

# Análise de expressões fixas e semifixas encontradas em textos da área da ciência da computação

(Analyzing Fixed and Semi-Fixed Expressions Found in Computer Science Texts)

**Silvelena Cosmo Dias**

Instituto de Biociências, Letras e Ciências Exatas – Universidade estadual Paulista  
(UNESP)

**diascosmo@yahoo.com.br**

**Abstract:.** This research study aims at examining the most frequent terms in English and their translation into Portuguese present in our corpus represented by texts of Computer Security, one of the sub-fields in the area of Computer Science. Our investigation has its theoretical starting point based on Corpus-Based Translation Studies, Corpus Linguistics, and Terminology. This study used WordList and Concord tools, provided by the software WordSmith Tools version 4.0. After listing and analyzing the terms of Computer Security, we compiled a glossary with the most frequent terms in English and the functionally equivalent terms translated into Portuguese followed by their cotexts in both languages.

**Keywords:** Computer Science; Translation; Corpus-Based Translation Studies; Corpus Linguistics, fixed and semi-fixed expressions.

**Resumo:** Este estudo objetiva examinar os termos de maior frequência em inglês e suas traduções em português presentes em nosso corpus representado pelos textos de Segurança de Computadores, uma das subáreas da área da Ciência da Computação. Nossa investigação fundamenta-se no arcabouço teórico-metodológico dos Estudos da Tradução Baseados em Corpus, Linguística de Corpus, e Terminologia. Este estudo usou as ferramentas *WordList* e *Concord*, disponibilizadas pelo software *WordSmith Tools* versão 4.0. Depois da listagem e análise dos termos de Segurança de Computadores, compilamos um glossário com os termos mais frequentes em inglês e seus termos funcionalmente equivalentes traduzidos para o português seguidos do seu cotexto em ambas as línguas.

**Palavras-chave:** Ciência da Computação; Tradução; Estudos da Tradução Baseados em Corpus; Linguística de Corpus; expressões fixas e semifixas.

## Introdução

Apesar da importância dessa área de estudo ser bem conhecida e o número de pessoas que lidam com seu estudo estar aumentando, a maioria dos livros relacionados a esse campo são escritos em língua inglesa, e apenas alguns deles são traduzidos para o português. Dessa maneira, a tradução desses livros especializados é necessária não somente para as pessoas que não tenham certo domínio da língua inglesa, mas também para pesquisadores, professores e alunos nesse campo.

São os seguintes os objetivos propostos para a presente pesquisa:

1. Observar aproximações e distanciamentos no uso de equivalentes funcionais dos termos de maior frequência, encontrados no texto traduzido da Segurança de Computadores para o português, em relação aos termos presentes no respectivo texto originalmente escrito em inglês;
2. Compilar uma amostra de um glossário com os termos de maior frequência traduzidos para o português e seus termos funcionalmente equivalentes em inglês, seguidos de seus contextos, extraídos dos três corpora.
3. Comparar os termos simples de maior frequência, expressões fixas e semifixas de Segurança de Computadores presentes no subcorpus principal do texto traduzido com os termos do subcorpus do texto original;
4. Comparar os termos de maior frequência de Segurança de Computadores no subcorpus principal de texto original em relação aos termos de maior frequência no corpus comparável de texto originalmente escrito em inglês;
5. Comparar os termos de maior frequência de Segurança de Computadores no subcorpus principal de texto traduzido com os termos de maior frequência do corpus comparável de textos originalmente escritos em português.

O objeto central da Linguística de Corpus é o corpus. De acordo com Baker (1995), o termo “corpus” refere-se a uma coleção de textos que podem ser processados por computador e analisados ambos automaticamente e semi - automaticamente de várias maneiras. Pode incluir não só textos escritos, mas também textos orais, dependendo dos objetivos do pesquisador. Em geral, um corpus consiste de uma grande quantidade de textos originais de muitas fontes baseados em uma seção vasta de assuntos. Isso significa que textos são selecionados de acordo com os objetivos pretendidos a atingir sob investigação. O corpus tem que ser representativo de um tipo particular da linguagem em uso.

O arcabouço teórico-metodológico proposto para pesquisas baseadas em corpora de textos traduzidos (Baker, 1993) exige análises de corpora realizados por ferramentas eletrônicas. Esse método procura “identificar características de textos traduzidos que nos ajudarão a entender o que é tradução e como ela funciona<sup>1</sup>” (BAKER, 1993, p. 243).

Dois tipos de corpora são usados nessa pesquisa: um corpus paralelo e outro comparável. O corpus paralelo é definido por Baker como sendo “uma coleção de textos fonte e suas respectivas traduções<sup>2</sup>” (1993 p. 248).

Já um corpus comparável, segundo Tognini Bonelli (2001) é constituído por um corpus comparável de textos originalmente escritos na língua 1 (L1, no caso da presente pesquisa, o inglês), um corpus de estudo com textos escritos originalmente na língua 2 (L2, no caso do presente trabalho, o português).

Munday define Estudos da Tradução como:

...uma nova disciplina relacionada ao estudo da teoria e dos fenômenos da tradução. Por sua natureza ela é multilíngüe e também interdisciplinar, abarcando as línguas, a

---

<sup>1</sup> [...] identify features of translated text which will help us understand what translation is and how it works.

<sup>2</sup> Parallel corpora, that is corpora of source texts and their translations.

lingüística, os estudos da comunicação, a filosofia e vários outros tipos de estudos culturais<sup>3</sup>. (MUNDAY, 2001, p. 1)

Até 1960, uma exigência pela precisão e fidelidade, e a noção de equivalência dominou todas as discussões sobre tradução. Grandes esforços foram feitos para que os textos traduzidos pudessem ser tão equivalentes quanto possíveis aos originais. O objetivo era atingir a tradução ideal em relação ao texto original.

Nos últimos anos, várias tentativas têm sido feitas para diminuir a noção de primazia do texto fonte. Um grande número de estudos tem sido produzido para mudar a orientação dos Estudos da Tradução de estudos prescritivos para estudos descritivos, do significado para o uso e de meramente equivalência para um tipo de equivalência dentro de certo contexto.

Sara Laviosa, ex-aluna na UMIST – *University of Manchester Institute of Science and Technology*, sob a supervisão de Baker, considera que:

Os Estudos da Tradução Baseados em Corpus representam uma área de pesquisa que tem atraído um número crescente de pesquisadores entusiastas que acredita firmemente em seu potencial de fornecer informação para projetos bem elaborados realizados no mundo todo bem como de reconciliar a pluralidade de necessidades e interesses dentro da disciplina<sup>4</sup>. (LAVIOSA, 2002, p.33)

Atualmente, muitos estudiosos na área da tradução têm usado corpora eletrônicos para suas investigações. A natureza da Lingüística de Corpus fundamenta-se a partir de uma base empirista, uma vez que favorece a análise de dados provenientes da observação da linguagem sob a forma de corpus computadorizado, possibilitando a análise de grandes quantidades de informações.

A Lingüística de Corpus é uma abordagem teórico-metodológica que surgiu a partir das investigações de Sinclair (1991). Lingüistas a empregam para investigar aspectos da linguagem em uso na maioria das áreas lingüísticas em geral. Isto significa que uma grande quantidade de textos é coletada, armazenada, manipulada e analisada.

Baker lançou sua proposta para os Estudos da Tradução Baseados em Corpus em 1993, e observou características específicas da linguagem da tradução. Essas características são: simplificação, explicitação, normalização (ou conservacionismo) e estabilização.

Baker declara que corpora de tradução e metodologia disponibilizada pela Lingüística de Corpus podem conduzir a detectar aspectos na linguagem produzidos pelos tradutores.

Com respeito à simplificação, uma das observações possíveis, considera a quantidade de *tokens* (itens: palavras corridas) e *types* (vocábulo: forma) em um texto.

---

<sup>3</sup> Translation Studies is the new academic discipline related to the theory and phenomena of translation. By its nature it is multilingual and also interdisciplinary, encompassing languages, linguists, communication studies, philosophy and a range of types of cultural studies.

<sup>4</sup> Corpus-Based Translation Studies represent an area of research that is attracting a growing number of enthusiastic scholars who genuinely believe in its potential for informing well thought-out projects throughout the world and for reconciling the plurality of needs and interests within the discipline.

Qualquer seqüência de letras com um espaço em ambos os lados é contada como palavra, ou mais precisamente, um *token*. Sendo assim, a palavra *dia* é contada como um *token* e nós podemos dizer que existem *x tokens da* palavra *dia* em um dado corpus. A forma de palavra *day* por si é um *type*, não importando quantas vezes aparece no texto<sup>5</sup>. (BAKER, 1995, p.236)

A comparação de *type/token* de um corpus original e de um corpus de texto traduzido pode ajudar o pesquisador a captar características específicas da linguagem de tradução que não são mostradas nem no texto original e nem em outros originais.

Nessa pesquisa, preocupada com os termos de Segurança de Computadores, estamos interessados em identificar os termos especializados de maior frequência nessa subárea. Esses termos têm uso específico dentro do campo, que não têm necessariamente o mesmo como no uso da linguagem geral. Assim adentramos no campo da Terminologia, que é o estudo dos termos e seus usos em específicos contextos ou em linguagens de especialidade. Seu estudo não é uma novidade, como profissionais de cada área têm sempre se preocupados com o vocabulário específico de seus campos de conhecimento. No entanto, esse estudo é relativamente recente, tendo começado na metade do século 20. (Krieger and Finato, 2004).

Tradução técnica ou especializada tem sido cada vez mais usada devido ao progresso da tecnologia e o desenvolvimento do relacionamento comercial entre os países de toda a parte do mundo. Então, um tradutor é sempre alguém que precisa da contribuição da Terminologia, especialmente se ele trabalha com os termos específicos de uma determinada área. Terminologia ajuda na prática da tradução.

No campo da Terminologia, define “termo simples” como “constituído de um só radical, com ou sem afixos” (ISO 1084, 1990: 7, *apud* BARROS, 2004, p.100).

Com referência a “expressões fixas” na disciplina dos Estudos da Tradução Baseados em Corpus, de acordo com Baker (1992, p.63) elas são padrões cristalizados (*frozen*) da língua que permitem pouca ou nenhuma variação na forma, por exemplo, *as a matter of fact, all the best*. Com respeito a “expressões semifixas”, situam-se em área menos extremas da escala de colocações, estas abarcando de um lado a flexibilidade de padrões, de outro a transparência de sentido (CAMARGO, 2005).

Nessa pesquisa, adotamos o termo “glossário”, baseado em Barros (2004):

Glossário (termo tolerado: dicionário bilíngüe, dicionário multilíngüe): pode situar-se tanto no nível do sistema como no da(s) norma(s). Sua principal característica é não apresentar definições, mas tão somente uma lista de unidades lexicais ou terminológicas acompanhadas de seus equivalentes em outras línguas. (BARROS, 2004, p.144)

No campo da Lingüística de Corpus, usando a ferramenta *Concord*, disponibilizada pelo software *WordSmith Tools*, cinco ou mais palavras à direita e cinco ou mais palavras à esquerda podem ser adicionadas aos termos simples no nosso glossário. Isto significa o cotexto em que os termos simples aparecem. De acordo com

<sup>5</sup> Any sequence of letter with an orthography space on either side is counted as a word or, more precisely, a token. Thus each occurrence of the word day is counted as an individual token and we can say that there are x tokens of day in a given corpus. The word-form day itself is a type, no matter how often it occurs. So we can say that there are x tokens of the type day in a corpus of y million words.

Berber Sardinha (2004, p.105), cotexto é o texto ao redor da palavra de busca, também chamado de nóculo.

De acordo com Camargo (2005), equivalência pode ser uma questão central em tradução, embora sua definição, relevância e aplicabilidade dentro do campo da tradução tem causado controvérsia.

Com relação à “equivalência funcional”, Halliday enfatiza que:

Se sentido é função em contexto, então, equivalência de sentido é equivalência de função em contexto. O que o tradutor faz quando traduz ou interpreta é tomar decisões, o tempo todo, a respeito de qual é o contexto relevante em que essa equivalência funcional se estabelece. (HALLIDAY, 1992. p.16)

Nessa investigação, temos um corpus paralelo, constituído por um livro originalmente escrito em inglês e sua respectiva tradução para o português. Nós também temos dois corpora comparáveis, um composto de um livro originalmente escrito em inglês, e o outro formado por três livros originalmente escritos em português. O domínio de todos os livros é da Segurança de Computadores, uma subárea da Ciência da Computação.

Para a compilação do corpus comparável em português, usamos mais livros porque eles contêm um número menor de palavras (palavras corridas ou *tokens*) do que o corpus comparável em inglês, formado por apenas um livro. Dessa maneira, obtemos os corpora comparáveis mais balanceados.

O corpus paralelo, constituído por um livro originalmente escrito em inglês, publicado em 2005, contém 117.126 palavras e sua tradução para o português, publicado um ano depois que seu original, tem 114.559 palavras. O corpus comparável composto de um livro originalmente em inglês (2006) tem 288.355 palavras. O outro corpus comparável é constituído de três livros originalmente escritos em português e contêm um total de 94.432 palavras. Eles foram publicados entre 1998 e 2003.

Os livros foram selecionados com a ajuda de um profissional na área da Ciência da Computação.

Também, dois dicionários foram usados para confirmar se os termos simples, expressões fixas e semifixas pertencem à área específica de Segurança de computadores.

Quanto aos procedimentos metodológicos adotados nessa pesquisa, tomamos por fundamentação a proposta de Baker (1993, 1995, 1996) e os trabalhos de Berber Sardinha (2000, 2004). O software usado nesse estudo foi o *WordSmith Tools*, criado por Mike Scott, da Universidade de Liverpool. É uma ferramenta computacional aplicada na análise de corpus e é considerado o mais completo e versátil conjunto de ferramentas para a Lingüística de Corpus.

Para o primeiro passo metodológico, os livros disponíveis *on-line* em formato de PDF foram salvos no tipo de texto sem formatação (“txt”), “*Text Only*”, usando o programa chamado “*PDF Text Reader*”. O formato “*Text Only*” é necessário para a extração de informações executada pelo *WordSmith Tools*. Também, os textos foram salvos em *Word*, para que pudessem ser visualizados na tela do computador com mais detalhes.

O desenvolvimento dessa pesquisa foi feito em três partes:

1. Extração das listas dos termos simples, expressões fixas e semifixas do subcorpus principal (corpus paralelo) de Segurança de Computadores;
2. Extração das listas dos termos de ambos os corpora comparáveis;
3. A organização do glossário.

1. Extração das listas dos termos simples, expressões fixas e semifixas do *corpus* principal

O corpus paralelo (textos dos subcorpora em inglês e português) foi analisado primeiro em relação aos corpora comparáveis. A ferramenta usada foi o *WordList* que permite ao pesquisador ver as palavras dos textos escolhidos exibidas em três listas: uma lista mostra as palavras mais freqüentes, uma lista de todas as palavras em ordem alfabética, e uma lista contendo as estatísticas dos *tokens* (itens: palavras corridas – *running words*) no texto, *types* (forma: vocábulo - palavras distintas), razão forma/item (FI: *type/token ratio*) e razão forma/item padronizada (*standardised type/token* - TTR).

A lista de freqüência do subcorpus principal de Segurança de Computadores foi observada a fim de identificar os termos de maior freqüência e contrastá-los com a lista de freqüência do corpus comparável formado por texto originalmente escrito em inglês. No entanto, nem todos os termos pertencem à área especializada do estudo. Dessa maneira, dois dicionários especializados foram usados para confirmar se eles podiam ser considerados termos especializados nesse campo.

Uma segunda lista de freqüência do subcorpus principal do texto traduzido foi gerada para mais tarde ser contrastada com a lista de freqüência do *corpus* comparável formado por textos originalmente escritos em português.

Outra ferramenta que nós usamos foi o *Concord* que permite o *WordSmith Tools* a especificar uma palavra de busca. Essa ferramenta também gera *clusters*, isto é, um grupo de palavras que vem junto. Eles são exibidos em uma lista que ajuda o pesquisador a identificar se um termo é simples ou se é uma expressão fixa ou semifixa. A ferramenta também mostra a freqüência desses termos simples ou expressões fixas e semifixas.

Dessa maneira, temos duas listas de termos simples seguido de suas freqüências respectivamente extraídas do texto original e do texto alvo do corpus de estudo de Segurança de Computadores. Também, temos duas listas de expressões fixas e semifixas, uma em inglês e a outra em português.

O próximo passo foi observar os termos simples, expressões fixas e expressões semifixas de maior freqüência nas listas. Então, extraímos listas de concordância com os termos simples em inglês, e listas em português, ambas as listas apresentando os termos inseridos em seus cotextos. Também, geramos listas de concordância com expressões fixas e semifixas com seus cotextos em inglês, bem como em português.

2. Extração das listas dos termos dos corpora comparáveis

Os mesmos procedimentos foram adotados a fim de obter as listas de termos simples, expressões fixas e semifixas de maior freqüência de Segurança de Computadores nos corpora comparáveis de textos originalmente escritos em inglês e em português, usando as ferramentas *WordList* e *Concord*. A fim de confirmar suas

especificidades, também usamos dicionários especializados. Desse modo, obtemos uma lista de termos de maior frequência em L1 e outra em L2.

O próximo procedimento foi verificar se os termos que foram selecionados no subcorpus principal (corpus paralelo) estavam presentes nos corpora comparáveis de textos originalmente escritos em inglês e em português. Três termos simples de maior frequência seguidos de suas expressões fixas e semifixas com seus cotextos foram selecionados para a organização de uma amostra de glossário.

### 3. Organização do glossário

Uma amostra de um glossário bilingüe (inglês/português) foi construída com os três termos de maior frequência no subcorpus principal de um texto originalmente escrito em inglês e sua respectiva tradução, e nos corpora comparáveis de textos originalmente escritos em inglês e em português. Os termos simples, expressões fixas e semifixas foram seguidas de suas traduções encontradas no subcorpus principal do respectivo texto traduzido, bem como os termos simples, expressões fixas e semifixas correspondentes encontradas nos corpora comparáveis de L1 e L2. Eles foram organizados em ordem alfabética para a comodidade do leitor.

As definições dos termos não foram apresentadas nessa pesquisa, porque nosso interesse é observar os termos em textos originalmente escritos em inglês e suas traduções em português e compará-los aos encontrados em textos originalmente escritos em inglês e português. No entanto, estudos futuros poderão ser desenvolvidos por terminólogos a fim de apresentar definições.

### **Análise dos Resultados**

A ferramenta *WordList* gerou listas de palavras por ordem de frequência, extraídas do corpus principal. A lista dos dez primeiros termos especializados de maior frequência presentes no subcorpus principal do texto original em inglês pode ser observada na tabela 1, abaixo:

Tabela 1: Dez termos simples do subcorpus principal em inglês:

	<b>N Word</b>	<b>Freq.</b>	<b>%</b>
1	SECURITY	309	0,263
		8	
2	COMPUTER	305	0,260
		4	
3	ACCESS	286	0,244
		2	
4	TIME	280	0,239
		1	
5	NETWORK	274	0,233
		9	
6	SYSTEM	264	0,225
		4	
7	INFORMATION	253	0,216

8	SERVER	249	6	0,212
9	COMPANY	239	1	0,204
10	PASSWORD	236	5	0,201

Os *clusters* que geramos, usando a ferramenta *Concord*, tabela 2 abaixo mostra as expressões fixas e semifixas com os três termos simples recorrentes de maior frequência do subcorpus principal em inglês.

Tabela 2: Expressões fixas e semifixas do subcorpus principal em inglês:

<p>SECURITY (309)</p> <ol style="list-style-type: none"> <li>1. computer security</li> <li>2. information security</li> <li>3. security policy</li> <li>4. security sources</li> <li>5. techno-security issue</li> </ol> <p>COMPUTER (305)</p> <ol style="list-style-type: none"> <li>1. college computer science major</li> <li>2. computer-based game</li> <li>3. computer-based products</li> <li>4. computer chip</li> <li>5. computer code</li> <li>6. computer equipment</li> <li>7. computer field</li> <li>8. computer hackers</li> <li>9. computer hardware</li> <li>10. computer-in-the-pocket</li> <li>11. computer moden</li> <li>12. computer network</li> <li>13. computer program</li> </ol>
---

14. computer-related jobs
15. computer science lab
16. computer security measures
17. computer technology frensics
18. computer wiretap

#### ACCESS (286)

1. administrative access rights
2. Internet access
3. secure Web access
4. unauthorized access
5. wireless access point

Selecionamos, em ordem alfabética, os termos simples “access”, “computer” e “security”, por estar em todas listas dos termos de maior frequência na subárea de Segurança de Computadores dos termos simples, expressões fixas e semifixas.

As expressões fixas e semifixas do termo “access” e suas respectivas traduções para o português são: “administrative access right”→“direitos de acesso administrativo”, “Internet access”→“acesso à Internet”, “secure Web access”→“acesso seguro à Web”, “unauthorized access”→“acesso não autorizado” e “wireless access point”→“pontos de acesso sem fio”.

Para o termo simples “computer”, selecionamos cinco expressões fixas e semifixas recorrentes: “computer-based products”→“produtos informatizados”, “computer chip”→“chip de computador”, “computer code”→“código de computador”, “computer hackers”→“hackers de computadores” e “computer technology”→“tecnologia de computadores”.

Com relação ao termo “security”, encontramos cinco expressões fixas e semifixas: “computer security”→“segurança de computadores”, “information security”→“segurança de informação”, “security policy”→“política de segurança”, “security sources”→“fontes de segurança” e “techno-security issue”→“questão de segurança tecnológica”.

De acordo com as definições dadas por Aubert (1998), essas traduções podem ser classificadas como “transposição”. O autor distingue “transposições” de “tradução literal”, considerando que:

*Tradução literal* é sinônimo de tradução palavra-por-palavra e em que, comparando-se os segmentos textuais fonte e meta, se observa: (i) o mesmo número de palavras, (ii) na mesma ordem sintática, (iii) empregando as ‘mesmas’ categorias gramaticais e (iv)

contendo as opções lexicais que, no contexto específico, podem ser tidas por sendo sinônimos interlingüísticos, como em: Her name is Mary→Seu nome é Maria. *Transposição* ocorre sempre que pelo menos um dos três primeiros critérios que definem a tradução literal deixa de ser satisfeito, ou seja, sempre que ocorrem rearranjos morfossintáticos. Assim, por exemplo, se duas ou mais palavras forem fundidas em uma única (como em **I visited**→Visitei) ou em várias unidades lexicais (por exemplo, **kindergarden**→Jardim de Infância), ou se a ordem das palavras for alterada (inversões e deslocamentos, como em **remedial action**→ação saneadora), ou se houver uma alteração gramatical (por exemplo, (**should he arrive late**→**se** ele chegar atrasado) ou quaisquer combinações dos anteriores, por mais ‘literais’ que os respectivos significados se apresentem, não constituirão segmentos textuais estruturalmente literais, sendo, assim, classificados como *transposições*. (AUBERT, 1998, p.106-107)

Dessa maneira, observamos que quase todos os termos nas listas do corpus paralelo (textos fonte e alvo) aparecem nas listas dos corpora comparáveis de textos originalmente escritos em inglês e português.

Abaixo, a tabela 3 mostra os dez primeiros termos especializados de maior frequência de Segurança de Computadores encontrados no corpus comparável de textos originalmente escritos em inglês:

Tabela 3: dez termos simples do corpus comparável em inglês:

<b>N</b>	<b>Word</b>	<b>Freq.</b>	<b>%</b>
1	SECURITY	1899	0,6
		586	
2	INTERNET	1108	0,3
		842	
3	NETWORK	995	0,3
		451	
4	INFORMATION	762	0,2
		643	
5	UNIX	757	0,2
		625	
6	FILE	690	0,2
		393	
7	USER	675	0,2
		341	
8	ACCESS	553	0,1
		918	
9	SERVER	500	0,1
		734	
10	PASSWORD	489	0,1
		696	

Comparando os dez primeiros termos de maior frequência do corpus comparável em inglês com os dez primeiros do subcorpus principal do texto original, seis deles estão presentes no corpus principal: “security”, “network”, “information”, “access”,

“server” e “password”. Dessa maneira, quase todos os termos no subcorpus principal ocorrem no corpus comparável em inglês.

Desde que os termos “access”→“acesso”, “computer”→“computador” e “security”→“segurança” registram alta frequência em todas as listas dos corpora paralelo e comparáveis, eles foram escolhidos para fazer parte do glossário. No entanto, eles não foram encontrados com todas as expressões fixas e semifixas nos dois corpora.

Por exemplo, as expressões fixas e semifixas “administrative access rights”→“direitos de acesso administrativo”, “secure Web access”→“acesso seguro à Web”, “wireless access point”→“pontos de acesso sem fio” não foram encontradas no corpus comparável de textos originalmente escritos em inglês e em português. Por outro lado, “Internet access”→“acesso à Internet”, “unauthorized access”→“acesso não autorizado” foram encontrados em ambos os corpora comparáveis de textos originalmente escritos em inglês e em português.

As expressões fixas e semifixas com o termo “computer”: “computer-based products”→“produtos informatizados”, e “computer chip”→“chip de computador” não foram encontrados em ambos os corpora de textos originalmente escritos em inglês e em português. “computer code”→“código de computador” e “computer hackers”→“hackers de computadores” não foram encontrados nos corpora comparáveis de textos originalmente escritos em português, mas foram encontrados no corpus comparável de texto originalmente escrito em inglês. Por outro lado, “computer technology”→“tecnologia de computadores” estava presente em ambos os corpora de textos originalmente escritos em inglês e em português.

As expressões fixas e semifixas “information security”→“segurança de informação”, e “security sources”→“fontes de segurança” extraídas do corpus paralelo não foram encontradas no corpus comparável de textos originalmente escritos em português, mas foram encontradas no corpus comparável de texto originalmente escrito em inglês. Por outro lado, “techno-security issue”→“questão de segurança tecnológica” não foi encontrada em ambos os corpora comparáveis de textos originalmente escritos em inglês e nem em português, e, “computer security”→“segurança de computadores”, “security policy”→“política de segurança” estavam presentes em ambos os corpora de textos originalmente escritos em inglês e em português.

Por meio desta pesquisa realizada, foi possível constatar que a disponibilidade do software *WordSmith Tools* capacita o pesquisador a realizar análise automaticamente através dos dados armazenados em um corpus de uma maneira mais rápida e mais precisa do que análises feitas apenas manualmente.

Com relação a aproximações observadas entre o corpus principal e os corpora comparáveis, podemos observar que a maioria dos termos simples levantados no texto original encontra equivalência no texto alvo, e também está presente nos corpora comparáveis como, por exemplo: “security”→“segurança”, “computer”→“computador”, “access”→“acesso”, “system”→“sistema”, “server”→“servidor”, “password”→“senha”.

Quanto aos distanciamentos entre os corpora da pesquisa de estudo, algumas expressões fixas e semifixas, usadas na organização do glossário, não estão presentes nos corpora comparáveis de inglês e de português como, por exemplo: “techno-security issue”→“questão de segurança tecnológica”.

Os resultados obtidos podem contribuir para as pesquisas no campo dos Estudos da Tradução Baseados em Corpus, ajudar alunos, professores, tradutores, bem como profissionais na área computacional.

Além de todas essas aplicações, podemos destacar que uma pesquisa dessa natureza pode também contribuir para a melhoria do ensino de Inglês para Fins Específicos em cursos tais como: Ciência da Computação, Engenharia da Computação e Sistemas de Informação.

## REFERÊNCIAS BIBLIOGRÁFICAS

AUBERT, F. H. *Introdução à metodologia da pesquisa terminológica bilingüe*. São Paulo: Humanistas Publicações - FFLCH/USP, 1996.

\_\_\_\_\_. *Translation Modalities: Theory and Practical Results*. *TradTerm*. n. 5, 1º. Sem. 1998, p.129-157.

BAKER. M. *In other words: a coursebook on translation*. London: Routledge, 1992.

\_\_\_\_\_. *Corpus linguistics and translation studies: implications and applications*. *Text and technology*: in honour of John Sinclair. Amsterdam/Philadelphia: John Benjamins, 1993, p.133-250.

\_\_\_\_\_. *Corpora in Translation Studies: an overview and some suggestions for future research*. *Target*. v. 7. n. 2. John Benjamins B. V. Amsterdam, 1995, p. 223-243.

\_\_\_\_\_. *Corpus-based translations studies: the challenges that lie ahead*. Somers, Harold, ed. *Terminology, LSP and translation studies in language engineering in honour of Juan C. Sager*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1996, p. 176-189.

\_\_\_\_\_. *A corpus-based view of similarity and difference in translation*. *Translation similarity and difference*. Manchester: St. Jerome, 2004, p. 1-17.

BARROS, L. *A Curso básico de terminologia*. São Paulo: USP, 2004.

BERBER SARDINHA, A. P. *Linguística de corpus: histórico e problemática*. *DELTA*, v. 16, n. 2, 2000, p. 323-367.

\_\_\_\_\_. *Linguística de corpus*. Barueri, SP: Manole, 2004.

CAMARGO, D. C. *Os textos jurídico e técnico e as modalidades tradutórias*. *Boletim do CITRAT II*. São Paulo: USP, 1995.

\_\_\_\_\_. *Corpus-based translation research on legal, technical and corporate texts*. *Across Languages and Cultures*. v. 2. n. 1. Budapest, 2001a, p. 113-115.

\_\_\_\_\_. *Estudos tradutológicos baseados em corpus de textos técnicos, corporativos e jornalísticos*. *Tradução e Comunicação Revista Brasileira de Tradutores*. São Paulo, v. 10. 2001b, p. 33-46.

\_\_\_\_\_. *Padrões de estilo de tradutores - PETra: investigação em Corpora de traduções literárias, especializadas e juramentadas*. 25/fev/2004. Projeto de pesquisa

apresentado como requisito parcial para a provação do Plano Trienal para o triênio 2004-2006, Instituto de Biociências, Letras e Ciências Exatas, UNESP/SJRP, 2004.

\_\_\_\_\_. *Padrões de estilo de tradutores: Um estudo de semelhanças e diferenças em corpora de traduções literárias, especializadas e juramentadas.* (Stylistic patterns of translators: A study of similarities and differences in corpora of literary, specialized and sworn translations). 2005. 512 p. Privatdozent Dissertation in Translation Studies – UNESP (University of the State of São Paulo), campus of São José do Rio Preto, São Paulo, Brasil, 2005.

DUBUC, R. *Manuel pratique de terminologie*. 2.ed.. Québec: Linguatex, 1985, 1992.

*INTERNATIONAL dictionary of english*. London: Cambridge University Press, 1996.

KRIEGER, M. G.; FINATTO, M. J. B. Finatto. *Introdução à terminologia: teoria e prática*. São Paulo: Contexto, 2004.

MAGALHÃES, C. M. *Metodologias de pesquisa em tradução*. Belo Horizonte: Ed. Da UFMG, (Estudos Lingüísticos, 3), 2001.

MUNDAY, J.. *Introducing translation studies: theories and applications*. New York: Routledge, 2001.

PAIVA, P. T. P. *Estudo baseado em corpora de traduções e três glossários bilíngües nas subáreas de anesthesiologia, cardiologia e ortopedia*. 2006. 223 f. Dissertação (MA Thesis) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2006.

SAWAYA, M. R. *Dicionário de informática e internet*. São Paulo: Nobel, 1999.

THING, L. *Whatis? com's: dicionário de tecnologia*. Trad. Por Bazán tecnologia e Lingüística e Texto Digital. São Paulo: Futura, 2003.

TOGNINI-BONELLI, E. *Corpus linguistics at work*. Amsterdam: John Benjamins, 2001.

\_\_\_\_\_. *Functionally complete units of meaning across English and Italian: Towards a Corpus-driven Approach*. Altenberg, Bengt and Grandier, Sylviane. *Lexis in contrast: corpus-based approaches*. Amsterdam: John Benjamins, 2002.