatar que, na frase (14), se *Phyllis Cameron* for substituído por *her* ou *him*, a TA é adequada: *Devo telefonar-lhe e pedir-lhe?* e *Devo ligar para ele e perguntar-lhe?* Contudo, a TA de *Shall I ring you and ask her?* é idiossincraticamente inapropriada: *Devo anel de você e pergunta a ela?*

Confronto

No primeiro momento do teste, as TAs das frases do *corpus* foram todas inadequadas. A TA da frase (13) foi: *Eu **anel** da Alfândega, um homem responde com uma voz rouca, eu peço para ela;* a TA da frase (14): *Devo **anel** a Phyllis Cameron e pergunta a ela?*; a TA da frase (15): *O sino para o intervalo seria de **anel** e que todos nós se fechar nossos livros sobre a doutrina Cristã.* Ressalta-se que, para essas TAs, a utilização de informações sintáticas contribuiriam para a resolução desse tipo de ambiguidade: a presença do pronome *I* do auxiliar *would*.

Descrição da TA: thought

O dicionário do *Google Tradutor* registra, para *thought*, as seguintes opções: os substantivos *pensamento, reflexão, atenção,* etc., e o "adjetivo" (sic) *pensado*.

Frase do inglês

*16) Sophy **thought**, briefly, about honesty*[45].

Leituras

(A) Sophy **pensou** por instantes na sinceridade.

(B) *Sophy **pensamento** por instantes na sinceridade.

(C) *Sophy **ideia** por instantes na sinceridade.

Traduções

TH: ... Sophy **pensou** por instantes na sinceridade.

TA: Sophy **pensamento**, brevemente, sobre a honestidade.

Frase do inglês

*17) He hugs the **thought** to himself*[46].

Leituras

(A) Ele abraça a **ideia**.

(B) *Ele abraça o **pensou.**

(C) Ele abraça o **pensamento.**

Traduções

TH: Abraça a **idéia...**

TA: Ele abraça o **pensamento** para si mesmo.

Avaliação

A TA da frase (16) demonstra que o sistema, por não reconhecer *Sophy* como um nome próprio, gera a TA inadequada, porque calcula que *thought* é núcleo de um SN e não de um SV. Ao se substituir *Sophy* por *Mary*, entretanto, a TA passa a adequada: *Maria **pensou**,*

---

[45] Frase extraída do *corpus: Leaning there, against the warm, rough stone wall, Sophy **thought**, briefly, about honesty.*

[46] Frase extraída do *corpus: He hugs the **thought** to himself with guilty glee.*

brevemente, sobre a honestidade. Essa constatação indica que a manutenção de um banco de nomes próprios é essencial para a resolução de ambiguidades.

A TA da frase (17) é adequada porque permite ao sistema recorrer à marca sintática *the*, que sinaliza a presença de um SN.

Confronto

A TA da frase (16) continuou inadequada neste segundo momento. Já a TA da frase (17), no primeiro momento, foi inadequada: *Ele abraça o* ***pensou*** *com alegria culpado*. Esse teste, mais uma vez, sugere que o *Google Tradutor* vale-se de estatísticas e *pattern matching* e não (tanto) de conhecimentos na sua tarefa de TA.

Descrição da TA: steps

O dicionário do *Google Tradutor* registra, para *steps*, as seguintes opções de tradução: os substantivos plurais *etapas* e *passos*; o verbo *step* na terceira pessoa do singular do presente simples (ele/ ela) *pisa*.

Frase do inglês

18) Leslie ***steps*** *forward with a smile*[47].

Leituras

(A) Leslie **avança** com um sorriso nos lábios.

(B) *Leslie **passos** com um sorriso nos lábios.

(C) Leslie **pisa** com um sorriso nos lábios.

Traduções

TH: Leslie **avança** com um sorriso nos lábios...

TA: Leslie **passos** para a frente com um sorriso.

Frase do inglês

19) She ***steps*** *toward me.*

---

[47] Frase extraída do *corpus: Leslie* ***steps*** *forward with a smile, introduces himself to the couple, and inspects their tickets and passports.*

Leituras

(A) **Avança** para mim.

(B) ***Passos** para mim.

(C) ***Pisa** para mim.

Traduções

TH: **Avança** para mim.

TA: Ela **passos** em direção a mim.

Avaliação

As TAs das frases (18) e (19) são inadequadas devido à análise sintática imprópria: *steps* deveria ser analisado como V (núcleo de SV) e não N (núcleo de SN). A TA da frase (18), com a substituição do nome próprio *Leslie* por *Mary*, não passa a adequada. Contudo, se o nome próprio for substituído pelo pronome *she*, a TA passa a ser adequada: *Ela **avança** com um sorriso*. Entretanto, trata-se de apenas uma estrutura *ad hoc*, armazenada no sistema, porque, ao se substituir *Mary* por *he*, you ou *it*, obtêm-se os seguintes resultados, respectivamente: *Ele dá um passo a frente com um sorriso*, *Você passo à frente com um sorriso* e *É um passo à frente com um sorriso*. Um analisador sintático poderia resolver essas inadequações.

Confronto

Não se constatou tradução distinta.

Em suma, das 10 frases discutidas para ilustrar a **Ambiguidade categorial**, 5 frases obtiveram TAs adequadas e 5 frases mantiveram as TAs inadequadas no segundo momento. Destas 5, 3 são frases que contêm nomes próprios e a substituição dos nomes próprios dessas frases por nomes próprios mais comuns ou colocações com pronomes mais frequentes resulta em TAs apropriadas, o que reforça a alta dependência do *Google Tradutor* de cálculos estatísticos e comparação de padrões (“pattern matching”).

4.3.1.3 (A3) Ambiguidade de transferência

Descrição da TA: wall

O dicionário do *Google Tradutor* registra, para o substantivo *wall*, opções de tradução como *muro, muralha, parede, paredão* e *lado.*

Frase do inglês

20) He stopped walking, and sat down on the pavement, his back against the ***wall*** *behind it.*

Leituras

(A) Parou, sentou-se no passeio e encostou-se no **muro**.

(B) Parou, sentou-se no passeio e encostou-se na **parede**.

Traduções

TH: Parou, sentou-se no passeio e encostou-se a um **muro**.

TA: Ele parou de caminhar, e se sentou na calçada, de costas contra a **parede** por trás dele.

Avaliação

Mesmo se esperando a tradução *muro*, a TA não é inadequada, posto que a frase não fornece elementos contextuais suficientes para a escolha dessa opção.

Confronto

Não se constatou tradução distinta.

4.3.2 (B) Ambiguidades estruturais

Dos 20 tipos de ambiguidade estrutural, serão discutidos estes 12[48]:

- B1.1 Ligação de sintagma preposicional a mais de um sintagma nominal ou verbal;
- B1.2 Ligação de oração relativa a mais de um sintagma nominal;
- B1.3 Ligação de sintagma preposicional a adjetivo predicativo;
- B1.5 Ligação de advérbio como modificador de sintagma verbal ou de frase;
- B1.6 Ligação de particípio ao sujeito estrutural da frase ou à frase;
- B1.7 Ligação simultânea de adjunto adverbial a verbos de duas frases distintas;
- B3.1 Detecção de partículas;
- B3.3 Diferenciação entre particípio presente e adjetivo;
- B3.4 Diferenciação entre particípio presente e substantivo;
- B3.8 Interpretação ambígua de particípios e de orações adjetivas no final de frase;
- B3.11 Delimitação dos diferentes tipos de estrutura formados com esta sequência de elementos: *be ADJ to V*;
- B4 Ambiguidade de escopo da quantificação;

4.3.2.1 (B1.1) Ligação de sintagma preposicional a mais de um sintagma nominal ou verbal

Instância com base em "with..."

Na frase (21), o SP *with the Bible* pode ligar-se ao SN *I* ou ao núcleo *kitchen* do SN *the kitchen with the Bible.* Já na frase (22), o SP *with the patchwork quilt* pode estar

---

[48] Essa restrição justifica-se pelo fato de não terem sido encontradas, no *corpus* da pesquisa, frases para ilustrar o tipo **D.** Ambiguidade temática, bem como estes 8 tipos de ambiguidade estrutural restantes: **B1.4** Ligação de sintagma preposicional ou advérbio a posições pertencentes à oração principal ou à oração encaixada; **B2** Ambiguidade de localização e preenchimento de lacunas; **B3.2** Diferenciação entre um sintagma preposicional e um sintagma adjetivo resultante de uma operação de alçamento e apagamento do verbo ser/estar aplicada ao complemento do verbo; **B3.5** Delimitação da extensão do sintagma nominal; **B3.6** Diferenciação entre oração relativa reduzida e sintagma verbal da oração principal; **B3.7** Delimitação da estrutura de um sintagma nominal complexo;**B3.9** Diferenciação entre frases clivadas e frases do tipo sujeito-verbo-objeto; **B3.10** Diferenciação entre particípio passado e um sintagma verbal incompleto, resultando na ambiguidade entre pergunta e ordem.

potencialmente ligado ao núcleo *bed* do SN *the bed with the patchwork quilt* ou ao SV *shared the bed with the patchwork quilt*.

Frase do inglês

21) I came back into the kitchen ***with the Bible***[49].

Leituras

(A) **Eu** voltei para a cozinha **levando uma bíblia** (ligação ao SN I=eu).

(B) Eu voltei para a **cozinha em que havia uma bíblia** (ligação ao SN kitchen=cozinha).

Traduções

TH: Voltei para a cozinha **com a Bíblia...**

TA: Voltei para a cozinha **com a Bíblia.**

Frase do inglês

22) There was no sign of Robin in the room, no evidence of his having shared the bed ***with the patchwork quilt****.*

Leituras

(A) Não havia sinal de Robin no quarto, nenhum indício de que ele tivesse partilhado **a cama sobre a qual havia uma colcha de retalhos** (ligação com o SN the bed= a cama).

(B) Não havia sinal de Robin no quarto, nenhum indício de que ele tivesse **partilhado**, **com uma colcha de retalhos, a cama** (ligação com o SV having shared the bed=ter partilhado a cama).

Traduções

TH: Não havia sinal de Robin no quarto, nenhum indício de que ele tivesse partilhado a cama **com a coberta (sic)**[50] **de retalhos**.

---

[49] Frase extraída do *corpus: I came back into the kitchen* ***with the Bible****, rather pleased with myself, and read the verse out to Sally.*

[50] Trata-se de tradução para o português europeu.

TA: Não havia nenhum sinal de Robin na sala, sem evidência de ele ter partilhado a cama **com a colcha de retalhos**.

Avaliação

A TH e a TA da frase (21) são coincidentes. A ambiguidade descrita não compromete a adequação da TA, sendo mantida nas traduções. Vale lembrar que, nas duas traduções, prevalece a leitura (A), já que essa é a leitura pragmaticamente favorecida. Além disso, o principio do *Minimal attachment,* apresentado na seção 4.2 ("criar um número mínimo possível de nós na fase da análise sintática"), bloqueia a preferência pela leitura B.

A avaliação das TAs da frase (22) é análoga a das TAs da frase (21). Acrescenta-se que, se fosse levada em conta a preferência contextual do verbo *share*, isto é, a preposição *with* ("partilhamento") por ele regida, a leitura preferida seria a (B), que é a pragmaticamente desfavorecida, fato que indica que, nem sempre, o uso de preferências contextuais para a otimização da TA deve ser acionada.

Confronto

Não se constatou tradução distinta.

4.3.2.2 (B1.2) Ligação de oração relativa a mais de um sintagma nominal

Instância com base em "...that..."

As ambiguidades resultantes da possibilidade de ligação múltipla de uma oração relativa nem sempre comprometem a adequação da TA, porque esse tipo de ambiguidade pode ser transferido para a língua-alvo ou resolvido na fase de análise sintática, como ilustra a frase (23): nela, a oração relativa (em negrito) pode estar ligada ao núcleo do SN *a much-told* ***tale*** *of her Australian infancy* (23 A: "história profética"), a preferência sintática, ou ao núcleo do SN *her Australian* ***infancy*** (23 B: "infância profética"), uma preferência semântica. O mesmo ocorre na frase (24), em que a oração relativa pode estar ligada ao núcleo do SN ***life*** (24 A: "a vida se encaixava") ou ao núcleo do SN *all the* ***ways*** *of life* (24 B: "os modos se encaixavam").

Frase do inglês

23) There was a much-told tale of her Australian infancy ***that was held to be prophetic in this respect.***

Leituras

(A) Havia uma **história**, muito contada, da sua infância na Austrália **que era considerada profética a esse respeito.**

(B) Havia uma história, muito contada, da sua **infância** na Austrália **que era considerada profética a esse respeito.**

Traduções

TH: Havia uma história, muito contada, da sua infância na Austrália **que era considerada profética a este respeito**.

TA: Houve muito disse-conto de sua infância australiano **que foi detido para ser profética, a este respeito**.

Frase do inglês

24) He knew all the ways of life ***that fitted into that category****.*

Leituras

(A) Ele conhecia todos **os modos** de vida que se **encaixavam** nessa categoria.

(B) Ele conhecia todos os modos de **vida** que se **encaixava** nessa categoria.

Traduções

TH: Ele conhecia todos os modos de vida **que se encaixavam nessa categoria**.

TA: Ele conhecia todos os caminhos da vida, **que se enquadrou nessa categoria.**

Avaliação

Conforme constata-se, dada a frase (23), com a substituição de *held* por *understood*, traduzido por *entendido*, o sistema restringe-se a traduzir o item léxico descontextualizado, isto é, sem levar em conta possíveis antecedentes e, portanto, ignorando a concordância. Dado a frase (24), o sistema comporta-se de modo análogo: ele se restringe à tradução do item léxico *fitted*, sem considerar quaisquer antecedentes. Essas duas traduções demonstram o comportamento idiossincrático do *Google Tradutor*, que, por ser fortemente baseado em cálculos estatísticos, conforme os estudos aqui realizados permitem deduzir, não considera, por exemplo, esta heurística: "fazer a ligação entre o verbo da oração relativa e o núcleo do sintagma antecedente (em negrito)": ***tale*** e *held*, em (23); ***ways*** e *fitted*, em (24).

Confronto

A TA da frase (23) alterou a tradução do verbo da oração relativa: *realizada* agora se traduz por *detido*. A TA da frase (24) foi também alterada: o verbo, antes traduzido por *encaixava*, passou a ser traduzido por *enquadrou*. O problema apontado na avaliação, entretanto, permaneceu.

4.3.2.3 (B1.3) Ligação de sintagma preposicional a adjetivo predicativo

Instância com base em "...seemed..."

A ambiguidade de fixação de constituintes também se manifesta quando há as possibilidades de ligação do SP ou a um adjetivo predicativo ou a um SV. Na frase (25), o SP *to her* pode ligar-se tanto ao adjetivo predicativo ***ludicrous***, gerando a interpretação "João Diogo parecia agir **de forma risível** <u>em relação a ela</u>", quanto ao núcleo (em negrito) do SV ***seemed*** *ludicrous to her*, gerando a interpretação "João Diogo, <u>na opinião dela</u>, **parecia agir** de forma risível".

Frase do inglês

25) João Diogo seemed ***ludicrous to her***[51].

Leituras

(A) João Diogo parecia agir **de forma risível** em relação a ela (ligação com o adjetivo ludicrous=de forma risível).

(B) João Diogo, na opinião dela, **parecia agir** de forma risível (ligação com o núcleo do SV seemed=parecia agir).

Traduções

TH: ...**teria achado João Diogo tão risível** quanto da outra vez.

TA: João Diogo **parecia ridículo para ela.**

Avaliação

A TA da frase (25) mantém a ambiguidade de ligação no português, não deixando de ser semanticamente aceitável. Apenas informação enciclopédica poderia resolver esse tipo de ambiguidade.

Confronto

No primeiro momento, a TA da frase *Were it not for the anguish constricting his chest, and his fear, João Diogo would have seemed as* ***ludicrous to him*** *as on the other occasion* foi também adequada e manteve a ambiguidade na frase em português: *Se não fosse a angústia de constrição no peito, e seu medo,* ***João Diogo teria parecido ridículo*** *para ela.*

---

[51] Frase extraída do *corpus: Were it not for the anguish constricting his chest, and his fear, João Diogo would have seemed as* ***ludicrous to him*** *as on the other occasion.*

4.3.2.4 (B1.5) Ligação de advérbio como modificador de sintagma verbal ou de frase

Instância com base em "sadly"

O advérbio *sadly,* presente na frase (26), pode modificar o SV *complained of the outrageous décolletage of some immodest woman* ("recalamei...") ou a frase toda, funcionando como um operador discursivo ("a reclamação foi infeliz").

Frase do inglês

26) Then, ***sadly****, I complained of the outrageous décolletage of some immodest woman*[52].

Leituras

(A) Depois, queixei-me **tristemente** do decote de uma senhora imodesta.

(B) **Tristemente**, depois queixei-me do decote de uma senhora imodesta.

Traduções

TH: Depois, **tristemente**, como um moralista magoado, queixei-me do nédio decote de uma senhora imodesta...

TA: Então, **infelizmente**, eu reclamei do decote absurdo de uma mulher sem recato.

Avaliação

A TA manteve a ambiguidade no português, colocando o advérbio entre vírgulas, o que torna a TA sintática e pragmaticmente aceitável. A escolha lexical, entretanto, não está adequada, por estar previsto o advérbio *tristemente* como opção de tradução para *sadly*.

Confronto

Não se constatou tradução distinta. A TA da frase, no primeiro momento, "transferiu" a ambiguidade para o português, mas também não apresentou escolha lexical adequada: *Então,* ***infelizmente****, eu reclamei do decote de uma mulher escandalosa, expondo seus braços*

---

[52] Frase extraída do *corpus: Then,* ***sadly****, in the tones of a wounded moralist, I complained of the outrageous décolletage of some particularly immodest woman, exposing her arms and her chest, making a display of all that splendid, irreligious flesh, a cause of grief to all righteous men and a great sorrow to the Church.*

e seu peito, fazendo uma exposição de tudo o que a carne, esplêndida irreligiosos, um motivo de tristeza para todos os homens justos e uma grande tristeza para a Igreja.

4.3.2.5 (B1.6) Ligação de particípio ao sujeito estrutural da frase ou à frase

Instância com base em "...considering..."

O particípio *considering*, parte da frase (27), pode ligar-se ao sujeito da frase, expresso pelo pronome *I,* ou ao à frase (*I was always light on my feet*).

Frase do inglês

27) I was always light on my feet, ***considering*** *the bulk they have to support*[53].

Leituras

(A) Eu, **considerando** o peso que os pés têm de suportar, sempre poupei os meus (pés).
(B) Eu sempre poupei os meus pés, **considerando** o peso que eles têm de suportar.

Traduções

TH: (eu) Sempre fui leve nos pés, **considerando** o peso que eles têm de suportar...
TA: Eu estava sempre a luz em meus pés, **considerando** o volume que tem que suportar.

Avaliação

Esse tipo de ambiguidade é transferido para o português sem o prejuízo para a TA.

Confronto

Não se constatou tradução distinta. A TA da frase foi confirmada: *Eu estava sempre a luz em meus pés,* ***considerando*** *o volume que tem que apoiar, um driblador bacana de uma bola de futebol e não um dançarino de salão ruim.*

4.3.2.6 (B1.7) Ligação simultânea de adjunto adverbial a verbos de duas frases distintas

Instância com base em "tomorrow afternoon"

---

[53] Frase extraída do *corpus: I was always light on my feet,* ***considering*** *the bulk they have to support, a nifty dribbler of a football and not a bad ballroom dancer.*

A ambiguidade é causada pelo fato de o adjunto adverbial *tomorrow afternoon*, parte da frase (28), poder modificar qualquer um dos dois verbos da frase, *tell* ou *expect.*

Frase do inglês

28) I told daddy to expect me ***tomorrow afternoon*** [54].

Leituras

(A) Diga ao pai p´ra me esperar lá **amanhã à tarde** (modificação de *expect*).

(B) Diga, **amanhã à tarde**, ao pai p´ra me esperar lá (modificação de *tell*).

Traduções

TH: -- Diz ao pai que eu apareço lá **amanhã** à tarde...

TA: Eu disse ao papai para me esperar amanhã à tarde.

Instância com base em "sometimes"

Na frase (29), a ambiguidade se deve ao fato de o adjunto adverbial *sometimes* poder modificar qualquer um dos dois verbos da frase, *seeing* ou *going out.*

Frase do inglês

29) The kid sees her mother ***sometimes*** *going out in the evening dressed to kill.* [55]

---

[54] Frase extraída do *corpus: Tell Daddy to expect me* ***tomorrow*** *afternoon, he said, and rang off.*

[55] Frase extraída do *corpus*: *I can't believe the kid doesn't suspect something, seeing her mother* ***sometimes*** *going out in the evening dressed to kill and carrying a container of home-made frozen food in one of her smartly gloved hands.*

Leituras

(A) A garota **às vezes** vê a sua mãe saindo de casa vestida para matar. (modificação de sees=vê).

(B) A garota vê a sua mãe saindo de casa **às vezes** vestida para matar. (modificação de *going out*).

Traduções

TH: Eu não acredito que a miúda não desconfie de qualquer coisa, vendo a mãe sair **de vez em quando** (sic) à noite vestida para matar...

TA: O garoto vê sua mãe **às vezes** sair à noite, vestida para matar.

Avaliação

Essas ambiguidades de ligação entre o adjunto adverbial e verbos da frase, em geral, são transferidas para a língua alvo, sem prejuízo da TA.

Confronto

Não se constatou tradução distinta nos dois momentos.

4.3.2.7 (B3.1) Detecção de partículas

Instância com base em "look up"

Em *look up*, a partícula *up* pode ser parte integrante do verbo *look* ou constituir uma preposição independente, causando ambiguidade na interpretação. Na frase (30), *up* pode integrar-se ao verbo *look* e formar uma unidade de sentido (consultar) ou dele se separar e encabeçar um SP Locativo (para cima).

Frase do inglês

*30) I **looked up** the original contract.*

Leituras

(A) Eu **consultei** o contrato original.

(B) Eu olhei **para cima** do contrato original.

Traduções

TH: Dei uma lida no contrato original.

TA: Eu **olhei para** o contrato original.

Avaliação

A TA de (30) não levou em conta a presença da partícula *up* nessa frase e traduziu apenas o verbo *look* por *olhar*. Contudo, esse comportamento não se verifica na totalidade dos casos, assim como pode ser observado com base na análise do Quadro 2, apresentado na seção 2.1.1.1.

Confronto

Não se constatou tradução distinta.

4.3.2.8 (B3.3) Diferenciação entre particípio presente e adjetivo

Instância com base em "singing"

Na frase (31), *singing* pode ser o particípio presente de *sing* ou um adjetivo (modificador de *nonsense words*).

Frase do inglês

31) A moment's attention revealed they were ***singing*** *nonsense words*[56].

Leituras

(A) Um momento de atenção revelou-me que eles estavam **cantando** palavras sem sentido (particípio presente).

(B) Um momento de atenção revelou-me que elas eram palavras sem sentido **feitas para cantar** (adjetivo).

Traduções

TH: Na verdade, um momento de atenção revelou-me que cantavam ambos palavras absurdas.

TA: Um momento de atenção revelaram que eles estavam cantando palavras sem sentido.

---

[56] Frase extraída do *corpus: Indeed, a moment's attention revealed they were both* ***singing*** *nonsense words.*

Instância com base em "carrying"

Na frase (32), *carrying* pode ser o particípio presente do verbo *carry* ou um adjetivo modificador de *placards* (modificador de *placards*).

Frase do inglês

*32) They were **carrying** placards*[57].

Leituras

(A) Eles estavam **carregando** cartazes. (particípio presente).

(B) Eles eram cartazes **feitos para carregar.** (adjetivo).

Traduções

TH: Carregavam cartazes...

TA: Eles estavam carregando cartazes.

Avaliação

As TAs das frases (31) e (32) evidenciam a preferência pela forma verbal e não adjetival da forma *–ing*.

Confronto

Não se constatou tradução distinta.

4.3.2.9 (B3.4) Diferenciação entre particípio presente e substantivo

Instância com base em "swimming"

Na frase (33), a ambiguidade decorre do fato de o item lexical *swimming* poder ser analisado como particípio presente (do verbo *swim*) ou como um substantivo derivado.

Frase do inglês

*33) It would be better to miss **swimming** and go straight down to the coast road*[58].

---

[57] Frase extraída do *corpus: They were **carrying** placards and singing, turning the protest into a party.*

[58] Frase extraída do *corpus: It would be better to miss **swimming** and go straight down to the coast road or look for my friends in their boarding houses.*

Leituras

(A) Seria melhor perder **nadando** e descer até a rodovia marginal (particípio presente).

(B) Seria melhor perder a **natação** e descer até a rodovia marginal (substantivo).

Traduções

TH: O melhor seria não tomar banho (sic), descer até a avenida marginal...

TA: Seria melhor **perder de natação** e ir direto para a estrada da costa.

Avaliação

A TA da frase (33) não é adequada porque não traduz *swimming*, quer como particípio presente quer como substantivo, resultando em uma frase agramatical. O sistema optou por traduzir *swimming* por um SP formado pela preposição *de*+substantivo, gerando uma frase anômala, problema que poderia ser solucionado considerando-se informações sobre o ambiente sintático do qual o item lexical é parte. Se o item lexical for uma forma do particípio presente, ele admitirá a ocorrência de advérbios e a partícula *not,* por exemplo. Se, contudo, for um substantivo, ele ocorrerá na presença do artigo *the* ou de adjetivos.[59]

Confronto

Não se constatou tradução distinta.

---

[59] Vale lembrar que para o exemplo extraído de Hirst (1992), *We discussed running*, já mencionado anteriormente, a TA também não foi adequada, porque *running* foi traduzido por *em execução*, um SP, situação que se repete na frase (33).

4.3.2.10 (B3.8) Interpretação ambígua de particípios e de orações adjetivas no final de frase

Instância com base em "smiling"

A ambiguidade se origina a partir do fato de que tanto o sujeito (*I*) quanto o objeto (*a gaunt face*) do verbo da oração principal (*imagined*) podem exercer a função de sujeito da oração predicativa (*smiling*).

Frase do inglês

34) I imagined a gaunt face ***smiling***[60].

Leituras

(A) Eu, **sorrindo**, imaginei um rosto descarnado.

(B) Eu imaginei um rosto descarnado que estava **sorrindo**.

Traduções

TH: ... e imaginei ver nas costas em forma de lira da cadeira uma face descarnada **a sorrir**.

TA: Imaginei um rosto descarnado **sorrindo**.

Avaliação

A TA da frase (34) transfere a ambiguidade para o português, sem comprometer a sua adequação.

Confronto

Não se constatou tradução distinta.

---

[60] Frase extraída do *corpus: Alex's desk was in the corner, and in the lyre-shaped back of her chair, I imagined a gaunt face* ***smiling***.

4.3.2.11 (B3.11) Delimitação dos diferentes tipos de estrutura formados com esta sequência de elementos: be ADJ to V

Instância com base em "was easy to please"

A sequência *be ADJ to V* pode estruturar-se de modos diferentes. Na frase (35), é necessário estabelecer se o sujeito (implícito) do verbo *to please* é correferente ao sujeito do verbo *was*.

Frase do inglês

35) Catherine ***was easy to please***[61].

Leituras

(A) **Era fácil de** se **agradar** Catherine (sujeitos distintos).

(B) *Catherine **agradava**-se a si mesma **de modo fácil** (sujeitos iguais).

Traduções

TH: Até então Catherine nunca fora de críticas severas; **era fácil de contentar...**

TA: Catherine **foi fácil de agradar.**

Avaliação

A TA da frase (35) é adequada, posto que, sugere-se, se trata de uma construção que foi o centro das discussões gerativistas da década de 60 e, portanto, não é difícil supor que o Google Tradutor tenha essa estrutura já prevista em seu banco de dados.

Confronto

Não se constatou tradução distinta.

---

[61] Frase extraída do *corpus: Catherine had hitherto not been a harsh critic; she* ***was easy to please*** *-- she liked to talk with young men.*

4.3.3 (B4) Ambiguidades de escopo da quantificação

Instância com base em "every"

Na frase (36), o escopo do quantificador universal *every* define-se em relação ao escopo do quantificador existencial *a*, conforme ilustram as leituras.

Frase do inglês

36) ***Every woman*** *in America is on a diet*[62].

Leituras

(A) **Todas as mulheres** estão fazendo **um mesmo regime**.
(B) **Cada uma das mulheres** está fazendo **um regime diferente**.

Traduções

TH: ... **todas as mulheres** da América fazem **dieta**.
TA: **Toda mulher** na América está em **uma dieta**.

Instância com base em "everybody"

Na frase (37), o escopo de *everybody* define-se em relação ao escopo do quantificador existencial *a*, conforme ilustram as leituras.

Frase do inglês

37) ***Everybody*** *was holding out a ticket*[63].

Leituras

(A) **Todas as pessoas** estavam mostrando **um mesmo bilhete**.
(B) **Cada uma das pessoas** estava mostrando **um bilhete diferente.**

---

[62] Frase extraída do *corpus: I have this theory that* ***every woman*** *in America is on a diet.*

[63] Frase extraída do *corpus: In a moment,* ***everybody*** *was holding out a ticket: they were about the same size as the people, and quite seemed to fill the carriage.*

Traduções

TH: ... **todos** mostraram um bilhete...

TA: **Todo mundo** estava segurando **um** bilhete.

Avaliação

Nas TAs das frases (36) e (37), a ambiguidade de escopo da quantificação foi transferida para a língua alvo, sem seu prejuízo.

Confronto

Não se constatou tradução distinta.

4.3.4 (C) Ambiguidades anafóricas/referenciais

Instância com base em "them"

Na frase (38), existe mais de um SN possível para ser o antecedente do pronome *them*. Os SNs candidatos são *young people*, *steps* e *sexes*, devido à concordância de número, que exclui *bank*.

Frase do inglês

38) I passed several young people, of both sexes, sitting on the steps of a bank, some of ***them*** *drunk, some waiting for the dealer, others waiting for customers, waiting for Godot, waiting for the film (me, certainly)*[64]*...*

Leituras

(A) Passei por vários **jovens** sentados nos degraus da porta de um banco, alguns **deles** bêbados.

(B) Passei por vários jovens sentados nos **degraus** da porta de um banco, alguns **deles** bêbados

---

[64] Frase extraída do *corpus: I passed several young people, of both sexes, sitting on the steps of a bank,* ***some of them*** *high, some waiting for the dealer, others waiting for customers, waiting for Godot, waiting for the film (me, certainly) , waiting for the night to end, for life to end.*

(C) Passei por vários jovens, de ambos os **sexos**, sentados nos degraus da porta de um banco, alguns **deles** bêbados.

Traduções

TH: Passei por vários jovens, de ambos os sexos, sentados nos degraus da porta de um banco, moças e rapazes, **alguns** drogados (sic), esperando o traficante, outros esperando o freguês, esperando Godot, esperando o filme (eu, certamente)...

TA: Passei por vários jovens, de ambos os sexos, sentado nos degraus de um banco, alguns **deles** bêbados, alguns à espera do comerciante, outros à espera de clientes, esperando Godot, esperando o filme (eu, certamente)...

Instância com base em "it"

Na frase (39), é preciso decidir qual é o antecedente adequado para o pronome *it*. Os antecedentes possíveis são os núcleos (em negrito) dos SNs *His mother's* ***hand***, *a sharp* ***slap*** *to the boy's forearm,* e *the boy's* ***forearm***.

Frase do inglês

39) His mother's hand delivered a sharp slap to the boy's forearm, knocking ***it*** *away*[65].

Leituras

(A) A mão da mãe disparou como uma flecha por cima da toalha e deu uma pancada forte no **antebraço do filho**, fazendo-**o** baixar.

(B) A mão da mãe disparou como uma flecha por cima da toalha e deu uma **pancada** forte no antebraço do filho, fazendo-**a** baixar.

(C) A mão da mãe disparou como uma flecha por cima da **toalha** e deu uma pancada forte no antebraço do filho, fazendo-**a** baixar.

Traduções

TH: ... a mão da mãe disparou como uma flecha por cima da toalha e deu uma pancada forte no **antebraço do filho**, fazendo-**o** baixar.

TA: Mão de sua mãe entregou um tapa afiada para o antebraço do rapaz, derrubando-o longe .

---

[65] Frase extraída do *corpus: Immediately, his mother's hand flashed across the tablecloth and delivered a sharp slap to the boy's forearm, knocking* ***it*** *away.*

Avaliação

A TA das frases (38) é adequada. Contudo, essa adequação é acidental, posto que, quando o SN *young people* é substituído pelo SN *girls*, a TA passa a inadequada: *Passei por várias meninas, sentado nos degraus de um banco, alguns* ***deles*** *bêbados, alguns à espera do comerciante, outros à espera de clientes, esperando Godot, esperando o filme (eu, certamente)*. Comentário análogo aplica-se à TA da frase (39). Note-se que o número do pronome não soluciona a ambiguidade da frase (39), porque o pronome e todos os seus antecedentes estão no singular; na frase (38), eliminaria apenas *bank* como possível antecedente, já que o pronome *them* deve retomar um SN no plural.

Confronto

Não se constatou tradução distinta.

Em suma, com base nas análises dessa subseção, registra-se que das 39 frases do *corpus* adaptadas e submetidas ao *Google Tradutor* no segundo momento, 31 são TAs adequadas; para 8 frases, as TAs foram inadequadas. Durante a discussão, sugeriram-se breves estratégias de resolução de ambiguidade para 7 frases. Apenas 1 frase com TA inadequada permanece sem discussão de resolução, a saber, a frase 30, que ilustra o tipo de ambiguidade **B3.1 Detecção de partículas**.

Não se apontaram estratégias de resolução para estes quatro tipos ambiguidade que foram analisados nessa subseção porque as TAs não foram inadequadas: **B1.6** Ligação de particípio ao sujeito estrutural da frase ou à frase; **B1.7** Ligação simultânea de adjunto adverbial a verbos de duas frases distintas; **B3.3** Diferenciação entre particípio presente e adjetivo e **B3.8** Interpretação ambígua de particípios e de orações adjetivas no final de frase.

O confronto de testes nos dois momentos registrou que das 39 frases, para 12 frases houve discordância de resultados: passaram de inadequadas a adequadas, são elas as frases (1); (2); (3); (4); (5); (7); (9); (10); (12); (13); (15) e (17). Essas 12 frases ilustram 2 tipos de ambiguidades, a saber, **A1** Ambiguidade devido à homonímia ou polissemia e **A2** Ambiguidade categorial.

Para as TAs das 27 frases restantes, no entanto, não houve discrepância nos testes, ou seja, as TAs foram as mesmas. Essas 27 frases distribuem-se dentre estes 16 tipos: **A1** Ambiguidade devido à homonímia ou polissemia; **A2** Ambiguidade categorial; **A3** Ambiguidade de transferência; **B1.1** Ligação de sintagma preposicional a mais de um

sintagma nominal ou verbal; **B1.2** Ligação de oração relativa a mais de um sintagma nominal; **B1.3** Ligação de sintagma preposicional a adjetivo predicativo; **B1.5** Ligação de advérbio como modificador de sintagma verbal ou de frase; **B1.6** Ligação de particípio ao sujeito estrutural da frase ou à frase; **B1.7** Ligação simultânea de adjunto adverbial a verbos de duas frases distintas; **B3.1** Detecção de partículas; **B3.3** Diferenciação entre particípio presente e adjetivo; **B3.4** Diferenciação entre particípio presente e substantivo; **B3.8** Interpretação ambígua de particípios e de orações adjetivas no final de frase; **B3.11** Delimitação dos diferentes tipos de estrutura formados com esta sequência de elementos: *be ADJ to V;* B4 Ambiguidade de escopo da quantificação; **C** Ambiguidade anafórica ou referencial.

Dessas 27 TAs: 19 TAs foram adequadas no primeiro e no segundo momentos; 8 TAs foram inadequadas nos dois momentos.

Enfatiza-se, por fim, que, por ser fortemente baseado em métodos estatísticos, envolvendo comparação de padrões e de traduções já existentes na sua base de conhecimento, o *Google Tradutor* apresenta comportamento instável e idiossincrático, que não permite que se avalie, com a precisão desejada, a qualidade do seu desempenho de modo absoluto, já que as TAs se alteram à medida que a sua base de conhecimento se altera. Por outro lado, os testes exploratórios aqui realizados são indicativos de deficiências do sistema, ao mesmo tempo em que permitem apontar conhecimentos que poderiam ser somados à base estatística do sistema e empregados na otimização das TAs geradas.

5. Considerações finais

Considerando os problemas amplamente estudados e postos pelas ambiguidades linguísticas no processo de TA inglês-português, não é possível desconsiderar as contribuições que conhecimentos de natureza linguística trazem para o aperfeiçoamento da TAs geradas pelo sistema. Com foco nesses problemas, essa pesquisa objetivou descrever, sistematizar e exemplificar os principais tipos de ambiguidade linguística do inglês apontados na literatura; avaliar, de modo exploratório a TA inglês-português gerada pelo *Google Tradutor,* e explorar estratégias para resolvê-las.

Para alcançar os objetivos, adotou-se a metodologia de pesquisa em PLN que distribui o estudo em três domínios complementares: o linguístico, o linguístico-computacional e o computacional (não abordado nessa pesquisa por incluir tarefas próprias dos cientistas da computação).

No domínio linguístico, apresentou-se, de modo geral, o fenômeno das ambiguidades linguísticas no inglês, de acordo com as discussões dos especialistas, e seu modo de manifestação no processo de TA, inglês-português, de frases e parágrafos. Ainda nesse domínio, catalogaram-se, descreveram-se e ilustraram-se com exemplos os tipos e subtipos de ambiguidade levantados da literatura sobre PLN/TA. Os quadros 6 e 7 são a síntese dessa tarefa.

Para o desenvolvimento da pesquisa, foi preciso recorrer ao auxílio de alguns recursos, como o *Google Tradutor,* a ferramenta de TA escolhida para os testes empíricos, por ser ele amplamente usado e de acesso gratuito, e o COMPARA, o *corpus* paralelo (feito por tradutores humanos) selecionado para fornecer as frases e parágrafos submetidos à TA e também, por conter traduções feitas por tradutores humanos, forneceu uma base de comparação para as TAs.

Como se mostrou no quadro 6, que se complementa com o Quadro 7, com exemplos de frases extraídas das literatura, do COMPARA ou livres, as ambiguidades linguísticas podem se manifestar em diversos níveis de análise linguística.

No domínio linguístico-computacional, com base no conhecimento sistematizado no domínio puramente linguístico, investigou-se o grau de adequação da qualidade da TA das frases e parágrafos do inglês para o português, analisando-se se e como a adequação da TA é comprometida por ambiguidades linguísticas. Essa investigação, feita por meio da

comparação das traduções feitas por tradutores humanos, e extraídas do *corpus*, com as TAs geradas pelo *Google Tradutor,* possibilitou a construção do Quadro 8, que forneceu argumentos à favor da afirmação de que os textos gerados pelo sistema de TA têm, na maioria das vezes, sua qualidade afetada por ambiguidades linguísticas, chegando a produzir resultados inaceitáveis (sintática e/ou semanticamente) para o português. Nessa etapa é que se investigaram as principais estratégias de resolução dos tipos de ambiguidade sistematizados no domínio linguístico.

Para avaliar o comprometimento da TA em decorrência de ambiguidades linguísticas e pontuar estratégias de resolução, foram selecionadas frases do inglês ilustrativas dos tipos e subtipos de ambiguidade sistematizados no *corpus* e submetidas à TA em dois momentos: março-setembro de 2010 e março de 2011. Dos 25 tipos e subtipos de ambiguidade considerados nesta pesquisa, 16 foram ilustrados com frases do *corpus.*

A comparação das TAs geradas pelo *Google Tradutor* nos dois momentos mostrou que 12 frases para as quais os resultados fornecidos no primeiro momento do teste eram inadequados, receberam TAs adequadas no segundo momento, sendo que todas elas ilustram ambiguidades no nível lexical. Esses números talvez se expliquem pelo fato de haver grande número de estudos a esse respeito e, consequentemente, grande quantidade de sugestões de estratégias para resolução.

Por outro lado, outras 8 frases haviam recebido resultados inadequados no primeiro momento do teste e esses resultados não se modificaram no segundo momento. O que chama a atenção nesse número é que 6 dessas frases também representam ambiguidades que se manifestam no nível lexical. Como mencionado anteriormente, muitos estudos se direcionam a essa temática e ainda assim esse tipo de ambiguidade parece ser bastante problemático. Para explicar esse fato, vale lembrar que as ambiguidades que se manifestam no nível lexical exigem sempre que o sistema escolha apenas um sentido entre os possíveis para um item lexical, o que não ocorre com outros tipos de ambiguidades (estruturais e de escopo da quantificação, por exemplo), que podem ser, muitas vezes, mantidos na TA gerada, sem que a frase seja sintática ou semanticamente anômala.

Resolver as ambiguidades linguísticas ou pelo menos prever seus efeitos sobre as TAs, além de otimizar grau de adequação das TAs, representaria um grande avanço para o PLN. Por isso, muitos esforços precisam ser dedicados para a descoberta de estratégias que possam ser aplicadas na resolução dos diversos tipos de ambiguidade.

Além da modelagem do contexto, que é essencial para fornecer informações, outras estratégias podem ser aplicadas, isoladas ou em conjunto, no processo de resolução das ambiguidades linguísticas: pré-edição do texto a ser traduzido ou pós-edição do resultado e modelagem e implementação no sistema de informações sintáticas e semânticas como a associação semântica entre os sentidos dos itens lexicais presentes no co-texto, mecanismos de inferência de informações contextuais e construção de léxicos e gramáticas mais explícitas e robustas, que prevejam as diferentes estratégias de desambiguição discutidas na subseção 4.2.

Como possibilidades para trabalho futuro, sugerem-se:

(i) Uma investigação mais aprofundada de diferentes tipos de frases e construções que possam trazer problemas para a TA em *corpora* mais abrangentes, para se modelarem estratégias de resolução para os diferentes tipos de ambiguidade que aqui não foram trabalhados, já que não se encontraram frases ilustrativas no *corpus*: **B1.4** Ligação de sintagma preposicional ou advérbio a posições pertencentes à oração principal ou à oração encaixada; **B2** Ambiguidade de localização e preenchimento de lacunas; **B3.2** Diferenciação entre um sintagma preposicional e um sintagma adjetivo resultante de uma operação de alçamento e apagamento do verbo ser/estar aplicada ao complemento do verbo; **B3.5** Delimitação da extensão do sintagma nominal; **B3.6** Diferenciação entre oração relativa reduzida e sintagma verbal da oração principal; **B3.7** Delimitação da estrutura de um sintagma nominal complexo; **B3.9** Diferenciação entre frases clivadas e frases do tipo sujeito-verbo-objeto; **B3.10** Diferenciação entre particípio passado e um sintagma verbal incompleto, resultando na ambiguidade entre pergunta e ordem e **D** Ambiguidade temática.

(ii) Uma investigação mais aprofundada dos tipos para os quais não se apontaram estratégias de resolução: **B1.4** Ligação de sintagma preposicional ou advérbio a posições pertencentes à oração principal ou à oração encaixada; **B1.5** Ligação de advérbio como modificador de sintagma verbal ou de frase; **B1.6** Ligação de particípio ao sujeito estrutural da frase ou à frase; **B1.7** Ligação simultânea de ajunto adverbial a verbos de duas frases distintas; **B2** Ambiguidade de localização e preenchimento de lacunas; **B3.1** Detecção de partículas; **B3.2** Diferenciação entre um sintagma preposicional e um sintagma adjetivo resultante de uma operação de alçamento e apagamento do verbo ser/estar aplicada ao complemento do verbo; **B3.3** Diferenciação entre particípio

presente e adjetivo; **B3.6** Diferenciação entre oração relativa reduzida e sintagma verbal da oração principal; **B3.7** Delimitação da estrutura de um sintagma nominal complexo; **B3.8** Interpretação ambígua de particípios e de orações adjetivas no final de frase; **B3.9** Diferenciação entre frases clivadas e frases do tipo sujeito-verbo-objeto; **B3.10** Diferenciação entre particípio passado e um sintagma verbal incompleto, resultando na ambiguidade entre pergunta e ordem e **B3.11** Delimitação dos diferentes tipos de estrutura formados com esta sequência de elementos: *be ADJ to V*.

(iii) Uma análise sistemática do comportamento idiossincrático do *Google Tradutor* e, sobretudo, das TAs inadequadas por ele geradas.

REFERÊNCIAS

ALFARO, C. **Descobrindo, Compreendendo e Analisando a Tradução Automática.** Rio de Janeiro, 1998. Disponível em: <http://www.tecgraf.puc-rio.br/~carolina/monografia/>. Acesso em: 25 out. 2010.

ALLEN, J. **Natural Language Understanding.** Redwood City, CA: Benjamin/Cummings, 1995.

ALMEIDA, J. Ambiguidade lexical. **Revista Alfa**, São Paulo. Vol 34, p. 187-193, 1990.

ALPAC. **Language and Machines**: Computers in Translation and Linguistics. Washington: NAS/NRC, 1966.

BALDWIN, T.; KORDONI, V.; VILLAVICENCIO, A. Prepositions in Applications: A Survey and Introduction to the Special Issue**. Computational Linguistics**, Cambridge, v. 35, n. 2, p. 119 – 149, Jun. 2009. Disponível em: < http://portal.acm.org/toc.cfm?id=1596839&coll=portal&dl=ACM&type=issue&idx=J25&part=affil&WantType=Affils&title=Computational%20Linguistics&CFID=37461972&CFTOKEN=28697371>. Acesso em: 04 out. 2010.

COLE, R. (Ed.) **Survey of the art in human language technology**. New York: Cambridge University Press, 1997.

CRUSE, A. **A Glossary of Semantics and Pragmatics**. Edinburgh University Press, 2006.

DALE, R.; MOISL, H.; SOMERS, H. **Handbook of natural language processing**. New York: Marcel Dekker, 2000.

DIAS-DA-SILVA, B.C.; MONTILHA, G.; RINO, L.H.M.; SPECIA, L.; NUNES, M.G.V.; OLIVEIRA Jr., O.N.; MARTINS, R.T.; PARDO, T.A.S. 2007. **Introdução ao Processamento das Línguas Naturais e algumas aplicações. Série de Relatórios Técnicos do NILC, NILC-TR-07-10**. São Carlos, ICMC, 121p.

DIAS-DA-SILVA, B. C. O estudo linguístico-computacional da linguagem. **Letras de hoje**, Porto Alegre, v. 41, p. 103-138, 2006.

DIAS-DA-SILVA, B. C. (2010). Brazilian Portuguese WordNet: a computational-linguistic exercise of encoding bilingual relational lexicons. **International Journal of Computational Linguistics and Applications**, v.1, p.137 - 150, 2010.

DIAS-DA-SILVA, B.C.; OLIVEIRA, M. F.; MORAES, H.R. Groundwork for the development of the Brazilian Portuguese wordnet. In: RANCHHOD, E.; MAMEDE, N. J. (Org.). **Advances in natural language processing**. Berlin / Heidelberg, 2002, v. 2389, p. 189-196.

_____. **A face tecnológica dos estudos da linguagem**: o processamento automático das línguas naturais. Araraquara, 1996. 272 f. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 1996.

FRANKENBERG-GARCIA, A.; SANTOS, D. COMPARA, um corpus paralelo de português e de inglês na Web. **Cadernos de Tradução IX,** Santa Catarina, p. 61-79, 2002.

FRANKENBERG-GARCIA, A.; SANTOS, D. Introducing COMPARA, the Portuguese-English parallel translation corpus. In**:** ZANETTIN, F.; BERNARDINI S.; STEWART, D. (Eds.). **Corpora in Translation Education.** Manchester: St. Jerome Publishing, 2003. p. 71-87.

GAZDAR, G., MELLISH, C. **Natural language processing in prolog: an introduction to computational linguistics**. New York: Addison-Wesley, 1989.

HALLIDAY, M., HASAN, R. **Cohesion in English.** London: Longman, 1976.

HATIM, B., MASON, I. **Discourse and the translator.** New York: Longman Inc., 1990.

HIRST, G. **Semantic interpretation and the resolution of ambiguity**. Cambridge: Cambridge University Press, 1992.

HOUAISS, A. **Webster's:** dicionário inglês-português. 15. ed. Rio de Janeiro: Record, 2005.

HUTCHINS, W. J. **Machine translation: past, presence, future**. Ellis Horwwod/Wiley, Chichester/New York, 1986.

_____. SOMERS, H. L. **An introduction to machine translation**. London: Academic Press, 1992.

_____. Machine Translation over fifty years. **Histoire, Epistemologie, Langage**, v. 22, n. 1, p. 7-31. 2001. Disponível em: <http://www.hutchinsweb.me.uk/history.htm>. Acesso em: 28 jul. 2010.

_____. Machine translation: general overview. In: MITKOV, R. (Ed.). **The Oxford handbook of Computational Linguistics.** Oxford: Oxford University Press, 2003. p. 501-511.

IDE, N.; VÉRONIS, J. Introduction to the Special issue on word sense disambiguation: The State of the Art. **Computational Linguistics**. Cambridge, v. 24, p. 2 – 40, Mar. 1998. Disponível em: <http://portal.acm.org/citation.cfm?id=972749.972751&coll=GUIDE&dl=%EF%BF%BD%C3%9C&idx=J25&part=affil&WantType=Affils&title=Computational%20Linguistics&CFID=96305886&CFTOKEN=64368211>. Acesso em: 28 jul. 2010.

INTRODUÇÃO ao COMPARA. Projecto Processamento Computacional do Português. Disponível em: <http://www.linguateca.pt/COMPARA/compara_intro.php>. Acesso em: 05 ago. 2010.

JENSEN, K.; BINOT, J. L. Disambiguating prepositional phrase attachments by using on-line dictionary definitions. **Computational Linguistics.** Cambridge, v. 13, n. 3-4, p. 251-260, Jul. 1987. Disponível em: <http://portal.acm.org/toc.cfm?id=48160&coll=portal&dl=ACM&type=issue&idx=J25&part

=affil&WantType=Affils&title=Computational%20Linguistics&CFID=37461972&CFTOKEN=28697371>. Acesso em: 04 out. 2010.

JURAFSKY, D.; MARTIN, J. **Speech and Language Processing**. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall, 2000.

KILGARRIFF, A. **Polysemy**. 1992. 146 f. PhD thesis – University of Sussex, School of Cognitive and Computing Sciences, 1992.

_____. What is word sense disambiguation good for?. In: Natural Language Processing in the Pacific Rim, 1997, Phuket, Thailand. **Proceedings…,** Phuket, Thailand, 1997. p. 209-214. Disponível em: <http://www.kilgarriff.co.uk/publications.htm>. Acesso em: 23 jul. 2010.

KURZWEIL, R. **The age of intelligent machines.** MIT Press, 1990.

LEVI, J. N. **The syntax and semantics of compound nominals**. New York: Academic Press, 1978.

LONGMAN Dictionary of Contemporary English. 3. ed. Nova Delhi: Gopsons Papers Limited, 2004.

MANNING, C. D.; SHÜTZE, H. **Foundations of Statistical Natural Language Processing.** Cambridge: MIT press, 1999.

MARCUS, M. P. **A theory of syntactic recognition for natural language.** Cambridge: The MIT press, 1980.

MARTINS, R. T.; NUNES, M. G. V. **Noções Gerais de Tradução Automática**. São Carlos: ICMC-USP, 2005.

MARTINS, R. T. Tradução automática e estudos da tradução: um conflito paradigmático. In: WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA,1, 2003, São Carlos. **Anais...** São Carlos, 2003. Disponível em:

<http://www.nilc.icmc.usp.br/til2003/oral/RonaldoMartins_31.pdf>. Acesso em: 25 out. 2010.

MITKOV, R.; CHOI, S. K.; SHARP, R. Anaphora resolution in machine translation. In: Sixth International Conference on Theoretical and Methodological Issues in Machine Translation**,** 6th.**,** 1995, Leuven, Belgium. **Proceedings…,** Leuven, Belgium: Katholieke Universiteit**,** 1995. p. 87-95. Disponível em: <http://www.mt-archive.info/TMI-1995-Mitkov.pdf>. Acesso em: 28 jul. 2010.

_____. Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches. In: Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts, 1997, Madrid, Spain. **Proceedings… Madrid, Spain,** 1997. p. 14-21. Disponível em: <http://portal.acm.org/citation.cfm?id=1598819.1598822&coll=GUIDE&dl=GUIDE,&type=series&idx=SERIES12862&part=series&WantType=Proceedings&title=ACL%20Workshops&CFID=96307541&CFTOKEN=23902547>. Acesso em: 28 jul. 2010.

_____(Ed.). **The Oxford handbook of computational linguistics.** Oxford: Oxford University Press, 2003.

NIRENBURG, S. Bar Hillel and machine translation: then and now. In: BISFAI'95 The Fourth Bar-Ilan Symposium on Foundations of Artificial Intelligence, 4th., 1995, Jerusalem, Israel. **Proceedings…**, Jerusalem, Israel: AAAI Press, 1996. p.300-305. Disponível em: <http://www.aaai.org/Papers/BISFAI/1995/BISFAI95-027.pdf>. Acesso em: 16 jul. 2010.

_____, SHAPIRO, S. **Machine translation**. San Mateo, Ca: Morgan Kaufmann, 1992.

PAVÃO JUNIOR, J. A lígua do Google. **Revista Veja**, São Paulo: Editora Abril, ano 43, n. 18, p. 122-131, maio 2010.

QUIRK, R.; GREENBAUM, S.; LEECH, G.; SVARTVIK, J. **A grammar of contemporary English.** London: Longman, 1972.

QUIRK, R. et al. **A comprehensive Grammar of the English language.** New York: Longman, 1985.

RUPPENHOFER, J. et al. **FrameNet II: Extended Theory and Practice**, 2010. Disponível em: < http://framenet.icsi.berkeley.edu/>. Acesso em: 20 fev. 2011.

SANTOS, D. **A fase de transferência de um sistema de tradução automática do inglês para o português**, 1988. 252 f. Dissertação (Mestrado em Engenharia Eletrotécnica e de Computadores) - Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, 1988. Disponível em: <http://www.linguateca.pt/Diana/public.html>. Acesso em: 15 jul. 2010.

_____. O computador e a tradução. In: II Seminário de Tradução Científica e Técnica em Língua Portuguesa**,** 2., 1999, Lisboa. **Actas do II Seminário de Tradução Científica e Técnica em Língua Portuguesa.** Lisboa, 1999. Disponível em: <http://www.linguateca.pt/Diana/download/SantosSeminTradTecnica99.pdf>. Acesso em: 15 jul. 2010.

SCHWARTS, L.; AIKAWA, T.; QUIRK, C. Desambiguation of English PP-attachment using multilingual aligned dada. In: MACHINE TRANSLATION SUMMIT IX, 9., 2003, New Orleans. **Proceedings…** New Orleans, 2003. Disponível em: <http://www.amtaweb.org/summit/MTSummit/papers.html>. Acesso em: 03 mar. 2010.

SHARP, R. CAT2: Implementing a Formalism for Multi-Lingual MT. In: **Proceedings of the TMI'88 Conference**, Pittsburgh, USA, 1988.

SOMERS, H. Machine translation. In: DALE, R.; MOISL, H.; SOMERS, H. **Handbook of natural language processing**. New York: Marcel Dekker, 2000. p. 329-346.

_____. Machine Translation: Latest Developments. In: MITKOV, R. (Ed.) **The Oxford Handbook of Computational Linguistics**. Oxford: Oxford University Press, 2003. p. 512-528.

SPECIA, L. **Uma abordagem híbrida relacional para a desambiguação lexical de sentido na tradução automática.** São Carlos, 2007. 245 f. Tese (Doutorado em Ciências) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2007.

SPECIA, L.; RINO, L. H. M. **Introdução aos Métodos e Paradigmas de Tradução Automática.** Série de Relatórios Técnicos do NILC, NILC-TR-02-04. São Carlos, 2002.

STEVENSON, M.; WILKS, Y. Word-sense Disambiguation. In: MITKOV, R. (Ed.). **The handbook of Computational Linguistics.** Oxford: Oxford University Press, 2003. p. 249-265.

TAGNIN, S. Corpora: o que são e para quê servem? (Corpora para terminologia). 2004. Disponível em: <http://www.fflch.usp.br/dlm/comet/Novo/Lexicografia.pdf>. Acesso em: 26 out. 2010.

_____. CorTrad: um corpus de traduções inglês-português on-line com mil e uma possibilidades de pesquisa**!** In: Encontro de Férias SBS, 10, 2010, São Paulo. Disponível em: < http://www.fflch.usp.br/dlm/comet/artigos/SBS_Encontro_de_ferias_2010.pdf>. Acesso em: 26 out. 2010.

TAYLOR, J. L. **Webster's**: Portuguese-English dictionary. 16. ed. Rio de Janeiro: Record, 2003.

VILELA, M. **Tradução e Análise Contrastiva**: Teoria e Aplicação. Lisboa: Caminho, 1994.

WILKS, Y. **Machine Translation**: Its Scope and Limits. Springer, New York, 2009.