

Análise das Colocações Especializadas na Área de Sensoriamento Remoto sob a Perspectiva da Fraseologia

ANALYSIS OF SPECIALISED COLLOCATIONS IN THE AREA OF REMOTE SENSING IN THE PERSPECTIVE OF PHRASEOLOGY

Adriane **ORENHA-OTTAIANO** *
Dalila dos Santos **HASMANN** **
Diva Cardoso de **CAMARGO** ***

Resumo: O objetivo desta pesquisa é construir e analisar um *corpus* paralelo na área de sensoriamento remoto para, por meio de frequência, identificar as colocações especializadas em inglês e proceder à busca de seus equivalentes em língua portuguesa. A investigação baseia-se na abordagem interdisciplinar dos Estudos da Tradução Baseados em *Corpus* (BAKER, 1995; CAMARGO, 2007), da Linguística de *Corpus* (BERBER SARDINHA, 2004; TOGNINI-BONELLI, 2001), da Fraseologia (ORENHA-OTTAIANO, 2009; PAVEL, 1993) e, em partes, da Terminologia (BARROS, 2004). Será utilizado o programa WordSmith Tools (SCOTT, 2012), versão 6.0. Para auxiliar este estudo, dois *corpora* comparáveis em inglês e português também

* Mestre em Estudos Linguísticos e Literários em Inglês, pela Universidade de São Paulo (2004), e Doutora em Estudos Linguísticos, pela UNESP (2009). Atualmente é Professora Assistente Doutora da UNESP, Câmpus de São José do Rio Preto. Contato: adriane@ibilce.unesp.br.

** Mestranda no Programa de Pós-Graduação em Estudos Linguísticos pela Universidade Estadual Paulista – UNESP, Câmpus de São José do Rio Preto, na linha de pesquisa de Estudos da Tradução. Contato: dalila_hasmann@yahoo.com.br.

*** Doutora em Tradução pela Universidade de São Paulo (1993). Pós-doutorado em Estudos da Tradução por The University of Manchester (2003). Livre-Docência em Estudos da Tradução pela Universidade Estadual Paulista “Júlio de Mesquita Filho” (2005). Atualmente é Professor Adjunto-MS5, aposentada da UNESP, Câmpus de São José do Rio Preto. Contato: divaccamargo@gmail.com.

foram construídos a partir de artigos publicados em periódicos nacionais e internacionais de sensoriamento remoto. Os resultados mostram que as colocações em português nesta área ainda parecem estar em processo de convencionalização, pois os tradutores fizeram uso de maior variação vocabular em suas traduções, o que pode ser uma maneira de tornar o texto mais claro para o leitor.

Palavras-chave: Colocações especializadas. Estudos da tradução baseados em *corpus*. Sensoriamento remoto.

Abstract: The aim of this research is to build and analyze a parallel *corpus* in the field of remote sensing in order to identify, according to its frequency, specialized collocations in English and then search for their equivalents in Portuguese. The research is based on the interdisciplinary approach of *Corpus-Based Translation Studies* (BAKER, 1995; CAMARGO, 2007), *Corpus Linguistics* (BERBER SARDINHA, 2004; TOGNINI-BONELLI, 2001), *Phraseology* (ORENHA-OTTAIANO, 2009; PAVEL, 1993), and some principles of *Terminology* (BARROS, 2004). For manipulating the *corpora*, the program *WordSmith Tools* (SCOTT, 2012) version 6.0 is used. To support this study, two comparable *corpora* in English and Portuguese were also built from articles published in both national and international journals in remote sensing. The results show that the collocations in Portuguese seem to be still in the process of conventionalization, as the translators made use of greater variation in their translational options, which can be a way to make the text clearer for the reader.

Key-words: Specialized collocation. *Corpus*-based translation studies. Remote sensing.

Introdução

Sensoriamento remoto é a ciência pela qual se obtêm informações sobre objetos ou fenômenos a partir de dados coletados sem o contato físico com tais objetos. Um instrumento de sensoriamento remoto coleta informação sobre um objeto ou fenômeno sem estar em contato físico direto com ele. Este tipo de instrumento pode estar localizado a poucos metros acima do solo, a bordo de uma aeronave ou mesmo satélite (JENSEN, 2009, p. 3).

Os dados resultantes desta técnica são úteis para o estudo de alguns dos temas que se tornaram comuns nestes últimos anos, como: aquecimento global, desastres ambientais, desmatamento da Amazônia, efeito estufa, geração de energia sustentável, mudanças climáticas, previsão do tempo, entre muitos outros, que são áreas que se utilizam e se beneficiam diretamente da tecnologia de sensoriamento remoto. Quando se trata de estudar áreas extensas e de condições extremas como a Amazônia, por exemplo, o avanço deste campo de pesquisa contribui concretamente para a obtenção de conhecimento dos ecossistemas dessa região.

Tais estudos e obtenção de dados sobre nosso território são de interesse não só do Brasil, mas do mundo. Segundo os próprios especialistas da área (NOVO et al., 2005, p. 270), “há uma urgente necessidade de uniformização de linguagem e protocolos no âmbito da comunidade de sensoriamento remoto, para que a comunicação, difusão e troca de informações e idéias se tornem mais profícuas”.

Compartilhando desse mesmo ponto de vista, já no ano de 1989, pesquisadores da *Sociedad de Especialistas Latinoamericanos en Percepción Remota* (Associação de Especialistas Latinoamericanos em Sensoriamento Remoto) – SELPER, se juntaram e produziram um dicionário trilingue de terminologia em inglês, espanhol e português. Esta empreitada teve a participação de diferentes países da América Latina e Caribe, envolvendo o Instituto Nacional de Pesquisas Espaciais (INPE) e a Agência Canadense de Desenvolvimento Internacional (*Canadian International Development Agency* – CIDA). A publicação deste dicionário foi uma contribuição da CIDA na transferência de tecnologia de sensoriamento remoto para países de línguas espanhola e portuguesa. Essa iniciativa é um dos componentes do Teledeteção II, um projeto de transferência de tecnologia em sensoriamento remoto do Canadá ao Peru.

Durante vários anos de experiência da CIDA na transferência de tecnologia para países em desenvolvimento, tornou-se evidente que uma das barreiras mais difíceis era a linguagem e, mais particularmente, a terminologia. O sensoriamento remoto iniciou e se desenvolveu principalmente, como uma tecnologia de idioma inglês. Mais recentemente, o francês tornou-se a segunda língua neste domínio, mas, como acontece com outras tecnologias, o inglês é o idioma principal de publicação. Embora o espanhol seja a quarta língua mais comum no mundo – depois de mandarim, inglês e russo –, apenas cerca de um por cento das publicações atuais relacionadas ao sensoriamento remoto é em espanhol.

A terminologia de sensoriamento remoto se divide em dois grandes grupos. O primeiro grupo compreende os termos que já existiam no inglês antes do desenvolvimento de sensores remotos (advindos de disciplinas-mãe), mas para alguns dos quais o uso corrente tem dado um significado não convencional e, muitas vezes, confuso (SELPER, 1989).

O segundo grupo consiste de novas palavras criadas pela tecnologia. Em muitos casos, não existem definições de qualquer tipo para essas palavras. Os usuários que não estão familiarizados com o idioma inglês, e que não têm acesso à extensa literatura técnica da área, muitas vezes têm dificuldade em inferir uma definição de trabalho. Neste grupo de termos, podemos incluir as siglas, que são amplamente utilizadas em sensoriamento remoto, mas são, em grande parte, ininteligíveis para o recém-chegado. Para os autores do glossário, “[o] lamentável resultado dos problemas de linguagem e terminologia é que somente os profissionais com sólido conhecimento da língua inglesa podem participar ativa e efetivamente em sensoriamento remoto”¹ (SELPER, 1989). Considerando-se esses obstáculos, a CIDA concordou em incluir um dicionário de termos de sensoriamento remoto no Projeto Teledeteção II.

Passados dezoito anos desta publicação, houve, no ano de 2007, uma segunda iniciativa dos pesquisadores que compõem a SELPER para atualizar o dicionário e disponibilizá-lo on-line. Sem dúvida, foi um enorme avanço poder tornar este trabalho acessível por meio da internet. Além da atualização dos termos, também foi adicionado um quarto idioma, o francês. Na interface virtual, o usuário, ao iniciar sua consulta, escolhe em que língua fará suas buscas. Sua divulgação é promovida e assistida pela SELPER Internacional e pelos membros que participaram de sua revisão e atualização. Para os autores:

A ‘área dos sensores remotos’ é um campo de estudo multidisciplinar a qual convergem ramos das ciências e da tecnologia. Como tal, requer, para sua atividade, o uso de termos científicos que ocorrem

¹ The unfortunate result of the problems of language and terminology is that only professionals with a solid knowledge of English can participate actively and effectively in remote sensing (SELPER, 1989) [Todas as traduções serão de nossa responsabilidade].

transversalmente. As ciências e a tecnologia, por sua vez, geram uma nova terminologia produto de sua relação com o campo dos sensores remotos. (SELPER, [s.d.]).

E assim justifica-se a escolha do sensoriamento remoto como foco principal deste estudo. Por meio de sua multidisciplinaridade, promover o conhecimento de sensoriamento remoto é disseminar informação e contribuir para diversas outras áreas do saber. É necessário haver mais uma atualização deste grande trabalho iniciado há mais de duas décadas, dessa vez não só sob a ótica da Terminologia, em busca de termos, mas também de expressões (colocações) que carregam sentidos próprios. E esta atualização é, acima de tudo, necessária, pois, como em qualquer outra tecnologia, a linguagem em sensoriamento remoto avança dia a dia com o desenvolvimento de novas técnicas, com o lançamento de novos foguetes, satélites, plataformas orbitais, estações espaciais, softwares, etc..

Desta forma, os objetivos deste estudo são: 1) proceder à compilação de um *corpus* paralelo da área de sensoriamento remoto formado por textos originalmente escritos em inglês e sua tradução para português; 2) proceder à compilação de dois *corpora* comparáveis, inglês e português, compostos por artigos de sensoriamento remoto; 3) levantar e analisar as colocações especializadas, que serão identificadas por meio de sua frequência; e 4) proceder à busca de seus equivalentes ou correspondentes em língua portuguesa. Por meio destes objetivos, tem-se a intenção de apresentar algumas colocações que podem vir a constituir desafios para os tradutores.

1 Fundamentação Teórica

Com o avanço da pesquisa em diversas áreas do conhecimento, houve um aumento no número de estudos relacionados aos discursos especializados. A tradução tem sido uma forte aliada e tem contribuído significativamente para a ampla divulgação do que está sendo produzindo atualmente. Como recurso auxiliar fundamental para o exercício do tradutor e dos próprios pesquisadores estão os glossários bi ou multilíngues e os dicionários terminológicos das mais diversas línguas de especialidade. No entanto, muitos deles apenas enfocam termos e, ao fazê-lo, deixam de fora as combinatórias de palavras, tais como, por exemplo, as *colocações* em contextos especializados

(ORENHA-OTTAIANO, 2009, p. 32). Um trabalho voltado para essas colocações mostra-se importante porque a busca por equivalentes e correspondentes torna-se, em geral, mais difícil para o tradutor do que a busca por equivalentes de termos simples. Por termo, ou unidade terminológica, entende-se:

... uma palavra (simples, derivada, composta) ou um grupo compacto de palavras (sintagma lexical, nominal, verbal ou adjetival) que designa um conceito de tipo objeto, ação ou propriedade, pertencendo ao sistema conceitual de uma especialidade. (PAVEL, 2003, p. 105).

Por sua vez, para língua de especialidade a autora traz a seguinte definição:

A língua de especialidade é um subconjunto da língua geral que serve para transmitir um saber atinente a um campo de experiência particular. Ela tem em comum com a língua geral a gramática e uma parte de seu inventário léxico-semântico (morfemas, palavras, sintagmas e regras combinatórias), mas faz deles um uso seletivo e criativo que reflete as particularidades dos conceitos em jogo e que apresenta variações sociais, geográficas e históricas. (PAVEL, 2003, p. 100).

Ao adentrarmos o ambiente da fraseologia, rapidamente percebemos haver uma série de perspectivas para seu tratamento, uma rica diversidade denominativa e conceitual, bem como uma grande dificuldade de delimitar e classificar seu objeto de estudo (ORENHA-OTTAIANO, 2009, p. 50). Por esta razão, com o intuito de uniformizar a nomenclatura utilizada neste trabalho, optamos por nos basear na pesquisa de Orenha-Ottaiano (2009). A autora refere-se a essas combinatórias sintagmáticas de unidades terminológicas, presentes nas línguas de especialidade, por meio do termo *colocações especializadas*, assim diferenciando-as das *colocações* da língua geral. Desta forma, daqui por diante, denominaremos nosso objeto de estudo como *colocações especializadas*.

De acordo com Anderson (2009, p. 53, grifos do autor), “[o] termo *colocação* refere-se à tendência das palavras de ocorrer no ambiente próximo

de outras determinadas palavras”.² A esse respeito, Firth (1957 apud ORENHA-OTTAIANO, 2009, p. 33), em seu artigo *Modes of Meaning*, também criou a frase: “you shall know a word by the company it keeps”, evidenciando que uma palavra é conhecida por aquelas que a acompanham.

Orenha-Ottaiano (2009, p. 33-34) afirma que as ligações entre os elementos que formam essas colocações não são semânticas, mas sintagmáticas e convencionais: “são regidas pela convenção que lhes foi imposta para que, desse modo, seus elementos possam se combinar. Há, dessa forma, uma restrição combinatória entre seus elementos”. Para a autora, assim como as colocações, que pertencem à língua geral, as colocações especializadas, pertencentes às línguas de especialidade, também podem ser definidas basicamente da mesma forma. Ambas são definidas pelo uso convencional; no entanto, a convenção, no caso das colocações especializadas, é estabelecida dentro de uma comunidade linguística específica, dentro de um grupo de especialistas (ORENHA-OTTAIANO, 2009, p. 44).

Segundo Pavel (2003), por fraseologia dentro de uma língua de especialidade entende-se a “combinatória sintagmática das unidades terminológicas decorrentes de uma estrutura conceitual coerente. As UT [unidades terminológicas] são tomadas como núcleo de coocorrências usuais ou privilegiadas nos textos de uma especialidade” (PAVEL, 2003, p. 106).

Pelo fato de as combinações lexicais especializadas estarem inseridas em uma comunidade linguística específica, como um subconjunto da língua geral, e pertencerem ao “seletivo e criativo inventário léxico-semântico desse subconjunto” (ORENHA-OTTAIANO, 2009, p. 44), os tradutores precisam adquirir conhecimento a respeito do comportamento de tais unidades terminológicas para, assim, poderem utilizá-las em um contexto adequado. A este respeito, Cowie (1998) justifica que a falta de domínio de expressões idiomáticas e colocações pode ter um efeito negativo sobre a eficácia da escrita [ou tradução], capaz de desviar a atenção do leitor do conteúdo da mensagem para a forma linguística, e, mais geralmente, esta deficiência pode acabar não atingindo as expectativas estilísticas de uma comunidade acadêmica³ (COWIE, 1998, p. 14). O autor defende que “a proficiência de

² The term collocation refers to the tendency of words to occur in the close environment of particular other words (ANDERSON, 2009, p. 53).

³ Imperfect control of idioms and collocations can have an appreciable effect on the effectiveness of student writing, deflecting the attention of the reader from message content to linguistic form, and more generally failing to meet the stylistic expectations of the academic community (COWIE, 1998, p. 14).

nativos em uma língua depende consideravelmente de um estoque de unidades pré-fabricadas – ou ‘prefabs’ – que variam em complexidade e estabilidade interna”⁴ (COWIE, 1998, p. 1). Desse modo, acreditamos que o tradutor precisa estar atento às chamadas unidades pré-fabricadas que, no caso desta investigação, chamamos de colocações especializadas. Ademais, quanto maior domínio tiver das referidas combinações de palavras, em uma dada língua de especialidade, mais precisas podem se tornar suas opções tradutórias.

No que se refere à inserção dessas combinatórias em dicionários, Carneado Moré (1983, p. 39) observa:

Há muito tempo lexicógrafos observaram que o significado de uma palavra isolada nunca se perde. Isto frequentemente acontece quando uma palavra, ao juntar-se com outra, forma uma combinação indivisível. Daí a necessidade de os dicionários incluírem combinações deste tipo.⁵

A fraseologia representa um grande desafio para a prática lexicográfica. A própria natureza de dicionários, com a sua organização linear, obriga as palavras a serem consideradas isoladamente. No entanto, a evidência de textos, incluindo *corpora*, mostra claramente que as palavras não são isoladas; pelo contrário, conectam-se umas com as outras através de uma padronização fraseológica e têm uma relação de significação a partir do contexto de uso. Além disso, os próprios modelos de produção de linguagem apontam para sua natureza fragmentada e fraseológica (MOON, 2008, p. 313).

Para Croitoru e Dumitraşcu (2006), o domínio no uso de colocações parece ser o ponto mais fraco de um falante não nativo. Este aspecto possibilita que um falante nativo da língua-alvo perceba ou identifique que o

⁴ The notion that native-like proficiency in a language depends crucially on a stock of prefabricated units - or ‘prefabs’ - varying in complexity and internal stability (COWIE, 1998, p. 01).

⁵ Hace tiempo los lexicógrafos notaron que la significación de una palabra aislada nunca se pierde. Esto sí sucede, frecuentemente, cuando una palabra, al unirse con otra, forma una combinación indivisible. De ahí la necesidad de incluir en los diccionarios combinaciones de este tipo (CARNEADO MOREÉ, 1983, p. 39).

tradutor não é um falante nativo da língua para a qual está traduzindo. Além disso, as colocações estranhas e não usuais que o tradutor acaba “criando” sob influência do texto-fonte serão interpretadas como colocações incomuns em sua própria língua materna (CROITORU; DUMITRAȘCU, 2006, p. 298). Quanto a uma estratégia de tradução que se baseia mais na forma (em vez do conteúdo) da mensagem, deve-se considerar que pode ser uma abordagem enganosa e até resultar em uma “falha pragmática”, ou seja, pode demonstrar uma incapacidade do tradutor de compreender o significado do que é dito pela maneira que é dito. Além disso, a transcodificação palavra por palavra pode nem sempre ser apropriada e, às vezes, até mesmo tornar-se totalmente inapropriada (CROITORU; DUMITRAȘCU, 2006, p. 298). Em razão de considerarmos o automatismo na produção fraseológica, também compartilhamos da visão de que a tradução de colocações especializadas costuma ser uma tarefa mais árdua para o tradutor, pois podem conter vários aspectos e transmitir significados que não podem ser inferidos a partir da soma dos elementos que os constitui (ORENHA-OTTAIANO, 2009, p. 52).

Para atingir os objetivos deste trabalho, elegemos a abordagem da Linguística de *Corpus*. Na visão de Baker (1995, p. 226), entende-se por *corpus* “um conjunto de textos naturais” em oposição a exemplos/sentenças criados com o propósito específico de mostrar um fenômeno linguístico, “organizados em formato eletrônico, passíveis de serem analisados, preferencialmente, de forma automática ou semi-automática (em vez de manualmente)”.⁶ Para a autora, um *corpus* paralelo é composto de textos originais em uma determinada língua de origem (neste caso, inglês) e suas respectivas traduções para outra língua (língua da tradução, sendo aqui o português). Esse tipo de *corpus* permite pesquisar traduções consagradas de certos itens lexicais ou estruturas sintáticas, peculiaridades de determinado(s) tradutor(es), diferenças entre traduções de um mesmo texto, produzidas em períodos diversos, normas tradutórias, entre outros.

O segundo *corpus*, o comparável, “como sugestão da própria palavra pode ser comparado de acordo com critérios externos e dar ideias sobre

⁶ *Corpus* mean[s] any collection of running texts (as opposed to examples/sentences), held in electronic form and analysable automatically or semi-automatically (rather than manually) (BAKER, 1995, p. 226).

dois sistemas linguísticos independentes”⁷ (TOGNINI-BONELLI, 2001, p. 133). Para a autora, o ideal é que o *corpus* comparável seja composto de: 1) um *corpus* comparável de textos originais na língua 1 (neste caso, o inglês); 2) um *corpus* de estudo com os textos traduzidos em questão (o qual pode ser ou não um *corpus* paralelo); e 3) um *corpus* comparável de textos originais na língua 2 (neste caso, o português) (TOGNINI-BONELLI, 2001, p. 134-135). Para Camargo (2007, p. 49), “esse tipo de *corpus* permite uma investigação mais completa de textos técnicos ou especializados”; possibilita identificar padrões que são ou restritos ao texto traduzido ou que ocorrem em frequências significativamente mais altas ou mais baixas no texto traduzido do que em relação aos textos originais. Outra contribuição desse tipo de estudo é para com a elaboração de glossários de termos especializados acompanhados de seu cotexto, os quais aparecem com maior frequência na linguagem de especialidade representada nos textos traduzidos e textos originais selecionados para análise (CAMARGO, 2007, p. 49).

2 Metodologia

Neste trabalho, usamos *corpora* paralelo e comparáveis, além de dois *corpora* de referência da língua geral. O *corpus* paralelo consiste de dois sub*corpora*: 1) sub*corpus* original, composto por um livro originalmente escrito em inglês; e 2) sub*corpus* traduzido, composto pela respectiva tradução para português. Ambos os livros foram digitalizados por meio de escaneamento, utilizando o Reconhecimento Ótico de Caracteres.

A obra original digitalizada foi: *Remote sensing of the environment: an Earth resource perspective*, de John R. Jensen, segunda edição, lançado pela Editora Pearson Prentice Hall, em 2007, contendo 592 páginas, e a versão traduzida *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres*, lançada pela Editora Parêntese, em 2009, com 672 páginas. Sua tradução foi feita por pesquisadores do Instituto Nacional de Pesquisas Espaciais (INPE): Drs. José Carlos Neves Epiphânio (Coordenador), Antonio Roberto Formaggio, Athos Ribeiro Santos, Bernardo Friedrich Theodor Rudorff, Cláudia Maria de Almeida e Lênio Soares Galvão, que optaram por colocar esse livro à

⁷ Comparable *corpora*, as the word suggests, can be compared according to external criteria and give insights into two independent linguistic systems (TOGNINI-BONELLI, 2001, p. 133).

disposição do estudante e do profissional brasileiro e latino-americano, um texto de excelência que servirá de base para sua formação continuada. O fato de a versão em português ter sido realizada por renomados especialistas do principal instituto de pesquisas no tema da América do Sul, o INPE, confere a esta edição do livro uma confiabilidade técnico-científica.

Já os *corpora* comparáveis, ambos são constituídos por artigos científicos de sensoriamento remoto, publicados em revistas nacionais e internacionais renomadas na área. Para a coleta do *corpus* comparável de língua inglesa encontramos a base de dados *Remote Sensing Magazine*, que oferecia importante contribuição para as necessidades da pesquisa. Ao final da recolha dos textos, obtivemos um total de 114 artigos, que contêm 862.192 palavras, representando mais que o triplo do tamanho do *corpus* de estudo. O *corpus* comparável de língua portuguesa foi coletado por meio da base de periódicos *SciELO*. Ao final, atingimos um total de 201 artigos coletados contendo 967.545 palavras corridas. Esta comparação, entre o *corpus* de estudo e os comparáveis, tem o propósito de examinar o uso das colocações especializadas mais frequentes encontradas no *corpus* paralelo e identificar seus padrões formais de uso.

No tocante ao *corpus* de referência, é formatado como uma lista de frequência de palavras e funciona como termo de comparação para a análise. “A sua função é fornecer uma norma com a qual se fará a comparação das frequências do *corpus* de estudo” (BERBER SARDINHA, 2004). É um *corpus* da língua geral normalmente composto por milhões de palavras e é utilizado para contrastar a frequência das palavras encontradas no *corpus* de estudo. Desse modo, se uma palavra tiver alta frequência no *corpus* de estudo, porém baixa frequência no *corpus* de referência, ou seja, alta frequência na área de sensoriamento remoto e baixa frequência na língua geral, esta palavra poderá ser considerada *chave* e passa a ser uma candidata a análise. O *corpus* de referência de língua inglesa que usamos nesta pesquisa é o *BNC (British National Corpus)*, que conta com 100 milhões de palavras do inglês britânico escrito e falado e, para o *corpus* de referência de língua portuguesa, foi utilizado o Lácio-Ref, que possui em torno de 8 milhões de palavras do português brasileiro contemporâneo escrito.

O manuseio dos *corpora* com rapidez e precisão é possível graças ao software WordSmith Tools (SCOTT, 2012), versão 6.0, que possui três ferramentas principais: WordList, KeyWord e Concord, que fornecem, respectivamente: a) produção de lista de palavras contendo todas as palavras

do arquivo selecionado; b) extração de palavras-chave; e c) extração de linhas de concordância. Este programa foi criado em 1996 por Mike Scott, professor da Universidade de Liverpool, Reino Unido, e tem contribuído para a divulgação da Linguística de *Corpus* no Brasil. Sua versão demo é obtida pela internet⁸ e o usuário, ao pagar a licença, recebe um código que a transforma em licença completa (BERBER SARDINHA, 2004).

3 Análise dos Dados

Tanto os Estudos da Tradução Baseados em *Corpus* quanto a Linguística de *Corpus* estão condicionados à tecnologia, a fim de possibilitar o armazenamento de *corpora* e também a sua exploração. Quanto aos primeiros resultados obtidos da digitalização do *corpus*, utilizamos a ferramenta WordList para criar as listas de palavras dos *subcorpora* em inglês e português e apresentamos, no Quadro 1, alguns resultados estatísticos:

Quadro 1 – Estatística simples a partir do *corpus* paralelo (inglês e português)

Subcorpus Paralelo inglês (original)		Subcorpus Paralelo português (traduzido)	
Itens	226.285	Itens	262.877
Formas	12.797	Formas	15.987
Razão Forma/Item	5.66	Razão Forma/Item	6.08
Razão Forma/Item Padronizada	38.48	Razão Forma/Item Padronizada	39.20

De acordo com resultados iniciais obtidos, podemos observar que o número de itens do texto traduzido (262.877 itens) foi maior (36.592 itens a mais) do que do texto original (226.285 itens), indicando que os tradutores empregaram mais palavras em suas traduções do que a quantidade que

⁸ O Software WordSmith Tools na versão 6.0 demo pode ser baixado gratuitamente nos seguintes endereços: <www.liv.ac.uk/~ms2928/>, <www.lexically.net/> e <www.oup.com/elt/global/isbn/6890/>.

constava no texto original. Este dado pode sugerir que lançaram mão de maior variação vocabular em relação ao texto traduzido.

O número de formas (vocábulos) do texto original (12.797) foi menor que o número de formas do texto traduzido (15.987), o que mostra que, ao traduzir, os tradutores podem ter variado no emprego de alguns termos, para garantir que o conceito transmitido fosse o mais próximo do texto original. Os tradutores, ao introduzirem novos conceitos, geralmente atuam para que as palavras ou expressões empregadas sejam aceitas pela comunidade científica e se universalizem dentro desse público, passando a constituir termos. Por esta razão, as listas de palavras serão monitoradas pelo *corpus* comparável, a fim de garantir que não estejamos baseando todo o trabalho na decisão tomada pelos tradutores da obra, mas estamos considerando, também, outras possibilidades de equivalências ou correspondências existentes.

Já a razão forma/item está relacionada à diversidade do uso de vocábulos observada em um dado *corpus*. Neste estudo, utilizamos a razão forma/item padronizada (*standardized*), que é mais adequada para neutralizar a influência do tamanho do texto no cálculo da razão, pois a primeira é sensível à extensão do material textual, não sendo confiável para uso em comparação entre textos de tamanhos diferentes (BERBER SARDINHA, 2004). Nos *corpora* de estudo de sensoriamento remoto, pudemos verificar que essa razão forma/item padronizada foi de 38.48 no texto original e de 39.20 no texto traduzido. Esses dados reafirmam que houve uma maior variação de itens na tradução.

Estas listas de palavras dos sub*corpora* de sensoriamento remoto, geradas pelo WordList, devem ser confrontadas com os *corpora* de referência (de língua geral em inglês e em português). Este procedimento compara as frequências das palavras do *corpus* de estudo com o de referência e considera apenas aquelas que possuem alta frequência no *corpus* de estudo (palavras-chave) e baixa frequência na língua geral. No Quadro 2, trazemos uma amostra das primeiras palavras-chave encontradas nos textos originais e nos textos traduzidos, a fim de procedermos a uma análise:

Quadro 2 – Listagem de palavras-chave extraídas de ambos os subcorpora

TO	Freq.	Keyness	TT	Freq.	Keyness
Infrared	860	9.267,94	Remoto	1.104	6.441,89
Sensing	871	8.814,77	Sensoriamento	1.009	6.053,69
Remote	1.106	8.683,53	Infravermelho	710	4.368,67
Aerial	727	6.821,61	Fotografia	678	3.558,13
Sensor	553	5.774,61	Dados	1.275	2.949,82
Data	1.223	5.717,70	Terreno	600	2.594,19
Color	497	5.403,62	Termal	392	2.544,72
Image	851	4.827,31	Aérea	442	2.506,91
Spectral	443	4.548,49	Reflectância	417	2.462,94
Reflectance	422	4.256,23	Imagem	828	2.436,73
Thermal	476	4.175,91	Sensor	427	2.383,12
Radar	447	3.889,91	Radar	429	2.296,73

Notamos que, nesta listagem de doze itens, podemos encontrar um total de dez equivalências, como: *infrared* > infravermelho; *sensing* > sensoriamento; *remote* > remoto; *aerial* > aérea; *sensor* > sensor; *data* > dados; *image* > imagem; *reflectance* > reflectância; *thermal* > termal; *radar* > radar. A partir da indicação que temos da alta frequência destas palavras, podemos conduzir um estudo de suas correlações em contexto. Num primeiro momento, ao analisarmos a coluna de frequência, percebemos que esta não segue uma ordem linear, pois, nesta lista, o fator que está sendo considerado para ranquear as palavras é o grau de “chavicidade” (*Keyness*). Este cálculo é feito automaticamente pelo WordSmith Tools.

Desta maneira, por uma questão de delimitação, nos ateremos a analisar apenas o primeiro termo da lista (*infrared* > infravermelho). Este termo tem 860 ocorrências no sub*corpus* paralelo em inglês. Utilizando a ferramenta Concord, do WordSmith Tools, inserimos *infrared* como palavra de busca (*Search Word*). Na aba *collocates* (colocados) pode-se visualizar mais claramente quais são as palavras que coocorrem com *infrared*. Conforme observa Berber Sardinha (2004), essas concordâncias possibilitam ao pesquisador ver a palavra que está sendo buscada, também chamada de nódulo (*node*), acompanhada de seu cotexto – palavras que ocorrem junto ao nódulo no *corpus*. Vejamos, no Quadro 3, algumas linhas de concordância a partir da palavra de busca *infrared* extraídas do sub*corpus* paralelo de textos originais em língua inglesa.

Quadro 3 – Linhas de concordâncias da palavra *infrared* extraída do sub*corpus* original

<i>thermal infrared</i>	352	<i>infrared reflectance</i>	24
<i>near infrared</i>	293	<i>visible infrared</i>	24
<i>middle infrared</i>	74	<i>infrared radiant flux</i>	21
<i>thermal infrared remote sensing</i>	70	<i>visible near infrared</i>	19
<i>color infrared</i>	69	<i>black and white</i>	18
<i>infrared energy</i>	57	<i>incident near infrared</i>	18
<i>infrared image</i>	55	<i>infrared portions</i>	17
<i>infrared data</i>	49	<i>infrared regions</i>	17
<i>red and near infrared</i>	46	<i>infrared radiation</i>	16
<i>thermal infrared data</i>	39	<i>near infrared bands</i>	15
<i>infrared bands</i>	35	<i>predawn thermal infrared</i>	15
<i>infrared imagery</i>	35	<i>near infrared radiant flux</i>	14
<i>infrared region</i>	34	<i>infrared aerial photography</i>	13
<i>thermal infrared imagery</i>	32	<i>thermal infrared image of</i>	13
<i>infrared band</i>	30	<i>color infrared aerial photography</i>	12
<i>infrared film</i>	30	<i>thermal infrared band</i>	12
<i>near infrared energy</i>	30	<i>infrared photograph</i>	11
<i>near infrared image</i>	30	<i>visible infrared imaging spectrometer</i>	11
<i>thermal infrared image</i>	30	<i>blue green red and near infrared</i>	10
<i>visible and near infrared</i>	27	<i>thermal infrared data collection</i>	10
<i>color infrared film</i>	25	<i>multispectral scanner</i>	8
<i>green red and near infrared</i>	25	<i>thermal infrared energy</i>	8
<i>infrared radiant</i>	25	<i>thermal infrared sensor</i>	8

O próximo passo foi realizar a observação das linhas de concordância, dos *clusters* (agrupamentos lexicais) e dos *collocates* (colocados). Apresentamos, no Quadro 4, a seguir, alguns dos *clusters* encontrados no *corpus*:

Quadro 4 – Clusters encontrados a partir da palavra de busca *infrared*

where the roof subsurface water is located, Thermal **infrared** remote sensing is the preferred method
 Plate 8-2 a) Density sliced display of predawn thermal **infrared** data of the Four Mile Creek thermal
 Figure depicts an aerial photograph and predawn thermal **infrared** image of the Solomon Blatt
 from view. Aerial photography and predawn thermal **infrared** imagery of downtown New York City
 away from the principal point. b) Predawn 1 x 1 m thermal **infrared** imagery. Note the intricate
 INDEX Clouds, remote sensing of height in thermal **infrared** imagery in visible imagery type
 0.15 and 30 cm (between GHz), well beyond the thermal **infrared** region. The microwave

A partir destes fragmentos identificados no *corpus* como sendo de alta frequência, nos guiamos para uma análise mais minuciosa das diferentes possibilidades de união de termos para formação de colocações especializadas. Para exemplificar, no Quadro 5, partiremos da união dos primeiros elementos da lista, os termos *thermal* e *infrared*, pois, juntos, apareceram em um significativo número de vezes (352 ocorrências) no *corpus*:

Quadro 5 – Expressões formadas a partir do colocado *thermal* mais o nódulo *infrared*

<i>thermal infrared</i> =	<i>thermal infrared data</i>	39
	<i>thermal infrared image</i>	30
	<i>thermal infrared regions</i>	07
	<i>thermal infrared imagery</i>	32
	<i>thermal infrared remote sensing</i>	70
	<i>thermal infrared remote sensing systems</i>	06
	<i>Thermal Infrared Multispectral Scanner (TIMS)</i>	02

Estas informações facilitam a busca para o pesquisador identificar quais palavras formam uma colocação. Ao examinar essas informações no *corpus* comparável de língua inglesa, constituído por artigos científicos da área, pôde-se verificar a frequência de utilização de tais colocações encontradas no *corpus* de estudo.

Percebemos também uma presença de acrônimos e siglas,⁹ embora com um número menor de ocorrências. Costumam aparecer em geral entre parênteses e com letras maiúsculas. No caso do nódulo de estudo exemplificado (*infrared*), detectamos, como mostra o Quadro 6, os seguintes acrônimos e siglas:

Quadro 6 – Acrônimos e siglas formados a partir do nódulo *infrared*

<i>Near</i>	<i>Infrared</i>	(NIR)
<i>Airborne Visible</i>	<i>Infrared</i>	<i>Imaging Spectrometer (AVIRIS)</i>
<i>Thermal</i>	<i>Infrared</i>	<i>Multispectral Scanner (TIMS)</i>
<i>High Resolution Visible</i>	<i>Infrared</i>	(HRVIR)
<i>Long-Wavelength</i>	<i>Infrared</i>	(LWIR)
<i>Short Wavelength</i>	<i>Infrared</i>	(SWIR)
<i>Visible</i>	<i>Infrared</i>	<i>Scanner (VIRS)</i>
<i>Forward-Looking</i>	<i>Infrared</i>	(FLIR)

Mais especificamente no caso deste último dado, encontramos ainda as seguintes variações, apresentadas no Quadro 7:

Quadro 7 – Variações de uso do acrônimo FLIR

<i>Forward-Looking</i>	<i>Infrared</i>	<i>Sensor (FLIR)</i>
<i>Forward-Looking</i>	<i>Infrared</i>	<i>Systems (FLIR)</i>
<i>Forward-Looking</i>	<i>Infrared</i>	<i>Imagery (FLIR)</i>

O próximo passo foi buscar equivalentes ou correspondentes para tais colocações no subcorpus paralelo de português, visto que, no texto traduzido, a palavra “infravermelho” teve 710 ocorrências, enquanto no texto original foram 860. Vejamos, no Quadro 8, alguns resultados encontrados:

⁹ As siglas se formam pela redução de um grupo de palavras às suas iniciais e são lidas letra por letra (ex.: IBGE, INSS) e os acrônimos se formam através da junção dessas mesmas letras iniciais de um grupo de palavras, mas com a diferença de serem pronunciados como uma palavra só (ex.: Inpe, Ibama). Fonte: *DT-Dicionário Terminológico para consulta em linha do Ministério da Educação de Portugal*.

Quadro 8 – Equivalentes/correspondentes encontrados no sub*corpus* paralelo de língua portuguesa

<i>thermal infrared</i>	352	infravermelho termal	230
<i>thermal infrared data</i>	39	dados do(de) infravermelho termal	15
<i>thermal infrared image</i>	29	imagem do infravermelho termal	4
<i>thermal infrared imagery</i>	32	imagens do infravermelho termal	4
<i>thermal infrared region(s)</i>	7	região(ões) do infravermelho termal	8
<i>thermal infrared remote sensing</i>	70	sensoriamento remoto no infravermelho termal	63
<i>thermal infrared remote sensing systems</i>	6	sistemas sensores de infravermelho termal	3

Neste quadro de possíveis equivalentes para as colocações, encontramos para a colocação *thermal infrared data* o equivalente “dados *do* infravermelho termal”, com 8 ocorrências e “dados *de* infravermelho termal”, com 7, mostrando que as duas formas podem ser usadas. No entanto, ao buscar esta expressão no *corpus* comparável de português, tivemos um resultado bastante diferente. Encontramos 128 vezes o uso de “*do* infravermelho” e apenas 3 vezes o uso de “*de* infravermelho”. Esta expressiva diferença, que só pôde ser percebida por meio do *corpus* comparável, mostra que o tradutor optou por uma tradução “possível” em seu trabalho, porém, o bloco todo, “dados *do* infravermelho termal”, parece estar mais consagrado, em comparação com o bloco “dados *de* infravermelho termal”. Desta forma, um *corpus* comparável, composto por textos escritos originalmente nas duas línguas, inglês e português no caso em questão, fornecerá a forma efetivamente usada por aquela comunidade. Para Tagnin (2007, p. 2), um *corpus* é capaz de fornecer “o termo mais ‘provável’ de ocorrer no contexto que o tradutor estiver pesquisando, ou seja, o mais recorrente”.

Já para as colocações *thermal infrared image* e *thermal infrared imagery*, de acordo com o Dicionário de sensoriamento remoto (SELPER, 1989), temos para cada termo os seguintes equivalentes: *image* > imagem e *imagery* > imagens.

Desta maneira, baseado tanto no subcorpus paralelo de português como no dicionário, conclui-se que, para a expressão *thermal infrared image*, temos o equivalente “imagem do infravermelho termal” e para *thermal infrared imagery* > “imagens do infravermelho termal”. Em busca feita no corpus comparável, obtivemos também o uso da expressão “conjunto de imagens” como equivalente para *imagery*.

Com respeito à colocação *thermal infrared remote sensing systems*, pôde ser encontrada de duas maneiras no subcorpus traduzido: “sistemas sensores de infravermelho termal”, sendo usada preferencialmente, com 3 ocorrências, e “sistemas de sensoriamento remoto de infravermelho termal”, com 1 ocorrência.

No Quadro 9, apresentamos os equivalentes/correspondentes encontrados para os acrônimos a partir do nóculo *infrared*:

Quadro 9 – Equivalentes/correspondentes encontrados para os acrônimos do nóculo *infrared*

<i>Near Infrared (NIR)</i>	Infravermelho Próximo
<i>Airborne Visible Infrared Imaging Spectrometer (AVIRIS)</i>	Espectrômetro Imageador Visível e Infravermelho Aerotransportado
<i>Thermal Infrared Multispectral Scanner (TIMS)</i>	Escaneador Multiespectral no Infravermelho Termal
<i>High Resolution Visible Infrared (HRVIR)</i>	Alta Resolução no Visível e Infravermelho
<i>Long-Wavelength Infrared (LWIR)</i>	Infravermelho de Ondas Longas
<i>Short Wavelength Infrared (SWIR)</i>	Infravermelho de Ondas Curtas (Mantém-se em inglês no subcorpus traduzido)
<i>Visible Infrared Scanner (VIRS)</i>	
<i>Forward-Looking Infrared (FLIR)</i>	Infravermelho de Visada Avante

Somente os acrônimos AVIRIS, NIR e SWIR foram encontrados por suas iniciais no corpus comparável de português. Entre 8 acrônimos contendo o termo “infravermelho”, apenas três foram encontrados no corpus comparável, que tem mais que o triplo do tamanho do corpus de estudo. Isso pode indicar que, embora haja utilização dessas expressões nos estudos em sensoriamento remoto, menos da metade delas aparece na forma de

acrônimo ou sigla em língua portuguesa. Quando partimos para uma busca por meio das palavras-chave de cada acrônimo, por exemplo, “ondas longas”, para LWIR, pudemos encontrá-las no *corpus* comparável. No que concerne à busca destes acrônimos no sub*corpus* traduzido, a maioria das siglas pôde ser encontrada com mais facilidade, pois os tradutores procuraram manter uma tradução bastante próxima do original; mesmo assim, algumas diferenças foram notadas.

Especialmente nos casos de AVIRIS e SWIR, percebemos que estes acrônimos foram amplamente adotados pela comunidade científica e são capazes de expressar um sentido por si só. Na maioria das vezes em que foram encontrados no sub*corpus* traduzido ou no comparável, não vinham acompanhados de descrição por extenso, mas apenas pelas iniciais. O mesmo não ocorre com o acrônimo TIMS, que foi encontrado no sub*corpus* traduzido, porém como colocação especializada e não como acrônimo, ou seja, foi localizado apenas escrito por extenso, sem nenhuma menção à sua forma reduzida de sigla.

Já no caso de HRVIR, embora a sigla tenha sido utilizada no texto traduzido, a descrição em português no sub*corpus* de estudo (traduzido) foi dada apenas uma vez e depois se manteve somente em inglês. Esta sigla não foi encontrada no *corpus* comparável de português. O acrônimo VIRS foi utilizado no sub*corpus* traduzido, mas apenas em língua inglesa. Os tradutores optaram por não traduzir esta colocação especializada.

Como visto anteriormente, o acrônimo FLIR, raras vezes, apareceu sozinho e, em alguns casos, novas palavras foram agregadas a ele e misturadas em sua descrição. Para suas variações, encontramos os seguintes correspondentes no Quadro 10:

Quadro 10 – Equivalentes/correspondentes encontrados para as variações do acrônimo FLIR

<i>Forward-Looking Infrared Sensor</i> (FLIR)	sensor(es) de infravermelho de visada avante
<i>Forward-Looking Infrared Systems</i> (FLIR)	sistema sensor de infravermelho de visada avante
<i>Forward-Looking Infrared Imagery</i> (FLIR)	Imagens FLIR (Infravermelho de Visada Avante)

Nota-se que, em inglês, mesmo que a expressão seja apresentada por extenso e com a inserção de novos vocábulos, mantém-se o uso do acrônimo na maioria dos casos; em português, isso nem sempre ocorre. Na maioria das ocorrências encontradas, não havia nenhuma menção ao acrônimo. A única exceção encontrada nos *corpora* desta pesquisa é “Imagens FLIR”, em que foi apresentada diretamente a forma traduzida da expressão. Esta aparente “elasticidade” e a capacidade do acrônimo FLIR de agregar novas palavras em sua composição torna a busca mais complexa.

Conclusão

O sensoriamento remoto é uma área predominantemente desenvolvida no exterior. Países como Estados Unidos, China, Alemanha e Rússia, por exemplo, têm avançados centros de pesquisa e desenvolvimento destas técnicas para serem aplicadas em prol do bem-estar e defesa do ser humano. Aqui no Brasil, por ser um país onde esta disciplina pode ser considerada emergente, a fixidez lexical parece ainda não ser observada, ou seja, ainda notamos uma variação colocacional. Isso se explica em razão de, conforme defende Pavel (2003, p. 107), a maioria dos grupos fixos deriva das disciplinas-mãe. Neste caso, as colocações em português ainda parecem estar em processo de convencionalização.

Para alcançar os objetivos deste trabalho, é inegável a eficiência das ferramentas e dos utilitários oferecidos pelo software WordSmith Tools. A análise de uma grande quantidade de dados gerados de forma automática fica consideravelmente mais rápida e confiável.

A consulta aos *corpora* comparáveis também se mostra necessária para dar suporte a pesquisas desta natureza. A possibilidade de comparação entre as opções feitas pelos tradutores da obra que constitui o *corpus* paralelo com o *corpus* comparável nos permitiu verificar outras possíveis formas de tradução para as colocações encontradas. Dessa forma, o contraste do *corpus* de estudo em relação aos *corpora* comparáveis em inglês e português permitiu uma investigação mais completa das colocações especializadas na área de sensoriamento remoto.

Referências

- ACRÔNIMO. In: *Dicionário Terminológico para consulta em linha do Ministério da Educação de Portugal* – DT. Disponível em: <dt.dgicd.min-edu.pt>. Acesso em: 20 mar. 2013.
- ANDERSON, W.; CORBETT, J. *Exploring English with online corpora: an introduction*. London: Palgrave Macmillan, 2009. p. 45-66.
- BAKER, M. *Corpora in translation studies: an overview and some suggestions for future research*. *Target*, Amsterdam, v. 7, n. 2, p. 223-243, 1995.
- BARROS, L. A. *Curso básico de Terminologia*. São Paulo: EDUSP, 2004.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri, SP: Manole, 2004.
- CAMARGO, D. C. *Metodologia de pesquisa em tradução e linguística de corpus*. São Paulo; São José do Rio Preto: Cultura Acadêmica; Laboratório Editorial do IBILCE; UNESP, 2007. v. 1.
- CARNEADO MORÉ, Z. Consideraciones sobre la fraseografía. In: CARNEADO MORÉ, Z.; TRISTÁ PEREZ, A.M. *Estudios de Fraseología*. Havana: Academia de Ciências de Cuba; Instituto de Literatura y Lingüística, p. 39-46, 1983.
- CASARES, J. *Introducción a la lexicografía moderna*. Madrid: CSIC, 1950.
- COWIE, A. P. Introduction. In: COWIE, A. P. (Ed.). *Phraseology: theory, analysis, and applications*. Oxford: Oxford University Press, 1998. p. 1-19.
- CROITORU, E.; DUMITRAȘCU, A. M. Collocations and colligations in translation and interpreting. *Annales Universitatis Apulensis*, Alba Iulia, t. 2, 2006.
- MAGALHÃES, C. M. Pesquisas textuais/discursivas em tradução: o uso de *corpora*. In: PAGANO, A. (Org.). *Metodologias de pesquisa em tradução*. Belo Horizonte: FALE-UFGM, 2001.
- MOON, R. Dictionaries and collocations. In: GRANGER, S.; MEUNIER, F. (Eds.) *Phraseology: an interdisciplinary perspective*. Amsterdam: John Benjamins, 2008. p. 313-336.

NOVO, E. M. L. M. et al. Técnicas avançadas de sensoriamento remoto aplicadas ao estudo de mudanças climáticas e ao funcionamento dos ecossistemas amazônicos. *Revista Acta Amazônica*, v. 35, n. 2, p. 259-272, 2005.

ORENHA-OTTAIANO, A. *Unidades fraseológicas especializadas: colocações e colocações estendidas em contratos sociais e estatutos sociais traduzidos no modo juramentado e não-juramentado*. 2009. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto.

PAVEL, S. A fraseologia na língua de especialidade: metodologia de registro nos vocabulários terminológicos. In: FAULSTICH, E.; ABREU, S.P. (Org.). *Linguística Aplicada à Terminologia e à Lexicografia: Cooperação Internacional: Brasil e Canadá*. Porto Alegre: UFRGS; Instituto de Letras; NEC, 2003. p. 99-132.

SCOTT, M. *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software, 2012.

SELPER – Sociedad de Especialistas Latinoamericanos en Percepción Remota. *Dicionário SELPER – Sensoriamento Remoto*. [S.l.]: SELPER, 1989.

SELPER – Sociedad de Especialistas Latinoamericanos en Percepción Remota. *Dicionário Latinoamericano de Sensoriamento Remoto*. [s.d.]. Disponível em: <<http://www.inpe.br/unidades/cep/atividadescep/dicionariospelper/portugues/home.htm>>. Acesso em: 6 mai. 2013.

SIGLA. In: *Dicionário Terminológico para consulta em linha do Ministério da Educação de Portugal* – DT. Disponível em: <dt.dgicd.min-edu.pt>. Acesso em: 20 mar. 2013.

TOGNINI-BONELLI, E. Working with *corpora* across languages. In: TOGNINI-BONELLI, E. *Corpus Linguistics at work*. Amsterdam; Atlanta: John Benjamins, 2001.

***Corpus* principal (paralelo) de sensoriamento remoto**

JENSEN, J. R. *Remote sensing of the environment: an Earth resource perspective*. 2. ed. Minnesota: Pearson Prentice Hall, 2007.

JENSEN, J. R. *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres*. Tradução José Carlos Neves Epiphanyo et al. São José dos Campos: Parêntese, 2009.

***Corpus* de referência**

BRITISH NATIONAL *CORPUS* (BNC). Disponível em <<http://www.natcorp.ox.ac.uk/>>. Acesso em: 25 jul. 2012.

PROJETO LÁCIO-WEB. Compilação de *corpus* do português do Brasil e implementação de ferramentas para análises linguísticas. Disponível em: <<http://www.nilc.icmc.usp.br/lacioweb/index.htm>>. Acesso em: 25 jul. 2012.