UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ESTUDOS DA LINGUAGEM
CURSO DE LETRAS

Vitória Pires Zampieri

ORDENAÇÃO DE ADJETIVOS NA TRADUÇÃO AUTOMÁTICA

Campinas
2016

Vitória Pires Zampieri

ORDENAÇÃO DE ADJETIVOS NA TRADUÇÃO AUTOMÁTICA

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Estudos da Linguagem da Universidade Estadual de Campinas como requisito parcial para a obtenção do título de licenciado(a) em Letras.

Orientadora: Profª. Drª. Sonia Cyrino

Campinas

2016

Agradecimentos

Em primeiro lugar, à minha mãe que me apoiou não apenas durante esses anos de graduação, mas toda minha vida e é meu porto seguro. À Alice, por sua doçura e grosseria, sem os quais a vida não seria tão divertida. Ao meu pai, que me mostrou a beleza das palavras.

À Cris, cuja hospitalidade e generosidade me permitiram cursar essa faculdade e que é uma segunda mãe para mim. À Lou, pelo carinho nos anos de graduação e em toda vida. À Giulia, que é minha irmã de coração.

À Beatriz, sobretudo pela amizade, pelas risadas, pelo companheirismo, pela paciência, pelos conselhos em momentos de crise e por sempre estar perto, mesmo estando longe.

À Larissa, pela amizade durante toda a graduação e por consultas sobre regras da ABNT.

À minha orientadora, Prof.ª Drª Sonia Cyrino, que me guiou não apenas nessa monografia, mas durante toda a graduação, sempre muito solícita e com muita paciência.

Resumo

A presente monografia tem como objetivo analisar o desempenho e a eficácia de tradução automática estatística, usando a ferramenta de tradução automática Google Tradutor no que tange a erros de ordenação na tradução de três tipos de estruturas ligadas ao sintagma nominal (SN): apenas um adjetivo, dois ou mais adjetivos e dois ou mais adjetivos com a presença de um enumerador (and, or, etc.). Para tanto, foram traduzidas com a ferramenta cerca de 300 ocorrências de adjetivos atributivos no livro Sherlock Holmes, de Arthur Conan Doyle, observando se algum dos fatores recortados interfere em um maior ou menor índice de erros. Essa análise baseia-se na Teoria de Movimento de Núcleo de Cinque sobre adjetivos pré e pós-nominais nas línguas germânicas e românicas e na teoria de Crisma que parte do mesmo pressuposto, mas considera a importância da interpretação na ordenação desses adjetivos.

Palavras-chave: Tradução automática, Ordenação de adjetivos, Google Tradutor

Abstract

This monograph aims to analyze the performance and effectiveness of statistical machine translation, using the machine translation tool Google Tranlate, focusing in ordering errors in the translation of three types of structures connected to SN: only one adjective; two or more adjectives; and two or more adjectives connected with an enumerator ("and", "or", etc.). To do that, 300 occurrences of attributive adjectives were taken from Sherlock Holmes, by Arthur Conan Doyle, and inserted in the tool, in order to analyze if any of these factors would result in a higher or lower error rate. This analysis is based on the Cinque's Head Movement Theory about pre- and post-nominal adjectives in Germanic and Romance languages and in Crisma's theory that is based on the same assumption, but assesses the importance of interpretation in the adjetive ordering.

Keywords: Machine translation, Adjective ordering, Google Translator

Lista de tabelas

Sumário

1. Apresentação

Neste trabalho tenho como proposta analisar a tradução de adjetivos do inglês para o português, no que diz respeito à sua qualidade, principalmente observando erros de ordenação de adjetivos. Para tal, usarei uma das ferramentas de tradução automática (TA) mais amplamente usadas, o Google Tradutor (GT).

Foi escolhido o livro de Arthur Conan Doyle Um Estudo em Vermelho (1887), da série Sherlock Holmes, que se encontra em domínio público. Foi realizado um levantamento de todas as sentenças que possuem algum tipo de adjetivo atributivo e elas foram inseridas na ferramenta.

Este trabalho tem como proposta analisar a tradução de adjetivos ligados a sintagmas nominais em uma ferramenta de tradução automática sob vários ângulos, especialmente quanto à ordenação correta ou não dos adjetivos, mas observando também erros de concordância, semântica e outros erros de sintaxe que possam ocorrer.

Para tanto, partirei da Teoria de Movimento de Núcleo de Cinque (1990, apud PRIM, 2010), segundo a qual línguas germânicas e românicas têm a mesma ordem linear pré-nominal de ordenação de adjetivos. Em línguas românicas, no entanto, ocorre um movimento de núcleo que move os adjetivos para a posição pós-nominal. Partindo desse pressuposto, considerarei a proposta formulada por Crisma (1990, apud PRIM, 2010) que propõe que a ordenação seja delimitada de acordo com a interpretação do adjetivo e, portanto, que cada adjetivo teria uma colocação específica e fixa na frase.

2. Sherlock Holmes

Um Estudo em Vermelho (1887), de Arthur Conan Doyle, é o primeiro livro da famosa série policial que conta as aventuras do detetive Sherlock Holmes e seu parceiro Watson. Nesse volume, os dois se conhecem e passam a morar juntos, logo após Watson voltar da guerra, na qual havia servido como médico. Eles se tornam amigos e Watson acompanha a investigação de Sherlock em um assassinato e se impressiona com as capacidades de observação e dedução do detetive enquanto resolve o mistério.

Os livros são contados pela perspectiva de Watson e têm grande influência cultural até hoje, resultando em muitas releituras e adaptações, entre elas o filme Sherlock Holmes (2009) e a série Sherlock (2010–presente). Sherlock teve mais de 70 atores interpretando a história em 200 filmes diferentes e, de acordo com o Guiness Book, é o personagem mais adaptado para filmes de todos os tempos. Desse modo, o livro, apesar de ter sido escrito há mais de 100 anos, continua sendo relevante e um fenômeno da cultura mundial.

Os capítulos escolhidos para análise são somente os três primeiros, que somam cerca de 10.000 palavras. Esses capítulos narram o primeiro encontro de Sherlock e Watson em que eles decidem dividir um apartamento, e as primeiras impressões de Watson sobre o detetive, apresentando assim muitas descrições, que resultam na maior presença dos adjetivos usados neste trabalho.

3. Tradução Automática e o Google Tradutor

Embora a ideia de tradução automatizada de textos seja antiga, o começo do desenvolvimento efetivo de uma ferramenta tradução automática remonta a 1947 em um memorando de Andrew D. Booth, um cristalógrafo inglês, a Warren Weaver da Função Rockefeller Foundation, em que ele descreve que tentará codificar um texto em russo como se tivesse sido escrito em inglês de forma a "extrair o código para recuperar informações contidas no texto" (HUTCHINS e SOMERS, 1992, p.12).

Desde essa primeira tentativa, muito foi realizado e a tradução automática popularizou-se e cresceu principalmente nas últimas décadas, paralelamente à popularização de diversas outras tecnologias de computação (computadores pessoais, celulares, internet).

Desse modo, a tradução automática é uma forma de mecanização da tradução que usa programas computacionais para sistematizar a linguagem e converter um idioma natural em outro. Para tal sistematização, há várias estratégias possíveis e destaco aqui a classificação descrita por Ibrahimo (2010), realizada no âmbito do maior ou menor uso de classificações gramaticais ou regras.

Há dois modos principais com os quais as ferramentas funcionam: aqueles baseados em corpora e aqueles baseados em regras. Entre as ferramentas baseadas em regras, há três subdivisões: o método binário, por transferência e por linguística universal.

O método binário baseia-se somente em regras gramaticais. Nesse método, há apenas regras pré-determinadas, sem nenhum processamento semântico e sintático mais aprofundado e utiliza um princípio de equivalência. Desse modo, esses programas "consistem apenas num grande dicionário bilíngue e num programa simples para a análise e geração de textos" e não há nenhuma análise aprofundada sobre as relações sintáticas entre as palavras e as possíveis polissemias semânticas (IBRAHIMO, 2010, p. 19).

De acordo com Ibrahimo (2010), a segunda abordagem, o método por transferência, tem três fases: análise, transferência e geração. Na análise, ela converte o texto para representações abstratas, na transferência converte as representações formais do texto em representações formais na língua de destino e na geração cria o texto na língua de destino.

A terceira abordagem baseia-se no conceito de uma linguística universal e converte o texto de partida em uma interlíngua para, em seguida, transformar para o texto de chegada. Esse método permite traduções para várias línguas diferentes ao mesmo tempo, tem vantagens em ambientes multilíngues e tem um bom desempenho também em frases complexas, pois ele

“requer necessariamente que as ambiguidades no texto de partida sejam completamente resolvidas” (IBRAHIMO, 2010, p. 20).

A principal diferença entre os três métodos é fundamentalmente o nível de profundidade da análise do texto. Enquanto o método binário consiste somente em análise morfológica, o método por transferência também realiza análise sintática e o método por interlíngua combina essas duas análises com uma análise semântica adicional.

Quanto aos métodos baseados em corpora, Ibrahimo (2010) explica que eles são divididos em dois tipos: o método de tradução automática baseado em exemplos e o método baseado em estatística.

No caso de tradução automática baseada em exemplos, usam-se traduções já realizadas por tradutores humanos como base para novas traduções e "se procura interpretar – em vários níveis linguísticos – o sentido do texto original (modelizando explicitamente todo o conhecimento linguístico)” (IBRAHIMO, 2010, p 26).

Por fim, a TA baseada em estatística, em que nos basearemos nesta monografia, usa modelos de aprendizado de máquina para “ensinar” à máquina as regras e exceções da língua, usando, para isso, um modelo baseado em estatística,.

Essa prática baseia-se em aprendizado de máquina, no qual o programa recebe apenas alguns parâmetros básicos de regras de ambos os idiomas e a partir desses parâmetros tem informações suficientes para deduzir por si só o restante das regras e as exceções, resultando em uma tradução mais otimizada e com menos erros contextuais.

Esse método foi idealizado pelo projeto Candide da IBM (KOERNER; ASHER, 1995 apud CASELI; NUNES, 2009) e inicialmente consistia em um método palavra por palavra, mas apresentava algumas falhas que o tornavam pouco efetivo.

De acordo com Lopez (2007, p.7), o modelo considerava que, para cada palavra, deveria haver um resultado equivalente na língua de destino, o que é problemático quando consideramos idiomas morfologicamente complexos como o alemão. Além disso, a escolha de palavras adequadas é de difícil obtenção com um modelo tão simples, o que gera também problemas estéticos. Por fim, esse modelo gera também um problema que Lopez chama de “salada de palavras” (bag-of-words), causado pelo fato de que a reordenação de palavras é realizada independentemente para cada palavra, o que causa complicações, pois muitas vezes na tradução conjuntos de palavras são traduzidos como uma unidade, fazendo com que a frase se torne apenas uma sequência de palavras sem conexão ou sentido.

Considerando esse problema, o modelo foi aperfeiçoado para um modelo de frase a frase e não palavra por palavra. Esse modelo tenta reconhecer cada elemento mínimo sintático e reorganizá-lo de acordo com as regras da língua de destino, usando informações fornecidas pelo corpus e analisadas estatisticamente para criar uma tradução adequada (LOPEZ, 2007).

Assim, esse método resolve o problema de ordenação sintática na língua de destino e tem a vantagem de utilizar estatística e aprendizado de máquina de modo a criar frases de forma mais precisa. Esse método atualmente é considerado o "estado da arte" (CASELI; NUNES, 2009, p. 4) em tradução automática.

Ainda assim, é importante destacar, a tradução automática ainda está muito longe de ser eficiente para traduções profissionais sem a necessidade de auxílio de algum tipo de revisão humana. Em um estudo de Caseli (2007), usando uma ferramenta que utiliza o método de tradução automática por estatística, 67% das frases apresentaram algum erro para o par de línguas inglês > português brasileiro, representando assim uma área que ainda precisa de muitos avanços.

E sse método de tradução automática por estatística é o utilizado pelo Google Tradutor, como já afirmou em vídeo a empresa Google, que desenvolveu o programa[1]. Desse modo, utilizando o vasto corpus ao qual o Google tem acesso, ele combina textos em inglês e suas traduções e correlaciona esses textos de modo a inferir estruturas sintáticas e traduções preferenciais, baseando-se primordialmente em estatísticas geradas por esses dados e parâmetros gramaticais base criados na ferramenta.

Para isso, ele conta com um banco de dados de milhões de textos do inglês para o português. É importante notar, como explicado pelo próprio Google, que um fator importante para a tradução automática estatística é o tamanho desse banco de dados. Assim, um par de idiomas que não tem muitos textos no banco de dados não fornecerá uma tradução tão adequada (uma tradução de uma língua arábica para o português, por exemplo, pode não ser tão bem-sucedida quando uma entre o par inglês>português).

A essa combinação de um corpus extenso traduzido por tradutores humanos e modelos gramaticais que se deve o sucesso do Google Tradutor (GT), que é atualmente uma das melhores ferramentas de tradução automática, especialmente quando consideramos outros programas de tradução por máquina que não possuem corpus ou banco de dados tão amplos quanto o Google.

[1] Retirado de: https://www.youtube.com/watch?v=_GdSC1Z1Kzs. Acesso em 5 de outubro de 2016.

Entretanto, é importante notar que o GT não é específico a nenhuma área, ele é uma ferramenta de tradução genérica, portanto pode não ter conhecimento terminológico específico de determinados temas, sendo sua função principal traduzir textos curtos e de necessidade imediata, sem maior profundidade e sem profissionalização.

É importante também ressaltar que, embora esse sistema represente o que há de mais avançado em tradução automática, ainda há muitos fatores que precisam ser aprimorados e erros de todos os tipos ainda são encontrados nessas traduções, como é possível observar na pesquisa de Caseli (2007) citada acima, por exemplo.

Destaca-se também o texto de Specia (2007), em que ela ressalta que os modelos híbridos usam regras de pré-processamento que são em certa medida simples, de forma a possibilitar que os sistemas de aprendizado de máquina permitam esse aprendizado, mas que, justamente por haver essas delimitações:

> "(...) tais abordagens utilizam um formalismo de representação e técnicas de modelagem limitados, ou seja, algoritmos de aprendizado de máquina baseados em representações vetoriais dos atributos. Isso dificulta a representação de conhecimento profundo, ou seja, que vai além de características extraídas de corpus como bag-of-words e colocações, ou fornecidas por ferramenta superficiais de PLN, como etiquetadores gramaticais, e a utilização desse conhecimento no processo de indução dos modelos de DLS." (p. 147)

Assim, pretendo fazer uma breve análise dos resultados obtidos nessa ferramenta no que tange à ordenação de adjetivos atributivos ligados a um sintagma nominal (SN), para observar melhor as falhas e os acertos dessa ferramenta tão utilizada atualmente, e averiguar se esse modelo apresenta uma carência ou não de marcadores gramaticais, como supõe Specia.

4. Ordenação de Adjetivos e a Hipótese de Movimento de Núcleo

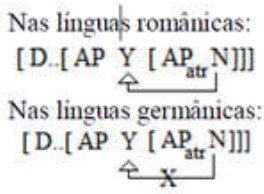
Embora estudos sobre as dificuldades da tradução automática no português brasileiro sejam escassos, usarei para este trabalho as teorias existentes sobre ordenação de adjetivos que tentam explicar uma relação entre as diversas línguas, a fim de observar se esses modelos seriam necessários para uma tradução automática ou se o método híbrido usado é suficiente.

Há muitas teorias que tentam explicar o fenômeno de ordenação de adjetivos, entre elas destaca-se a Hipótese de Movimento de Núcleo, adotada por Cinque (1993, apud PRIM, 2010) e Crisma (1993), que postula que a estrutura sintática da ordenação de adjetivos das línguas germânicas e românicas é a mesma e a diferença que ocorre na ordenação de adjetivos nesses blocos linguísticos é que em línguas românicas é possível haver um movimento de núcleo, que coloca o nome numa posição superior ao adjetivo (PRIM, 2015).

De acordo com Cinque (1993, apud PRIM, 2010), a posição dos adjetivos atributivos em línguas germânicas e românicas provém da mesma estrutura de adjetivação pré-nominal, fazendo com que os adjetivos nas duas línguas tenham o mesmo tipo de ordenação base e sejam, portanto, comparáveis (PRIM, 2010).

Ele defende que o movimento no DP (sintagma determinante, do inglês determiner phrase) em línguas românicas ocorre somente se os adjetivos são especificadores de projeções funcionais, caso contrário esse movimento não ocorre, o que pode ser resumido no seguinte esquema[2]:

(1)



[2] Retirado de: PRIM, 2015, p. 18.

O ponto interessante proposto por Cinque é que mesmo línguas com flexibilidade na ordenação dos adjetivos, mantêm a ordem das línguas germânicas pré-nominais, ou seja, temos a seguinte ordem (PRIM, 2015):

(2) a. Línguas [Adj N] → adjetivo avaliativo, tamanho, cor, **nome**
b. Línguas [N Adj] → **nome**, cor, tamanho, adjetivo avaliativo
c. Línguas [Adj N Adj] → adjetivo avaliativo, tamanho, **nome**, cor

Desse modo, como observa Crisma (1993, p. 82), Cinque mostra que há um "c-comando assimétrico nos argumentos do núcleo N, com os argumentos internos sempre mais baixos do que os externos, independente da ordem linear" (CRISMA, 1993, p. 82, tradução nossa)[3], ou seja, há um paralelismo entre línguas germânicas e românicas.

Essa observação é especialmente relevante no âmbito da tradução, pois propõe um modelo de espelhamento que pode ser útil ao criar parâmetros de tradução automática, como aqueles usados no GT.

Entretanto, Crisma (1993), embora concorde com Cinque que os "adjetivos entram nos DPs como especificadores de projeções funcionais específicas e, assim, são sempre XPs" (PRIM, 2010 p.33), acredita que o que rege a ordenação dos adjetivos é a interpretação, dessa forma, a relação entre a posição e a interpretação é direta (PRIM, 2010).

Crisma aponta algumas falhas de proposições tradicionais, como de Giorgi e Longobardi (1991), Valois (1991) e Zamparelli (1993), como o fato de que elas propõem que qualquer adjetivo pode ser colocado na posição pós-nominal em línguas germânicas, o que não ocorre; que não há diferença semântica em adjetivos que podem ocorrer em ambas as posições; e, além disso, esses autores não oferecem explicação para as regras de ocorrência de adjetivos pré-nominais, a não ser pela explicação de que eles seriam "ornamentais", o que não se sustenta, já que a omissão de um adjetivo pré-nominal implica a mesma perda semântica de um pós-nominal.

Como solução para esse problema, Crisma propõe, em paralelo com estruturas adverbiais, uma análise de adjetivos que se comportam de maneira semelhante em línguas românicas e

[3] C-comando (commando de constituinte) é um conceito criado pela relação binária entre nós em uma estrutura de árvore em que A é diferente de B, A não domina B e B não domina A e todo o X que domina A também domina B (CHOMSKY, 1986).

defende que a posição dos advérbios está intrinsecamente ligada à sua interpretação, que se baseia na classificação dos advérbios proposta por Belletti (1990, apud CRISMA, 1993).
Para tanto, ela analisa adjetivos ligados a nominais eventivos em quatro categorias: orientado para o falante e orientado para o sujeito, adjetivo de modo e a classe 'merely' para advérbios, que equivale a adjetivos que não se adequam nos outros grupos, como "mero".
Crisma (1993) acredita que adjetivos que permitem ambas as posições, mas que apresentam significados diferentes em cada uma das posições, na verdade têm uma posição fixa de acordo com sua interpretação, seja ela orientada para o falante ou orientada para o sujeito, assim:

> (...) em línguas românicas, um adjetivo que é compatível com uma interpretação orientada para modo e para o falante poderá aparecer antes e depois do substantivo, recebendo a primeira interpretação ao preceder o nome e a segunda ao segui-lo. (p. 90)

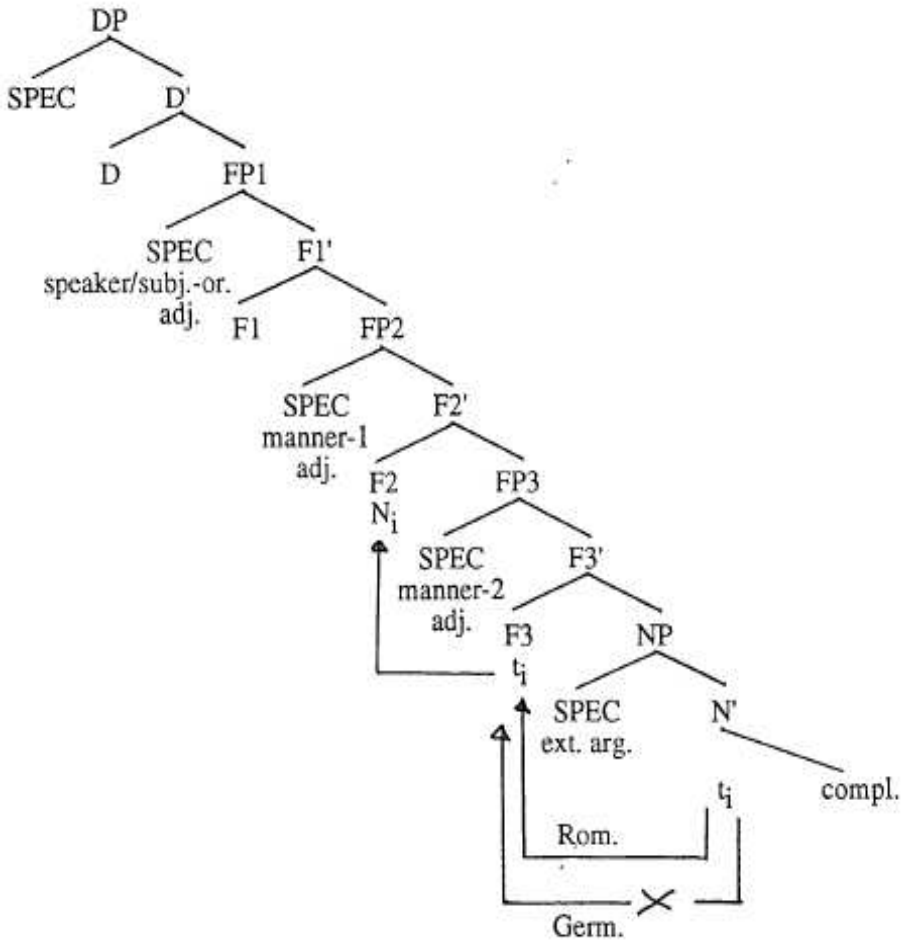
Para adjetivos que transitam livremente entre as posições pré e pós-nominais com pouca alteração de sentido, a autora acredita que a proposta tradicional de que na posição pré-nominal há uma preferência por interpretação explicativa enquanto na posição pós-nominal há uma leitura restritiva, é uma classificação que "trata do uso dos adjetivos, não de um atributo deles" (CRISMA, 1990, p.116), como mostrado no exemplo a seguir de Prim (2010, p. 114):

(3) As perfumadas flores (leitura não restritiva)
As flores perfumadas (leitura restritiva)

Para tanto, ela propõe que eles podem ser explicados dividindo os adjetivos de modo da seguinte forma: tipo 1 e tipo 2, no qual "adjetivos de modo tipo 2 podem ser usados em contextos genéricos e adjetivos de modo 1 só podem ser usados quando o DP se refere a um evento ou série de eventos específico" (CRISMA, 1993, p. 99).
Desse modo, ela propõe a seguinte estrutura no italiano[4]:

(4)

[4] Retirado de CRISMA, 1993, p. 97



Ela também destaca a classe de adjetivos “merely” (‘mero’, em português) que parece ter uma posição mais alta que o núcleo, embora pareça também competir quanto à posição com os adjetivos de modo de tipo 1.

Essa proposta, embora trate somente dos adjetivos ligados a nomes eventivos, traz vantagens para análise de dados de tradução automática, pois ela delimita uma ordem possível para os parâmetros da ferramenta de tradução e também oferece uma solução para o problema de sistematização da ordenação pré e pós-nominal, inclusive quando não há diferença de sentido entre as duas posições, que seria explicado pela posição de adjetivo de modo 2.

Assim, temos no inglês uma estrutura sintática dos adjetivos exclusivamente pré-nominal [Adj N] e no português temos uma estrutura preferencialmente pós-nominal, mas que admite adjetivação pré-nominal [Adj N Adj] e inclusive tem adjetivos exclusivamente pré-nominais.

Estreitando a análise para o português, vemos que o conjunto de adjetivos exclusivamente pré-nominais é o de menor prevalência e alguns estudos já até tentaram delimitá-lo, entre eles Muller, Negrão e Nunes-Pemberton (2002, apud PRIM, 2010), que chegaram ao seguinte

conjunto após a análise do corpus mínimo do Projeto da Gramática do Português Falado: último, diverso, excelente, verdadeiro, mau, suma, máxima, justas, respectivos, magnânimo, inúmeras, sucessivo, ilustre, extrema, possível, senhora, vil, célebre, constante e relevante. E ampliado por Prim (2010, p. 18): baita, bastante, bel, big, adjetivos cardinais, ledo, meio, mero, mesmo, pretenso, primeiro, segundo, etc, prisco, puta, reles, senhor, sumo, suposto, último, nesse.

Esses não são os únicos adjetivos exclusivamente pré-nominais do português, mas eles serão tomados como base nesta análise, além de outros que possam surgir ao examinar os dados.

Assim, temos no português adjetivos pré, pós-nominais e aqueles que podem ser colocados em ambas as posições. Entre os adjetivos restritos à posição pós-nominal, Prim (2010) descreve que há algumas categorias exclusivamente pós-nominais: adjetivos de cor, nacionalidades e determinados adjetivos terminados em –anso e –ense.

Além desses dois grupos, temos os adjetivos que transitam entre os dois grupos, que podem ser classificados da maneira proposta acima por Crisma como modo 2.

Considerando todas essas classificações e suas dificuldades, me proponho a analisar como a tradução automática lida com essas diversas variáveis e fatores e se a ferramenta se beneficiaria ou não na adoção de uma teoria gramatical sobre ordenação de adjetivos nos parâmetros de tradução automática baseada em estatística.

Além disso, pretendo observar também se a maior quantidade de adjetivos e, portanto, maior complexidade sintática da oração é um fator que interfere na tradução dos adjetivos ou quais outros fatores podem ser causadores de erros.

5. Metodologia

Neste projeto, para analisar a qualidade e as dificuldades da tradução automatizada de adjetivos do inglês para o português, foi usado um corpus com frases de escrita em um contexto real literário em uma ferramenta de tradução automática. A ferramenta escolhida foi o Google Tradutor e o corpus escolhido foi o livro Sherlock Holmes - Um estudo em vermelho.

Para esse fim, foi realizado um levantamento de todas as sentenças que possuem adjetivos atributivos ligados ao SN e elas foram inseridas na ferramenta.

A escolha do livro Sherlock Holmes foi feita pois proporcionava o uso de um corpus real e não frases fabricadas, por acreditar que os resultados se assemelhariam mais aos de uso real da ferramenta. Para tanto, procurei um livro que fosse amplamente conhecido, estivesse em domínio público e tivesse um estilo que propiciasse maior ocorrência de adjetivação.

Delimitei o corpus para somente os três primeiros capítulos do livro, que está disponível em domínio público, totalizando em aproximadamente 10.000 palavras.

Os erros de ordenação encontrados foram analisados de acordo com algumas características sintáticas e morfológicas, partindo do pressuposto de que um maior número de adjetivos ligados a um sintagma nominal e a presença ou não de adjetivos mais complexos (como os adjetivos compostos) possam ser fatores influenciadores na tradução automática. Os critérios para selecionar os contextos com adjetivos a serem analisados foram:

1. Mais de um adjetivo ligado ao SN
2. Adjetivos compostos
3. Mais de um adjetivo ligado ao SN com enumeradores (‘,’, ‘and’, ‘or’)

A distinção entre mais de um adjetivo ligado ao SN, adjetivos compostos e enumeradores foi importante, pois mudanças sutis na pesquisa em ferramentas de tradução, como texto em caixa alta, vírgulas, e a presença ou não de hifens e travessão parecem ser um fator importante na escolha lexical e sintática, como pode ser observado nos exemplos a seguir:

(5) a. There was nothing on it except a **tiny golden key**.[5]
    b. Não havia nada sobre ele, exceto uma **pequena chave de ouro**.

[5] Retirado de Alice's Adventures in Wonderland, de Lewis Carrol. Disponível em: https://www.adobe.com/be_en/active-use/pdf/Alice_in_Wonderland.pdf

(6) a. There was nothing on it except a **tiny, golden key**.

b. Não havia nada sobre isso, exceto uma **pequena chave dourada**.

Nesse excerto temos uma distinção estilística que não afeta diretamente o entendimento ou a semântica da frase; ainda assim é interessante notar como uma diferença mínima no texto fez com que houvesse uma escolha lexical diferente, ainda que só represente uma alteração no campo estilístico e não resulte em nenhum erro.

Além disso, levanta-se a hipótese de que a coordenação de um ou mais adjetivos pode ser um fator de dificuldade na tradução automática e que ele pode apontar dificuldades comuns em modelos de TA.

Os três tipos de contexto adjetival acima foram analisados quanto à presença de um ou mais dos seguintes erros:

1. Erro semântico (S/N)
   Por ex.: little pinch -> pouco pitada[6]
2. Palavra ausente (S/N)
   Por ex.: skilful workman -> artífice
3. Não tradução do adjetivo (S/N)
   Por ex.: ineffable twaddle -> twaddle inefável
4. Erro ortográfico (S/N)
   Por ex.: inconsequent remark -> observação inconseqüente
5. Erro na ordenação dos adjetivos (S/N)
   Por ex.: ungodly hours -> ímpios horas
6. Erro de concordância (S/N)
   Por ex.: mystified eyes -> olhos mistificado
7. Erro sintático (S/N)
   Por ex.: bottom steps -> etapas fundo

Essa distinção baseia-se nas classificações de Caseli e Nunes (2009), que discorrem sobre anotação manual de erro de TA. Elas se baseiam na classificação usada por Vilar et al., (2006), que separa os erros em uma tipologia de vários níveis, no nível mais alto dessa

[6] O corpus foi editado em alguns exemplos para maior fluência, as frases utilizadas nas análises estão disponíveis na íntegra no anexo, no fim desta monografia.

tipologia, temos: palavra ausente, ordem da palavra, palavra incorreta, palavra não conhecida e pontuação. Além desses parâmetros, que foram usados somente no âmbito do conjunto adjetivo + sintagma nominal, incluí também erros de concordância e outros erros sintáticos, que não fossem nem de concordância nem de ordenação, de forma a observar outros problemas que pudessem surgir.

Para observar melhor os erros sintáticos causados por frases complexas, optei por padronizar as buscas realizando a pesquisa com frases completas (independentemente do seu tamanho), de forma que não houvesse perda sintática. Mesmo se a frase em questão fosse muito grande, optei por mantê-la inteira, pois qualquer modificação é muito significativa durante a TA.

Além disso, foram destacados os adjetivos compostos. De acordo com McCarthy et al (2008, p.24) adjetivos compostos são duas ou mais palavras ligadas por um hífen que formam um adjetivo. Normalmente eles são formados por um adjetivo e um particípio, por exemplo: absent-minded, left-handed, curly-haired. Eles normalmente são usados para descrever aparência ou caráter do sujeito, mas há também casos em que eles são formados por preposições[7]:

(7) built-up area [cheia de prédios]
broken-down car [que não funciona]
built-in furniture [não pode ser removido]

Os adjetivos compostos têm como função facilitar a interpretação do texto e evitar ambiguidades; portanto, acredito que seja interessante observar se a ocorrência desse tipo de adjetivo provoca uma diferença no desempenho da tradução automática em relação a adjetivos simples. Além disso, observo se o hífen existente em adjetivos compostos fará com que haja uma leitura preferencial que é seguida corretamente ou se ele trará confusão na interpretação da máquina nessa situação de análise proposta.

Partindo do pressuposto, portanto, de que há adjetivos exclusivamente pré-nominais e adjetivos exclusivamente pós-nominais, para esta análise desconsiderarei diferenças de ordenação em adjetivos que não resultem em erros semânticos. Para tanto, não considerarei flutuações semânticas que possam ser causadas pela alteração na ordenação de adjetivos usadas em ambas as posições, considerando que essa ambiguidade também pode estar presente no inglês.

[7] Retirado de MCCARTHY, Michael; O'DELL, Felicity & SHAW, Ellen. Vocabulary in use - Upper Intermediate. 10ª edição. Nova York: Cambrigde, 2008. 311 p.

Por exemplo, a distinção entre “lindo homem” e “homem lindo” que possui uma distinção semântica restritiva ou não, não será considerada, mas a distinção entre “novo homem” e “homem novo”, que resulta em uma diferença de sentido mais clara fará parte da interpretação.

Entretanto, é importante notar também que essa movimentação com variação semântica dos adjetivos muitas vezes é resolvida no segmento de origem em uma tradução, delimitando o conjunto de possibilidades tradutórias. Por exemplo, no inglês a homonímia de (8) e (9) não é possível:

(8)

a. Um novo homem chegou à cidade

b. A new man came to town

(9)

a. Um homem novo chegou à cidade

b. A young man came to town

Assim, em uma tradução da frase (8b) só seria possível a interpretação (8a), assim como em (9a) e (9b).

Outra consideração importante em relação à posição do adjetivo (somente pré-nominal, somente pós, pós/pré sem perda de sentido ou pós/pré com alteração de sentido) é que há delimitações na ordenação causadas também pelo próprio texto de origem. Por exemplo:

(10)

a. Worn with pain, and weak from the prolonged hardships which I had undergone, I was removed, with a great train of **wounded sufferers**, to the base hospital at Peshawar.

b. Desgastado com dor, e fraco das dificuldades prolongadas que eu tinha sofrido, I foi removido, com uma grande comitiva dos **doentes feridos**, para o hospital de base em Peshawar.

Embora "feridos doentes" e "doentes feridos" sejam duas opções gramaticais na língua portuguesa, nesse caso apenas "doentes feridos" seria possível, devido à natureza do texto original que delimita o sentido, então temos dois fenômenos: o fato de que o adjetivo “ferido”

é pré e pós-nominal no português e de que há a delimitação dessa regra na tradução, devido ao universo possível de sentido da frase original.

Um fenômeno semelhante ocorre com “pleasant thing” em:

(11)

a. The sight of a friendly face in the great wilderness of London is a **pleasant thing** indeed to a lonely man.

b. A visão de um rosto amigável na grande deserto de Londres é uma **coisa agradável**, na verdade, um homem solitário.

Na frase, a tradução "agradável coisa" é agramatical, mas se adjetivo “agradável” em si estivesse ligado a outro nome, como "agradável homem" ele será aceito em ambas as posições.

Portanto, para analisar esses adjetivos precisamos levar em consideração que eles já estão delimitados de alguma forma, não cabendo aqui o conjunto total de opções do idioma, pois eles baseiam-se em uma língua de origem, cujas diferentes opções semânticas resultam em polissemias e homonímias também diferentes, fazendo com que haja um conjunto menor de possibilidades de ordenação.

6. Análise dos Dados

6.1. Ordenação

Após coletar e classificar todos os dados, comecei a análise pela observação da quantidade de acertos gerais entre os três parâmetros (um adjetivo ligado ao SN, dois ou mais adjetivos ligados ao SN ou dois ou mais adjetivos ligados ao SN com um enumerador). Em um total de 319 ocorrências, foram encontrados os seguintes resultados para cada parâmetro:

Tabela 1. Taxa de acerto geral.

	Acertos/Ocorrências	Taxa de acerto geral
SN + Adj1	184/234	78,63%
SN + Adj1 + Adj2...	43/64	67,18%
SN + Adj1 + Enumeradores + Adj2...	10/21	47,61%

Esse resultado sozinho parece indicar uma tendência de menor acerto conforme maior complexidade da frase, com destaque ao grupo de adjetivos com enumeradores, que apresenta uma taxa de acerto menor que 50%.

Observei, em seguida, se esse padrão se repetia em erros de ordenação. É importante notar, em primeiro lugar, que a grande maioria das ocorrências foi traduzida sem nenhum erro de ordenação: foram apenas 15 erros em 319 ocorrências. Isso indica que a ferramenta tem um mecanismo eficiente para a ordenação correta de adjetivos. Entretanto, continuei a análise para observar os erros que ocorreram e confirmar essa hipótese.

Assim, os erros de ordenação foram distribuídos da seguinte maneira conforme os fatores de análise:

Tabela 2. Erros de ordenação de acordo com os três fatores de análise.

Tipo de adjetivação	Número de erros/total de segmentos	Taxa de erro de ordenação
SN + Adj1	5/234	2,13%

SN + Adj1 + Adj2...	4/64	6,25%
SN + Adj1 + Enumeradores + Adj2...	6/21	28,57%

Alguns exemplos desses erros são:

(12)

- SN + Adj1

  ungodly hours → ímpios horas

  my late habits → minha tarde hábitos

- SN + Adj1 + Adj2...

  stalwart police constable → policial robusto polícia

  practical medico-legal discovery → descoberta mais prática médico-legal

- SN + Adj1 + E + Adj2...

  pompous and self-satisfied manner → pomposo e auto forma satisfeito

  writhing, unnatural posture → contorção, postura anormal

Quando a prevalência desses erros é analisada, nota-se novamente que há maior tendência de erro em segmentos com maior complexidade, mas é importante notar que a diferença aqui é muito mais significativa. Enquanto apenas 2,13% dos adjetivos ligados ao SN têm erros e 6,25% das orações com mais de um adjetivo, o número dispara para 28,57% quando há enumeradores. Isso indica que o fator de enumeração é relevante para a análise dos erros.

É possível observar, portanto, que enquanto a ocorrência de erros com adjetivação direta apresenta uma quantidade de erros relativamente parecida, a presença de um separador, seja uma vírgula ou a palavra 'and', os erros aumentam significantemente. Isso pode significar que esses erros são decorrentes de um não reconhecimento da função sintática de cada elemento quando há uma complexidade maior no segmento e não exatamente de um erro de ordenação. É importante notar que esse não reconhecimento pode ser decorrente do fato que o GT não tem mecanismos de reconhecimento da função sintática, pois apenas baseia-se em frases semelhantes e probabilidades de ocorrência e, portanto, quanto mais complexa a frase mais difícil é ele se basear em uma ocorrência anterior..

Entretanto, é preciso destacar novamente que os erros de ordenação apontados são pontuais, já que eles representam apenas 4,7% dos erros totais e que há diversos exemplos em que essas construções foram traduzidas corretamente, mesmo quando muito complexas como, por exemplo, em:

(13)

a. pure-blooded well-trained foxhound → foxhound bem treinado puro-sangue

b. old white-haired gentleman → velho cavalheiro de cabelos brancos

c. grey-headed, seedy visitor → visitante de cabelos grisalhos, decadente

d. studious and quiet habits → hábitos de estudo e tranquilos

e. red wax candle → vela de cera vermelha

f. imitation white marble → imitação de mármore branco

Entre esses acertos, destacam-se "vela de cera vermelha" e "imitação de mármore branco" que sequem a ordenação normalmente encontrada no português:

> Opinião geral > opinião específica > dimensão > formato > idade > cor > nacionalidade > material > nome (MOREIRA, 2015, p.143)

Desse modo, embora não seja o propósito deste trabalho analisar ordenações em que não há agramaticalidade ou perda de sentido, o GT parece seguir a ordem preferencial do português brasileiro, mesmo quando ela é apenas preferencial, não obrigatória.

Feito esse parênteses, os erros de ordenação foram analisados para tentar explicar suas causas e foi observado que há muitos casos em que erros semânticos são causadores dos erros de ordenação. Um exemplo disso é:

(14)

a. They consisted of a couple of **comfortable bed-rooms** and a single large airy sitting-room, cheerfully furnished, and illuminated by two broad windows.

b. Eles consistiu de um par de **confortáveis quartos de cama** e uma sala de estar-single amplo e arejado, alegremente decorados, e iluminado por duas grandes janelas.

A tradução incorreta de “bed-rooms” causa um erro em todo o segmento, fazendo com que o adjetivo “confortável”, que em português é pós-nominal, seja colocado em posição pré-nominal, como consequência da tradução não adequada do substantivo. Essa hipótese é fortalecida, quando alteramos o substantivo para sua ortografia mais usual, “bedrooms”, e obtém-se o seguinte resultado:

(15) Eles consistiu de um par de **quartos confortáveis** e uma sala de estar-single amplo e arejado, alegremente decorados, e iluminado por duas grandes janelas.

Assim, quando o substantivo é traduzido corretamente, o erro de ordenação não ocorre mais e “confortáveis” é colocado em uma posição gramatical na oração. Podemos observar que erros semânticos resultam em erro de ordenação, mesmo que esses erros semânticos sejam decorrentes, por sua vez, de erros ortográficos.

Outro erro observado que decorre de um erro semântico é “Russian leather card-case”, traduzido como “couro russo cartão de caso”. No segmento houve o erro de tradução de “card-case”, que significa “caixa para cartão”. O não reconhecimento do termo parece acarretar em um erro de sintaxe, em que ‘card-case’ não é reconhecido como o sujeito da oração e não há coordenação entre os itens.

Realizei o mesmo procedimento com esse termo, realizando a busca para “card case” sem o hífen e obtive: “caso de cartão de couro da Rússia”, nesse caso também há uma melhora da tradução e a ordenação está correta, mas o termo continua incorreto, já que “caso” não é uma escolha lexical adequada.

Assim, é possível perceber que embora o erro ortográfico contribua para uma maior dificuldade de interpretação mesmo em palavras conhecidas e comuns, como “bedroom”, ele não é um fator determinante de erro, já que a correção do texto nem sempre resulta em uma interpretação correta da palavra.

Foi observado também que 53,3% dos segmentos que apresentam erros de ordenação também apresentaram algum erro semântico, enquanto a média geral de erros semânticos é de apenas 10,8% nos segmentos sem nenhum erro de ordenação. Isso faz sentido quando consideramos a proposta de Crisma, de que a interpretação do adjetivo que determina sua ordem na oração, de forma que, quando o GT não consegue analisar adequadamente o significado do adjetivo, sua ordenação fica comprometida.

Portanto parece que temos um movimento de [erro semântico] → [erro de ordenação] e não o contrário, o que leva à hipótese de que a ferramenta de máquina tenha um grande índice de acerto sobre ordenação de adjetivos quando em condições ideais, ou seja, quando todos os componentes semânticos são interpretados corretamente.

Além disso, como é possível observar na distribuição dos erros de ordenação, fatores sintáticos como a coordenação de adjetivos e a presença ou não de "and" ou "or" também são fatores relevantes no que diz respeito à ordenação. Por exemplo:

(16)

a. This malignant and terrible contortion, combined with the low forehead, blunt nose, and prognathous jaw gave the dead man a singularly simious and ape-like appearance, which was increased by his **writhing, unnatural posture**.
b. Esta contorção maligna e terrível, combinado com a baixa testa, nariz sem corte, e da mandíbula prognata deu o morto uma aparência singularmente simious e de macacos, que foi aumentada por sua **contorção, postura anormal**.

Nesse exemplo, o GT não reconhece que "writhing" e "unnatural" estão coordenados em relação ao substantivo e isso resulta em um erro de ordenação. Outro aspecto que confirma essa hipótese é que "unnatural", que é um adjetivo apenas pós-nominal, está ordenado corretamente.

Outro exemplo representativo ocorre a seguir. Embora os dois adjetivos possam estar nas mesmas posições, a coordenação faz com que a única posição possível seja a pós-nominal, pois não há coordenação pré-nominal:

(17)

a. Like all other arts, the Science of Deduction and Analysis is one which can only be acquired by **long and patient study** nor is life long enough to allow any mortal to attain the highest possible perfection in it.

b. Como todas as outras artes, a ciência da dedução e Análise é aquela que só pode ser adquirida por **longa e paciente estudo** nem é vida longa o suficiente para permitir que qualquer mortal para alcançar a perfeição máxima possível na mesma.

Nesse caso também parece haver um não reconhecimento de que ambos os adjetivos estão ligados ao mesmo SN, o que pode ser confirmado por “longa” estar no feminino, concordando com “análise”. Esse erro, assim como no caso anterior, faz com que haja uma ordenação não gramatical.

No único outro erro com ordenação [+AND], essa ordenação pré-nominal se repete, como em “pompous and self-satisfied manner” que é traduzido como “pomposo e auto forma satisfeito”.

Analisei também se havia diferença na taxa de acertos entre frases com adjetivos separados somente por vírgula [-AND] e frases separadas por “and” ou “or” [+AND]. Encontrei 4 erros de ordenação em 8 ocorrências do fator (50% de taxa de erro) em [-AND] e 2 erros entre 13 ocorrências de [+AND] (15,3% de taxa de erros).

Dessa forma, podemos observar que, além de fatores semânticos, o fator de enumeração, em especial enumerações realizadas com vírgula, é o maior fator de confusão da ferramenta, que não consegue diferenciar com tanta eficácia quando a vírgula é um enumerador e quando ela tem outras funções na frase.

Portanto, é seguro afirmar que, assim como dificuldades lexicais, as enumerações, embora sejam facilitadores da interpretação em determinados casos, são também um fator de erro em traduções do GT e isso pode ocorrer pois a máquina tem a dificuldade de compreender orações complexas, que não podem ser são facilmente assimiladas por aprendizado de máquina pois necessitam de elementos contextuais para serem compreendidas.

Também entre os erros de ordenação, destaco o seguinte erro:

(18)

a. Its **somewhat ambitious title** was "The Book of Life," and it attempted to show how much an observant man might learn by an accurate and systematic examination of all that came in his way.

b. É um **pouco ambicioso título** era "O Livro da Vida", e tentou mostrar o quanto um homem observador pode aprender através de um exame preciso e sistemático de tudo o que veio em seu caminho.

Nesse caso, o que parece trazer confusão no texto é o pronome possessivo “Its” que é traduzido como pronome + verbo “It is”. Desse modo, mudei o pronome para “his” e realizei novamente a pesquisa, obtendo:

(19) Seu **título um tanto ambicioso** foi "O Livro da Vida", (...)

Assim, é possível observar novamente que o problema de ordenação pode decorrer de outros fatores no processamento do texto. No caso, há novamente um problema sintático que é decorrente de um problema semântico, já que o pronome “Its” foi interpretado como “It’s”, o que faz com que a ferramenta não consiga processar o texto e traduzi-lo com a ordem adequada.

Outro fator que reforça a hipótese de que erros semânticos e sintáticos são causadores dos erros de ordenação é que foi encontrada apenas uma ocorrência sem nenhum erro sintático ou semântico nos 15 erros de ordenação e nas 319 ocorrências de adjetivação analisadas:

(20)

a. At present my attention was centred upon the **single grim motionless figure** which lay stretched upon the boards, with vacant sightless eyes staring up at the discoloured ceiling.

b. Neste momento a minha atenção foi centrada sobre a **figura imóvel sombria única** que estava estendido sobre as placas, com olhos cegos vagos olhando para o teto descolorido.

A palavra “única”, como observado anteriormente só poderia ser colocada na posição pós-nominal se tivesse o sentido de “unique”, o que não é o caso. Ainda assim, é um erro “sutil” quando observamos os outros erros destacados, pois ele não prejudica totalmente o entendimento do texto e a frase de modo geral está quase perfeita em termos de sentido.

Embora esse erro pudesse ter sido beneficiado com a inserção de regras de ordenação nos parâmetros base da ferramenta, por ser apenas um erro em todo o corpus, essa afirmação não se justifica, já que a ferramenta analisada apresenta alta taxa de acerto no quesito ordenação, especialmente quando são excluídos erros de ordenação causados por outros fatores, que representam 93,3% dos erros.

Portanto, tanto o número como a análise dos erros parece reforçar a teoria de Crisma de que a interpretação é fundamental para a ordenação de adjetivos. Além disso, indica também que não há necessidade de mais parâmetros gramaticais no que diz respeito à ordenação de adjetivos para a finalidade e o uso do GT.

A análise revelou que outros erros são responsáveis por erros de ordenação, principalmente os erros de semântica e erros na interpretação de uma estrutura coordenada na enumeração de adjetivos. Esses dois tipos de erros geram um "efeito cascata" que resulta em erros de ordenação e em traduções "bag-of-words" (Specia, 2007), nas quais as palavras são apenas traduzidas e não se ligam sintaticamente.

Assim, é possível observar que os erros de ordenação em adjetivos atributivos ocorrem principalmente com a combinação de outros erros, especialmente semânticos, que geram um efeito cascata que compromete toda a tradução do segmento.

Quanto à análise dos adjetivos compostos, observa-se que em 26,6% dos erros de enumeração são encontrados em adjetivos compostos. Essa descoberta segue uma tendência de maior erro conforme maior complexidade da oração. Observam-se também 2 casos em que o termo composto é o substantivo, ambos apresentando erros de ordenação.

Esses erros estão na mesma oração e são:

(21)

a. They consisted of a couple of **comfortable bed-rooms** and a **single large airy sitting-**room, cheerfully furnished, and illuminated by two broad windows.

b. Eles consistiu de um par de **confortáveis quartos de cama** e uma **sala de estar-single amplo e arejado**, alegremente decorados, e iluminado por duas grandes janelas.

Nos dois casos, é importante notar, a grafia usada não é a usual e isso pode ter contribuído para os erros de ordenação, inclusive porque no segundo caso o hífen que não deveria estar na tradução é colocado de forma inadequada e provavelmente é o motivo que "single" não foi traduzido.

Além disso, o adjetivo "único" em português tem mudança de sentido nas posições pré e pós-nominal. Quando ele ocorre na posição pré-nominal normalmente é interpretado com o sentido de "single" (sozinho) e quando ocorre na posição pós-nominal tem o sentido de

“unique” (exclusivo). Dessa forma, esse seria um caso em que a posição do adjetivo é pré-determinada pelo texto de partida, mas a não tradução de “single” impossibilita a ordenação adequada.

Entretanto, é importante notar que os erros de ordenação em que há a presença de adjetivos compostos, nem sempre esses erros têm uma correlação direta, como é possível ver nos exemplos abaixo:

(22)

a. “practical medico-legal discovery” → “descoberta mais prática médico-legal”

b. “one little sallow rat-faced, dark-eyed fellow” → “pequena andorinha, companheiro de olhos escuros com cara de rato”

Nos dois casos, os adjetivos compostos foram traduzidos corretamente e não é possível afirmar que a presença ou não de um adjetivo composto seja o fator de erro nesses casos. Portanto, não é possível afirmar que a presença de adjetivos compostos contribua para maior ou menor prevalência de erros, embora eles contribuam para maior complexidade da frase, especialmente quando a grafia usada não é a usual.

6.2 Concordância

Embora a análise tenha se concentrado na ordenação de adjetivos, é importante destacar outro achado desta investigação: os erros de concordância encontrados no corpus. Ao contrário dos erros de ordenação, que decorrem principalmente de erros sintáticos e semânticos, 56% dos segmentos com erros de concordância não apresentam nenhum outro erro.

Observando alguns desses casos, constata-se que não é possível encontrar qualquer justificativa para eles e também que na maioria das vezes eles estão ligados a substantivos comuns na língua portuguesa, como “limites”, “olhos”, “leitores”, “expressão”, o que deveria indicar que o banco de dados do GT teria condições de indicar o gênero dessas palavras. Observa-se também que não há problemas com a concordância dos artigos.

(23)

a. Yet his zeal for certain studies was remarkable, and within **eccentric limits** his knowledge was so extraordinarily ample and minute that his observations have fairly astounded me.

b. No entanto, o seu zelo para certos estudos foi notável, e dentro dos **limites excêntricas** seu conhecimento era tão extraordinariamente ampla e minuto que suas observações foram bastante me surpreendeu.

(24)

a. Lestrade grabbed it up and stared at it with **mystified eyes.**

b. Lestrade agarrou-se e olhou para ele com **os olhos mistificado**.

(25)

a. The same afternoon brought a grey-headed, seedy visitor, looking like a Jew pedlar, who appeared to me to be much excited, and who was closely followed by a **slip-shod elderly woman**.

b. Na mesma tarde trouxe um visitante de cabelos grisalhos, decadente, parecendo um mascate judeu, que me pareceu ser muito animado, e que foi seguido de perto por uma **mulher idosa desleixado**.

É interessante notar que, no GT, ao clicar na tradução da palavra aparecem novas opções de tradução para ela e ao buscar somente os termos destacados e observando essas novas opções são encontrados alguns dados interessantes: em todos os casos, ao clicar na palavra obtive uma ou mais opções referentes à concordância.

Por exemplo, ao pesquisar somente “slip-shod elderly woman”, ao clicar na entrada “slip-shod”, são encontradas as opções: “slipshod”, “indolente”, “descuidada”, “desleixada”, “desleixado”. Assim, embora “mulher” seja uma palavra exclusivamente feminina, o GT oferece pelo menos uma opção masculina, o que por si só já seria uma escolha inadequada.

Entretanto, quando realizamos o mesmo procedimento em “idosa” vemos que “mulher idosa” é tratado como uma única unidade e as opções disponíveis são: “idosa” e “senhora idosa”, ambas escolhas no feminino.

Esse fato pode ocorrer por dois motivos: um é sintático, por “elderly” estar ligado diretamente a “woman” o GT considera que ele só tem a opção feminina, mas ele não consegue determinar a mesma relação entre “slip-shod” a “woman”, talvez devido à distância. Outra hipótese é a de que não há ocorrências suficientes em “slip-shod” para fornecer uma tradução adequada.

Realizando a pesquisa com somente “slip-shod woman” encontra-se o resultado “mulher desajeitada”, o que favorece a primeira hipótese. Entretanto, analisando os dados em que há apenas um adjetivo ligado ao SN, a hipóteses do corpus ser insuficiente é favorecida, pois não é possível atribuir esse erro à maior complexidade gramatical.

Outra hipótese possível é que o GT pesquise em seu banco de dados em primeiro lugar toda a frase e quando não encontra ocorrências ele passa a pesquisar individualmente as palavras, levando em conta mais uma vez a prevalência e não suas características morfológicas (gênero/número), realizando assim uma seleção puramente estatística.

Dessa forma, é possível que o erro seja decorrente novamente de erros sintáticos, como visto em “slip-shod” ou de uma seleção morfológica das palavras baseada somente em estatística, fazendo com que palavras com maior número de entradas ou frequência maior sejam conjugadas adequadamente, já que a ferramenta tem o input necessário para tal. Esses resultados sugerem também que não há uma etapa de pós-processamento da tradução, que eliminaria esses erros de concordância simples.

Entretanto, como o propósito do meu trabalho não é analisar a concordância, não me aprofundarei nesse assunto e seria necessária uma investigação mais detalhada para afirmar qual das duas hipóteses é a correta, mas esse parece ser um problema relevante na tradução automática estatística e que pode se beneficiar de maior marcação gramatical.

7. Conclusão

Após analisar os dados, observa-se que o principal problema encontrado na ordenação de adjetivos na tradução automática diz respeito a erros semânticos e a estruturas de coordenação, pois a ferramenta tem dificuldade de realizar uma seleção semântica adequada ao contexto e de compreender quando há estruturas coordenadas.

Desse modo, a hipótese original de que o GT teria dificuldade em compreender a ordenação de adjetivos e sequências de adjetivos ligados ao SN e reorganizá-los seguindo a ordem adequada do português brasileiro não se concretizou, já que há uma taxa de acerto nesse quesito de 95,29% usando dados naturais sem um controle estrito de quantidade e ordenamento.

Dessa forma, o questionamento proposto se a ferramenta se beneficiaria de uma maior utilização de marcações gramaticais nos adjetivos seguindo uma teoria gramatical, como a teoria proposta por Crisma (1993), embora esteja de acordo com o que foi descoberto, que erros semânticos resultam em erros de ordenação, não parece ser necessário na ferramenta estudada, pois ela já apresenta uma margem de erro muito baixa para o uso proposto da ferramenta.

O sucesso da ferramenta para ordenar adjetivos provavelmente se deve ao modelo de tradução automática por estatística, que os ordena seguindo a ordem que tem maior prevalência no corpus, fazendo com que não seja necessário um marcador gramatical nesses casos.

É interessante que esse modelo pode ser mais eficiente estatisticamente até do que os modelos gramaticais propostos até o momento, já que, embora diversos teóricos tentem explicar o fenômeno, especialmente Cinque e Crisma que são a base teórica desse trabalho, nenhuma teoria abrange todos os casos e exceções possíveis nas línguas e esse problema parece ter sido bem resolvido na tradução automática estatística, por considerar a prevalência da posição nominal usando dados reais e estatística. Assim, a ferramenta tem sucesso, pois "imita" de certa forma a intuição do falante ao ordenar adjetivos, usando as opções com maior frequência estatística.

Entretanto foi possível observar alguns outros fatores que apresentaram maior dificuldade e que resultaram indiretamente em erros de ordenação nos segmentos traduzidos pelo GT.

Em primeiro lugar, foi possível observar que erros de ordenação decorrem principalmente de erros semânticos ou de palavras desconhecidas, que resultam em um "efeito cascata" de erros. Esses erros semânticos decorrem de dois fenômenos: desconhecimento da palavra, o que resulta em não tradução, ou escolha incorreta do léxico para determinado contexto. O desconhecimento da palavra pode ser resolvido com ampliação do léxico, inserções realizadas por usuários e maior interação com dicionários.

Para solucionar o problema de escolha de léxico, destaca-se a tese de Specia (2007), que propõe uma maneira mais eficaz para desambiguação lexical em sistemas de tradução automática. De acordo com Specia, a maior parte das abordagens híbridas tem um conhecimento linguístico superficial, pois há uma dificuldade de "integrar conhecimento linguístico profundo com os algoritmos de aprendizado de máquina tradicionalmente utilizados para DLS [Desambiguação Lexical de Sentido]" (p.4).

Assim, ela propõe um novo modelo de desambiguação lexical com base em Programação de Lógica Indutiva (PLI) que,

> Ao contrário da lógica computacional tradicional, que é baseada na inferência por dedução a partir de fórmulas providas por um usuário, a PLI investiga a indução como mecanismo de inferência, visando obter programas lógicos a partir de exemplos e de conhecimento de fundo. (p. 152)

Quanto à coordenação de adjetivos, embora não tenha encontrado nenhuma pesquisa especificamente sobre esse assunto, a complexidade sintática de orações é um dos fatores mais problemáticos da tradução automática. De acordo com Zamagni (2014), isso ocorre pois a interpretação depende de vários fatos que não são puramente linguísticos, como informações "para-linguísticas, não linguísticas e contextuais" (2014, p. 41).

Assim, para solucionar esse problema, seria necessário ensinar à máquina como captar essas informações paralinguísticas, o que não é simples. Uma proposta interessante é a de Piruzelli (2011) que, embora não tenha tratado especificamente de enumeração, também analisa os erros sintáticos comuns e propõe algumas soluções que utilizam maior marcação gramatical para resolver esses problemas, como por exemplo, a solução para problemas de ambiguidade anafórica ou referencial. Ele toma como base o modelo de Mitkov et al. (1995) e propõe que para a resolução desses problemas seja usada (i) maior marcação de número/pessoa/gênero,

> (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 papéis 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. (PIRUZELLI, 2011, p.99).

Entre esses fatores, o conceito de restrição semântica pode ser benéfico no caso de enumerações, já que uma maior correlação semântica dos componentes da frase pode beneficiar aspectos sintáticos. Por exemplo, em "pompous and self-satisfied manner" ou "long and patient study**"**, a tradução seria beneficiada se a ferramenta marcasse a relação semântica entre "long" e "patient" e "pompous" e "self-satisfied", classificando-os como adjetivos, por exemplo, que provavelmente fazem parte da mesma estrutura. Essa hipótese, entretanto, não foi testada nem analisada com profundidade e este trabalho limita-se apenas a elencá-la como uma possibilidade a ser vista no futuro.

Por fim, sobre os erros de concordância, com base na hipótese que a marcação de nome/número seja formada apenas por análise estatística e não por marcadores morfológicos em cada palavra, alguns erros encontrados também poderiam se beneficiar de maior marcação gramatical, já que a etapa de pré-processamento dos textos que define os parâmetros base para aprendizado de máquina delimita determinadas regras (SPECIA, 2007, p. 4), fazendo com que o sistema fique limitado a modelos que não são aprofundados, uma vez que é fornecido apenas um conhecimento superficial da língua.

Desse modo, embora o fator de ordenação de adjetivos seja bem resolvido na tradução automática, ainda há um longo caminho a ser traçado e a tradução automática ainda tem um longo percurso antes que possa se assemelhar verdadeiramente a uma tradução humana.

8. Referências

ARNOLD, D., et al. **Machine translation: An introductory guide**. Oxford: NCC Blackwell, 1994. Acesso em 16 de outubro de 2016. Disponível em: < http://promethee.philo.ulg.ac.be/engdep1/download/bacIII/Arnold%20et%20al%20Machine%20Translation.pdf>.

BELLETTI, A. **Generalized Verb-Movement**. Turim: Rosenberg & Sellier, 1990 apud CRISMA, P. On Adjective Placement in Romance and Germanic Event Nominals. In: Rivista di Grammatica Generativa, 18:61-100 (1993). Acesso em: 5 de outubro 2016. Disponível em: <http://arcaold.unive.it/bitstream/10278/410/1/3.4.pdf>.

CASELI, H. M.. **Indução de léxicos bilíngües e regras para a tradução automática**. 2007, 158 p. Tese (Doutorado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2007. Acesso em: 10 outubro de 2016. Disponível em: http://www.nilc.icmc.usp.br/nilc/download/tese_HelenaCaseli_defendida.pdf>.

_____, H.M.; NUNES, I.A. (2009). **Tradução Automática Estatística baseada em Frases e Fatorada:** Experimentos com os idiomas Português do Brasil e Inglês usando o toolkit Moses (NILC-TR-09-07). Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional (NILC), 2009. Acesso em: 5 de outubro de 2016. Disponível em: <http://www.nilc.icmc.usp.br/nilc/download/NILCTR-09-07.pdf>.

CHOMSKY, N. **Knowledge of Language:** Its Nature, Origin, and Use. Londres: Greenwood Publishing Group, 1986.

CINQUE, G. **Types of A' dependencies**. Cambridge, MA: MIT Press, 1990, apud PRIM, Cristina de Souza. A sintaxe de adjetivos nas posições pré- e pós-nominal. 2010. 109p. Dissertação (Mestrado em Linguística) – Universidade Federal de Santa Catarina, Florianópolis, 2010.

_____. **On the Evidence for Partial N-movement in the Romance DP**. University of Venice Working Papers in Linguistics, 3.2, 21-40. Veneza: Centro Linguistico Interfaculta, 1993 apud PRIM, C. de S. A sintaxe de adjetivos nas posições pré- e pós-nominal. 2010. 109p. Dissertação (Mestrado em Linguística) – Universidade Federal de Santa Catarina, Florianópolis, 2010.

_____. **Typological Studies: Word Order and Relative Clause**s. (Routledge leading linguists) New York: Routledge, 2013. 400 p.

_____; RIZZI, L. **The Cartography of Syntactic Structures**. Working Papers, STiL - Studies in Linguistics (CISCL WP, Vol.2). Acesso em: 5 outubro 2016. Retirado de: <http://www.ciscl.unisi.it/doc/doc_pub/cinque-rizzi2008-The_cartography_of_Syntactic_Structures.pdf>.

CRISMA, P. **On Adjective Placement in Romance and Germanic Event Nominals**. In: Rivista di Grammatica Generativa, 18:61-100, 1993. Acesso em: 5 de outubro 2016. Disponível em: <http://arcaold.unive.it/bitstream/10278/410/1/3.4.pdf>.

DOYLE, A. C. **A study in Scarlet**. 1887. Acesso em: 5 de outubro 2016. Disponível em: <https://sherlock-holm.es/stories/html/stud.html>.

GOOGLE. Inside Google Translate. 2010. Vídeo. Disponível em <https://www.youtube.com/watch?v=_GdSC1Z1Kzs>.

HUTCHINS, W. J.; SOMERS, H. L. **An introduction to machine translation**. London: Academic Press, 1992.

LOPEZ, A. **Survey of Statistical Machine Translation**. Prince George: University of Maryland, 2007. Acesso em: 5 de outubro 2016. Disponível em: < http://www.dtic.mil/dtic/tr/fulltext/u2/a466330.pdf>.

IBRAHIIMO, N. **Para uma Tradução Automática baseada em Conhecimento:** especificação da modificação e da predicação adjectival. 2010, 110 p. Dissertação (Mestrado em Tradução). Faculdade de Letras, Universidade de Lisboa, Lisboa: 2010.
Acesso em: 5 de outubro 2016. Disponível em: < http://repositorio.ul.pt/bitstream/10451/4125/1/ulfl081176_tm.pdf>.
KOERNER, E. F. K.; ASHER, R. E. (1995). **Concise history of the language sciences:** from the Sumerians to the cognitivists, p. 431–445. Pergamon Press, Oxford. apud CASELI, H.M.; NUNES, I.A. (2009). **Tradução Automática Estatística baseada em Frases e Fatorada:** Experimentos com os idiomas Português do Brasil e Inglês usando o toolkit Moses (NILC-TR-09-07). Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional (NILC), 2009. Acesso em: 5 de outubro de 2016. Disponível em: <http://www.nilc.icmc.usp.br/nilc/download/NILCTR-09-07.pdf>.
MCCARTHY, M.; O'DELL, F.; SHAW, E. **Vocabulary in use - Upper Intermediate**. 10ª edição. Nova York: Cambrigde, 2008. 311 p.
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 edition, Lovaina: Katholieke Universiteit, 1995. p. 87-95. Acesso em: 5 de outubro de 2010. Disponível em: < http://www.mt-archive.info/TMI-1995-Mitkov.pdf>.
MOREIRA, T. L. D. **A Sintaxe dos Adjetivos Atributivos**. 2015, 214 p. Dissertação (Mestrado) – Setor de Ciências Humanas, Universidade Federal do Paraná. Curitiba, 2015. Acesso em: 5 outubro 2016. Disponível em: < http://acervodigital.ufpr.br/bitstream/handle/1884/41196/R%20-%20D%20-%20THAIS%20LUISA%20DESCHAMPS%20MOREIRA.pdf?sequence=2&isAllowed=y>.
MULLER, A.; NEGRÃO, E.; NUNES-PEMBERTON, G. **Adjetivos no português do Brasil:** predicados, argumentos ou quantificadores? In: ABAURRE, Maria Bernadete M.; RODRIGUES, Angela C. S. (Org.). Gramática do Português Falado: novos estudos descritivos. Campinas: Editora da UNICAMP, 2002, v. VIII, p. 317-344 apud PRIM, C. de S. A sintaxe de adjetivos nas posições pré- e pós-nominal. 2010. 109p. Dissertação (Mestrado em Linguística) – Universidade Federal de Santa Catarina, Florianópolis, 2010.
PIRUZELLI, M. P. F. **Ambiguidades Linguísticas no Inglês e a Tradução Automática Inglês-Português:** um estudo exploratório. 2011, 156 p. Dissertação (Mestrado em Tradução). Faculdade de Ciências e Letras, Universidade Estadual Paulista. Araraquara, 2011. Acesso em 17 de outubro de 2016. Disponível em: < http://portal.fclar.unesp.br/poslinpor/teses/Maria_Paula_Piruzelli.pdf>
PRIM, C. S. **A sintaxe de adjetivos nas posições pré- e pós-nominal**. 2010. 109p. Dissertação (Mestrado em Linguística) – Universidade Federal de Santa Catarina, Florianópolis, 2010.
_____. **A Sintaxe dos Adjetivos em Português Brasileiro**. 2015. 175p. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, Campinas, 2015.
SPECIA, L. **Uma abordagem híbrida relacional para a desambiguação lexical de sentido na tradução automática**. 2007, Instituto de Ciências Matemáticas e de Computação, Universidade do Estado de São Paulo. Tese (Doutorado em Ciências de Computação e Matemática Computacional). São Carlos, 2007. Disponível em: < http://www.teses.usp.br/teses/disponiveis/55/55134/tde-05122007-205308/pt-br.php>.
VILAR, D.; et al. **Error analysis of statistical machine translation output**. Madri: Departamento de Engenharia eletrônica, Universidade Politécnica de Madrid, 2006. Acesso em 17 de outubro de 2016. Disponível em: <https://www-i6.informatik.rwthaachen.de/publications/download/281/VilarDavidXuJiaDHaroLuisFernoNeyHermann--%7BErrorAnalysisofMachineTranslationOutput%7D--2006.pdf>.

ZAMAGNI, A. **Italiano controlado para a tradução automática (italiano – português) Linguagem especializada:** informática. 2014, 136 p. Dissertação (Mestrado em Tradução). Faculdade de Letras, Universidade de Lisboa. Lisboa, 2014. Acesso em 17 de outubro de 2016. Disponível em: < http://repositorio.ul.pt/bitstream/10451/18390/1/ulfl179137_tm.pdf>.

9. Anexo

Tabela 3. Resultados utilizados na análise

Texto original	Texto traduzido pelo GT
"What ineffable twaddle!" I cried, slapping the magazine down on the table, "I never read such rubbish in my life."	"O que twaddle inefável!" Eu chorei, batendo a revista sobre a mesa ", eu nunca li tais lixo em minha vida."
I could imagine his giving a friend a little pinch of the latest vegetable alkaloid, not out of malevolence, you understand, but simply out of a spirit of inquiry in order to have an accurate idea of the effects.	Eu podia imaginar a sua dando um amigo um pouco pitada do último alcalóide vegetal, não por maldade, você entende, mas simplesmente fora de um espírito de investigação, a fim de ter uma ideia precisa dos efeitos.
Now the skilful workman is very careful indeed as to what he takes into his brain-attic.	Agora, o artífice é muito cuidadosos sobre o que ele tem em seu cérebro-sótão.
With which inconsequent remark he strode on into the house, followed by Gregson, whose features expressed his astonishment.	Com o qual a observação inconseqüente ele caminhou sobre a casa, seguido por Gregson, cujas características expressou seu espanto.
I laughed at this cross-examination. "I keep a bull pup," I said, "and I object to rows because my nerves are shaken, and I get up at all sorts of ungodly hours, and I am extremely lazy.	Eu ri neste interrogatório. "Eu mantenho um cachorro touro," eu disse, "e eu opor-se a linhas, porque os meus nervos estão abalados, e eu me levanto em todos os tipos de ímpios horas, e eu sou extremamente preguiçoso.
Lestrade grabbed it up and stared at it with mystified eyes.	Lestrade agarrou-se e olhou para ele com os olhos mistificado.
"We have it all here," said Gregson, pointing to a litter of objects upon one of the bottom steps of the stairs.	"Temos tudo aqui", disse Gregson, apontando para uma ninhada de objetos sobre uma das etapas fundo das escadas.
Worn with pain, and weak from the prolonged hardships which I had undergone, I was removed, with a great train of wounded sufferers, to the base	Desgastado com dor, e fraco das dificuldades prolongadas que eu tinha sofrido, I foi removido, com uma grande comitiva dos doentes feridos, para o hospital de base em

hospital at Peshawar.	Peshawar.
The sight of a friendly face in the great wilderness of London is a pleasant thing indeed to a lonely man.	A visão de um rosto amigável na grande deserto de Londres é uma coisa agradável, na verdade, um homem solitário.
The landlady had become so accustomed to my late habits that my place had not been laid nor my coffee prepared.	A dona da casa tornou-se tão acostumados a minha tarde hábitos que meu lugar não haviam sido previstas, nem o meu café preparado.
The garden was bounded by a three-foot brick wall with a fringe of wood rails upon the top, and against this wall was leaning a stalwart police constable, surrounded by a small knot of loafers, who craned their necks and strained their eyes in the vain hope of catching some glimpse of the proceedings within.	O jardim foi delimitada por uma parede de tijolos de três pés com uma franja de trilhos de madeira sobre o alto, e contra esta parede estava encostado um policial robusto polícia, rodeado por um pequeno grupo de mocassins, que esticaram o pescoço e se esforçou os olhos na vã esperança de pegar algum vislumbre do processo dentro.
"Why, man, it is the most practical medico-legal discovery for years.	"Ora, o homem, é a descoberta mais prática médico-legal durante anos.
Sherlock Holmes chuckled to himself, and appeared to be about to make some remark, when Lestrade, who had been in the front room while we were holding this conversation in the hall, reappeared upon the scene, rubbing his hands in a pompous and self-satisfied manner.	Sherlock Holmes riu para si mesmo, e parecia estar prestes a fazer alguma observação, quando Lestrade, que tinha sido na sala da frente enquanto estávamos segurando esta conversa no corredor, reapareceu em cena, esfregando as mãos em um pomposo e auto forma satisfeito.
This malignant and terrible contortion, combined with the low forehead, blunt nose, and prognathous jaw gave the dead man a singularly simious and ape-like appearance, which was increased by his writhing, unnatural posture.	Esta contorção maligna e terrível, combinado com a baixa testa, nariz sem corte, e da mandíbula prognata deu o morto uma aparência singularmente simious e de macacos, que foi aumentada por sua contorção, postura anormal.

As I watched him I was irresistibly reminded of a pure-blooded well-trained foxhound as it dashes backwards and forwards through the covert, whining in its eagerness, until it comes across the lost scent.	Ao observá-lo eu estava irresistivelmente lembrou de um foxhound bem treinado puro-sangue, uma vez que corre para trás e para a frente através da encoberta, lamentando-se em sua ânsia, até que ele se depara com o perfume perdido.
On another occasion an old white-haired gentleman had an interview with my companion; and on another a railway porter in his velveteen uniform.	Em outra ocasião, um velho cavalheiro de cabelos brancos fez uma entrevista com o meu companheiro; e em outro um porteiro ferroviária em seu uniforme veludo.
The same afternoon brought a grey-headed, seedy visitor, looking like a Jew pedlar, who appeared to me to be much excited, and who was closely followed by a slip-shod elderly woman.	Na mesma tarde trouxe um visitante de cabelos grisalhos, decadente, parecendo um mascate judeu, que me pareceu ser muito animado, e que foi seguido de perto por uma mulher idosa desleixado.
"I should like to meet him," I said. "If I am to lodge with anyone, I should prefer a man of studious and quiet habits.	"Eu gostaria de conhecê-lo," eu disse. "Se estou a apresentar com ninguém, eu preferiria um homem de hábitos de estudo e tranquilos.
On one corner of this was stuck the stump of a red wax candle.	m um canto desse estava preso o toco de uma vela de cera vermelha
Opposite the door was a showy fireplace, surmounted by a mantelpiece of imitation white marble.	Em frente à porta havia uma lareira vistosa, encimada por uma lareira de imitação de mármore branco.
They consisted of a couple of comfortable bed-rooms and a single large airy sitting-room, cheerfully furnished, and illuminated by two broad windows.	Eles consistiu de um par de confortáveis quartos de cama e uma sala de estar-single amplo e arejado, alegremente decorados, e iluminado por duas grandes janelas.
Like all other arts, the Science of Deduction and Analysis is one which can only be acquired by long and patient study nor is life long enough to allow any mortal to attain the highest possible perfection in it.	Como todas as outras artes, a ciência da dedução e Análise é aquela que só pode ser adquirida por longa e paciente estudo nem é vida longa o suficiente para permitir que qualquer mortal para alcançar a perfeição

	máxima possível na mesma.
Its somewhat ambitious title was "The Book of Life," and it attempted to show how much an observant man might learn by an accurate and systematic examination of all that came in his way.	É um pouco ambicioso título era "O Livro da Vida", e tentou mostrar o quanto um homem observador pode aprender através de um exame preciso e sistemático de tudo o que veio em seu caminho.
At present my attention was centred upon the single grim motionless figure which lay stretched upon the boards, with vacant sightless eyes staring up at the discoloured ceiling.	Neste momento a minha atenção foi centrada sobre a figura imóvel sombria única que estava estendido sobre as placas, com olhos cegos vagos olhando para o teto descolorido.
There was one little sallow rat-faced, dark-eyed fellow who was introduced to me as Mr. Lestrade, and who came three or four times in a single week.	Houve uma pequena andorinha, companheiro de olhos escuros com cara de rato que me foi apresentado como o Sr. Lestrade, e que veio três ou quatro vezes em uma única semana.
Yet his zeal for certain studies was remarkable, and within eccentric limits his knowledge was so extraordinarily ample and minute that his observations have fairly astounded me.	No entanto, o seu zelo para certos estudos foi notável, e dentro dos limites excêntricas seu conhecimento era tão extraordinariamente ampla e minuto que suas observações foram bastante me surpreendeu.