

4

A Linguística Sistêmico-Funcional e a Linguística de *Corpus*

Este capítulo objetiva fazer uma breve apresentação do cenário geral do funcionalismo, dos funcionalistas e dos tipos de funcionalismo mais presentes na literatura estudada, com ênfase no funcionalismo britânico, e cuja importância no panorama dos estudos da linguagem se fez notar em função do legado de estudos e pesquisas na área. Além disso, pretende levantar alguns pressupostos relacionados à abordagem de *Corpus* utilizada nesta pesquisa.

4.1. Cenário Geral

Cabe apresentar, inicialmente, algumas concepções de Funcionalismo, cuja origem está na Antropologia, através das pesquisas de Malinowski (1922) e Redcliffe Brown (1952) (*apud* Macedo, 1998. p. 71-88).

Por meio das tentativas de explicação dos fatos culturais em função de estruturas sociais mais abrangentes, Malinowski, nos estudos dos trobrianeses do Pacífico Sul, apresentava o que se denomina postura “teleológica”. Esse termo, tido como chave no Funcionalismo, emprega-se em estudos que buscam explicar e analisar uma coisa em função de outra. Assim, a língua, na perspectiva funcionalista, está associada ao preenchimento de funções de tal forma que a estrutura da gramática pode ser explicada como sendo resultado de funções de outras esferas como as dos níveis cognitivos e comunicativos, por exemplo. Evidenciar, portanto, de que modo a estrutura gramatical espelha a situação comunicativa é o ponto central do enfoque funcionalista.

Halliday, um dos precursores do funcionalismo britânico, aponta que “a natureza da língua está intimamente relacionada às necessidades que lhe impomos, com as funções a que deve servir” (Halliday, 1970, p. 141). É ainda desse autor a ideia de que as redes relativamente independentes de escolhas em torno das quais a língua se organiza correspondem a certas funções básicas da linguagem. Compreendendo que acima dos sistemas linguísticos e dos falantes há

a realidade da linguagem, Halliday indica que esta cumpre três funções fundamentais, além da comunicativa: expressão de conteúdos, estabelecimento e manutenção de relações sociais e organização e codificação de significados de forma linear e coerente.

O pressuposto básico da teoria funcionalista em geral é o fato de a estrutura gramatical espelhar parcialmente a situação comunicativa. Consequentemente, a linguagem é entendida como instrumento de interação social. A língua, nesse âmbito, será, então, resultante das necessidades/criatividades dos falantes.

Pelo fato de a ênfase no estudo da língua recair sobre seu usuário, nas situações comunicativas buscar-se-ão as motivações para os usos diversos desse instrumento. Na verdade, as regularidades observadas na utilização da língua em situações concretas de comunicação podem revelar condições discursivas específicas para a ocorrência de determinadas estruturas em lugar de outras.

Como uma perspectiva funcionalista de análise linguística considera, necessariamente, as situações reais de comunicação, os aspectos extralinguísticos revelam-se também foco de preocupação nas descrições dos procedimentos linguísticos. Importante ainda observar que essas descrições estarão centradas nas regularidades e ainda nas irregularidades presentes nesses discursos.

O Funcionalismo em linguística vem propor uma abordagem da língua que não a considere um elemento autônomo para cuja análise se prescindia da observação do fenômeno em situações reais de comunicação. Ao contrário disso, em uma abordagem funcionalista, há que se levar em conta que as pressões oriundas dos eventos comunicativos influenciarão inexoravelmente a estrutura gramatical que alicerça qualquer idioma.

A teoria sistêmica postula que compreender a língua sob o ponto de vista sistêmico-funcional implica conceber que cada forma linguística resulta da função que o item ou a estrutura gramatical desempenham no evento comunicativo oral ou escrito. Se a língua é constituída por uma rede de sistemas, podem-se fazer escolhas significativas que cumprem funções.

Para interagir socialmente, os usuários de uma língua fazem uso desse sistema uma vez que a troca no meio social se dá essencialmente via linguagem verbal. Dentre vários estudiosos que adotaram uma perspectiva funcionalista de abordagem gramatical, Michael Alexander Kirkwood Halliday legou aos estudos da linguagem a Gramática Sistêmico-Funcional, em cuja base repousa a noção de

que a forma está subordinada à função e a de que a organização interna da linguagem se dá em termos de funções que ela deve desempenhar na vida social.

4.2. A Gramática Sistêmico-Funcional

A Gramática Sistêmico-Funcional (doravante GSF) propõe-se a descrever a língua inglesa a partir da estreita relação entre forma e função, no sentido de que a forma assumida pelo sistema gramatical e as necessidades sociais e pessoais preenchidas pela língua são fortemente ligadas, como se pode depreender de Halliday (1970). Concebe, pois, que a língua toma a forma de uma série de estruturas sistêmicas que são representativas das escolhas semânticas associadas com algum tipo de elemento lexicogramatical (Halliday, 1967, p.37).

Trata-se, assim, de um modelo de base semântica, o que orienta, evidentemente, para uma perspectiva de língua como potencial de significado. Em Hallidaye Matthiessen (2004), é patente a noção de que os significados são produzidos via língua, sendo esta, portanto, um recurso para esse fim. Os falantes de uma língua, ao desejarem comunicar situações particulares por eles vivenciadas, operam escolhas atreladas tanto ao que querem comunicar quanto às situações em que se encontram.

Na GSF, um aspecto que interessa especialmente a este trabalho é o Sistema da Transitividade, que está relacionada à realização da Metafunção Ideacional, já que é o funcionamento desse sistema que demonstra as relações possíveis entre os elementos constitutivos e constituintes do mundo externo e interno do usuário da língua.

Para além do aspecto forma/função, é preciso se considerar ainda que as formas linguísticas e as funções, que lhes dão origem, estão necessariamente atreladas à esfera social em que a comunicação via linguagem verbal se realiza. Consequentemente, uma análise do objeto que o isole do ambiente no qual ele efetivamente é usado não pode contribuir para a compreensão de seu funcionamento.

Retomando a ideia de que na base da GSF subjaz a noção de que a forma está subordinada à função e a de que a organização interna da linguagem se dá em

termos de funções que ela deve desempenhar na vida social, Halliday (1979) aponta que são quatro as funções que a linguagem cumpre em uma cultura:

- a) interpretação do conjunto das experiências dos falantes;
- b) expressão de certas regularidades lógicas elementares;
- c) expressão da participação dos falantes/ouvintes na situação de discurso;
- d) capacidade de realizar as três funções anteriores concomitantemente, organizando-se como discurso relevante.

Dessa forma, Halliday (1994) evidencia, baseando-se nessas funções, as Metafunções que constituem a estrutura interna da língua. Trata-se da Metafunção Ideacional, da Metafunção Interpessoal e da Metafunção Textual. Cada uma dessas Metafunções possui um sistema que viabiliza a realização de seus significados. A Metafunção Ideacional é realizada pelo sistema da Transitividade e atua, justamente, na interpretação do conjunto das experiências dos falantes (subfunção experiencial) e na expressão de certas regularidades lógicas elementares (subfunção lógica); já a Metafunção Interpessoal (responsável pela expressão da participação dos falantes/ouvintes na situação de discurso) é realizada pelo sistema Modo e a Metafunção Textual (capacidade de realizar as três funções anteriores, concomitantemente, organizando-se como discurso relevante) é realizada pelo sistema de Tema.

Neste trabalho, será focalizada a Metafunção Ideacional, cuja instanciação se dá pelo Sistema da Transitividade, que envolve o usuário em um circuito que pressupõe a escolha de processos, participantes e circunstâncias. Essa Metafunção está relacionada à possibilidade que a língua oferece ao seu usuário de expressar verbalmente a imagem mental que apreende do mundo que o cerca e do mundo que se constrói internamente. Em outras palavras, permite que se dê sentido ao que se passa no exterior e no interior do indivíduo, codificando-o.

4.2.1. O Sistema de Transitividade

Da mesma forma que a linguagem é utilizada para interação entre as pessoas, é usada também para criar a experiênciade mundo ao redor dos

indivíduos, os acontecimentos que eles vivenciam, as qualidades atribuídas a si mesmos e a outrem. Elabora-se igualmente por meio da linguagem o chamado “mundo interno”, ou seja, os pensamentos, as crenças e os sentimentos.

Estar a par dessa perspectiva leva a que se focalize, primeiramente, o significado a ser criado, o conteúdo da mensagem, considerando-se especialmente o propósito para o qual o falante a expressou. Nesse conteúdo, podem-se antever as relações estabelecidas, constituintes da mensagem. Tais significados são expressos por meio de processos que “organizam”, “projetam” certos tipos de participantes — e não outros — que estão em certas circunstâncias, ou não, como apontam os exemplos abaixo¹:

Eu	estou oferecendo	aV.Sas	o mesmo prazo
Participante	Processo	Participante	Participante

CAREC3

(eu) Solicito	o reparo	em uma colméia do teto
Participante	Processo	Participante

CAREC4

Na perspectiva experiencial, os verbos são os itens linguísticos que realizam os processos na oração, configurando-se, assim, como o elemento central da mensagem. O sistema da Transitividade, pois, organiza o mundo da experiência em um conjunto de tipos de processos que, conforme já foi apontado, é constituído do processo em si, dos participantes e das circunstâncias. Importa assinalar que se fala aqui de categorias semânticas expressas por estruturas linguísticas.

Processos

É por meio da Metafunção Ideacional que se expressa, como já foi afirmado, o modo como concebemos o mundo. Essa Metafunção se associa tanto ao mundo

¹ Exemplos retirados do *corpus* referente às cartas de reclamação (CAREC3 e CAREC4)

externo (eventos, acontecimentos etc.) quanto ao nosso mundo interno (pensamento, crenças, sentimentos etc). O Sistema da Transitividade possibilita, por sua vez, a sua instanciação – que pode ser entendida como o contínuo entre um significado potencial e suas diferentes formas de expressão na língua – nas realizações linguísticas. Nesse Sistema, pressupõe-se a escolha de processos e de seus argumentos, participantes e circunstâncias.

Pode-se entender Transitividade, de acordo com Halliday e Matthiessen (2004), como um sistema que possibilita construir, por meio de escolhas linguísticas, o mundo interno e externo ao falante e tais escolhas implicarão possibilidades significativas diferentes a partir do momento em que o foco da representação esteja no processo em si ou nos participantes desse processo, o que promove efeitos de sentido diferenciados.

O modo como experienciamos o mundo que nos cerca e as ações e estados nele presentes é passível de ser traduzido em processos. Nesse sentido, há os denominados Materiais, os Mentais, os Relacionais, considerados os principais e os intermediários que são os Comportamentais, os Verbais e os Existenciais, conforme assinalam Halliday (1994, p. 107), Thompson (1996) e Halliday e Matthiessen (2004).

Em que consiste cada um desses processos? Os Materiais estão ligados à expressão das ações do mundo físico, exterior; já os Mentais àquelas que ocorrem no mundo interior; os Relacionais, por sua vez, dizem respeito a significados ligados à identificação, à classificação e à posse; os Comportamentais representam ações com caráter físico e mental; os Verbais são aqueles do dizer e, por fim, os Existenciais refletem o reconhecimento da existência de uma entidade, estando relacionados ao existir. Importa esclarecer que, atrelados a esses processos, estão os participantes e em alguns casos as circunstâncias. Outro aspecto que cabe considerar é o fato de não haver prioridade de um processo sobre o outro, já que estão, segundo Halliday (1994), interligados de maneira contínua dentro de um círculo, como se pode evidenciar na figura (Halliday, 1994) reproduzida a seguir:



Figura 1-Transitividade: os tipos de processos (Adaptado de Halliday, 1994:108)

A seguir, os processos serão mais detalhados em suas características, incluindo-se exemplos retirados do *corpus* desta pesquisa.

Processos Materiais

Os processos materiais expressam situações do âmbito do fazer, do acontecer e do criar, conforme aponta Halliday (1994), constituindo-se, dessa forma, como aqueles processos mais próximos dos usuários da língua, já que esse tipo de processo é o que diz respeito às ações e aos acontecimentos experienciados. Seus participantes são o Ator, a quem pode estar restrito o processo, como no exemplo 1 abaixo. Todavia, pode haver, igualmente, uma extensão desse processo ao participante designado Meta, como nos exemplos 2 e 3, que foram retirados das cartas de reclamação e do atendimento em central telefônica.²

1

<i>Essas câmeras</i>	<i>queimaram</i>	<i>entre agosto e setembro de 96</i>
Ator	Processo	Circunstância

CAREC7

² A maior parte dos exemplos foi retirada do *corpus*, mas alguns foram criados por mim.

2

<i>Coletamos</i>	<i>água</i>	<i>de um dos vazamentos</i>
Processo	Meta	Circunstância

CAREC8

3

<i>Encaminhamos</i>	<i>algumas reivindicações</i>
Processo	Meta

CAREC7

Além dos participantes mencionados, é possível ainda se considerarem outros dois, que são a Extensão — participante não exclusivo desse tipo de processo — e o Beneficiário. Uma das funções da Extensão, segundo Eggins (1994, p. 233), está relacionada à reafirmação ou à continuação do processo, e outra diz respeito à definição de seu escopo. Já o participante Beneficiário, como indicado no próprio nome que o designa, se beneficia do processo. Eggins (1994, p. 35) aponta que esse participante é como se fosse um cliente ou receptor, uma vez que ou algo é feito por ele ou para ele. Os exemplos 4 e 5 abaixo, respectivamente, ilustram esses participantes.

4

<i>A empresa</i>	<i>fez</i>	<i>a troca das pastilhas</i>
Ator	Processo	Extensão

CAREC4

5

<i>Estou oferecendo</i>	<i>a Vossa Senhoria</i>	<i>o mesmo prazo prometido</i>
Processo	Beneficiário	Meta

CAREC3

Não parece tarefa simples distinguir Meta de Extensão, como aponta Eggins (1994, p. 233). Aqui, buscou-se diferenciar esses participantes seguindo a noção de que, no primeiro, há a ideia de que se trata de participante modificado

pela ação, como no exemplo 2, em que o participante Meta “água”, ao ser coletado, é modificado de alguma forma.

Já em relação ao participante Extensão, considerou-se a ideia de reafirmação ou continuidade do processo no participante. Em 4, por exemplo, “fazer a troca” das pastilhas é “trocar”. É possível, no entanto, que tal correspondência que se tentou verificar a fim de diferenciar Meta de Extensão não encontre equivalência na totalidade dos casos que, porventura, se queira analisar, mas parece importante buscar a diferenciação, já que se trata, também segundo Egins, de funções distintas, logo, configuração semântica também diferenciada.

É importante ter em conta que toda oração com processo material tem um resultado final proveniente do desenrolar desse processo no tempo. Nesse sentido, verifica-se que podem ocorrer mudanças nas características de um dos participantes. Surge daí um contraste que pode ser observado nos processos materiais. De um lado, o processo se caracteriza como criativo e de outro como transformativo. No criativo, o Ator ou a Meta é trazido à existência enquanto o processo se desenrola. No transformativo, o Ator ou a Meta é transformado com o desenrolar do processo.

6

<i>O condomínio</i>	<i>expediu</i>	<i>a notificação</i>
Ator	Processo Criativo	Meta

CAREC26

7

<i>O electricista</i>	<i>conseguiu diminuir</i>	<i>tais defeitos</i>
Ator	Processo Transformativo	Meta

CAREC7

No exemplo 6, *a notificação*, que se configura como participante Meta, foi trazida à existência no desenrolar do processo. Já o exemplo 7 é processo do tipo transformativo, pois a Meta se transformou no desenrolar do processo.

É importante observar que todo processo implica a existência de um Ator, como assevera Thompson (1996, p.78), ainda que ele não seja mencionado. Logo, pode-se inferir a obrigatoriedade desse tipo de participante,

Processos Mentais

Esses processos expressam a apreciação humana do mundo, porque lidam com o modo como o sujeito elabora a experiência na sua consciência, no seu mundo interior. Por meio deles, criam-se as experiências do sentir, como a percepção (ver, ouvir, perceber), da cognição (pensar, lembrar, compreender, querer) e da afeição (gostar, amar, detestar, odiar). Sua análise permite a identificação de crenças, valores e desejos representados nas orações. Os participantes são o Experienciador e o Fenômeno.

8

<i>Eu</i>	<i>(não) vejo</i>	<i>outra solução</i>
Experienciador	Processo Mental	Fenômeno

CAREC11

Como se pode observar por meio do exemplo acima, o Experienciador é um ser consciente, representado por um grupo nominal; o Fenômeno, por sua vez, pode portar qualquer traço, sendo o elemento percebido ou sentido pelo Experienciador.

Halliday (1994) evidencia um conjunto de características que os processos mentais compartilham, tais como o participante Experienciador ser dotado de consciência, sendo, assim, humano ou representado como tal, o que aponta para uma restrição maior no tocante à classe semântica desse participante. O participante Fenômeno pode ser uma “coisa” ou um “fato”, ampliando-se, nesse sentido, a classe semântica das entidades que podem atuar como complementos nos processos mentais.

9

<i>Eu</i>	<i>(não) sei</i>	<i>sua função</i> (ATEND91)
<i>O Senhor</i>	<i>(não) entendeu</i>	<i>a minha pergunta</i> (ATEND1)
<i>Eu</i>	<i>(não) lembro</i>	<i>o total</i> (ATEND1)
<i>Eu</i>	<i>(não) tô ouvindo</i>	<i>t e</i> (ATEND5)
Experienciador	Processo Mental	Fenômeno

No quadro acima, observa-se que os Experienciadores (*Eu/O Senhor*) são conscientes. Alterando-se, por exemplo, a oração *O Senhor não entendeu a minha pergunta* para *A minha pergunta não entendeu o Senhor*, verifica-se a incongruência da construção, dada a restrição semântica do processo mental *m* *jogo* (*entender*) que acaba por delimitar a natureza do participante Experienciador.

O participante Fenômeno (*sua função, a minha pergunta, o total, te*) pode ser realizado, como indica o exemplo, por entidades de natureza mais diversa: pessoas, atos, objetos (abstratos e concretos), fatos.

Processos Relacionais

Esse tipo de processo indica uma relação entre duas entidades, por classificação ou identificação. São os chamados processos do “ser”. O sistema (inglês) de construção dos processos relacionais divide-se em:

- 1) Intensivo – quando uma qualidade é atribuída a uma entidade;
- 2) Circunstancial – quando uma circunstância (tempo, lugar, modo...) é atribuída a uma entidade;
- 3) Possessivo – quando há uma relação de posse.

Esses três tipos dividem-se em dois modos:

- a) Atributivo;
- b) Identificativo.

10

Tipo	Modo	
	Atributivo	Identificativo
Intensivo	<i>O gotejamento é constante</i> (CAREC2)	<i>A empresa é conceituada</i> <i>Conceituada é a empresa</i> (exemplo criado)
Circunstancial	<i>dois banheiros de um</i> <i>apartamento que estão com</i> <i>vazamento</i> (CAREC1)	<i>Amanhã é a vistoria</i> <i>A vistoria é amanhã</i> (exemplo criado)
Possessivo	<i>os trabalhos tem qualidade</i> <i>deficiente</i> (CAREC42)	<i>O vazamento é do</i> <i>apartamento 102</i> <i>O apartamento 102 tem</i> <i>vazamento</i> (exemplo criado)

Processos Comportamentais

Situam-se entre os processos materiais e os mentais, expressando, de acordo com Halliday e Matthiessen (2004, p. 248), comportamentos físicos e psicológicos, muitas vezes, concomitantemente. Do conjunto de processos, o comportamental é o que menos apresenta um conjunto de características particularizantes. Pelos traços que apresenta, aproxima-se, por vezes, das marcas dos processos materiais e, de outras, daquelas inerentes ao processo mental. O participante que “se comporta” é chamado Comportante e é um participante consciente, contendo, portanto, traço mais humano.

O tipo mais comum de experiência construída por processo comportamental dispõe apenas do participante Comportante e do processo em si. Opcionalmente, pode haver um participante que estende o processo, a Extensão.

11

<i>A cliente</i>	<i>chorou</i>	<i>como prejuízo</i>
Comportante	Processo Comportamental	Extensão

(exemplo criado)

Ainda que as fronteiras do processo comportamental sejam imprecisas, Halliday e Matthiessen (2004) identificam os seguintes verbos que estão a serviço dos Processos Comportamentais:

	Próximo ao mental	Processos de consciência representados em forma de comportamento	Olhar, assistir, ouvir, pensar, preocupar-se, sonhar
	Próximo ao verbal	Processos verbais em forma de comportamento	Conversar, falar, fofocar, discutir, murmurar, resmungar
	-	Processos fisiológicos manifestando estados de consciência	Chorar, rir, sorrir, suspirar, soluçar, rosnar, gemer
	-	Outros processos fisiológicos	Respirar, espirrar, tossir, bocejar, dormir
	Próximo ao material	Posturas corporais e passatempos	Cantar, dançar, deitar-se, sentar, levantar

Processos Verbais

Thompson (1996) assevera que entre os processos que estão nas fronteiras dos considerados centrais, ou seja, os materiais, os mentais e os relacionais, há os considerados menos centrais, que compartilham com aqueles algumas de suas características. Dentre eles, o processo verbal é, segundo o autor, o mais importante. É um processo que expressa o dizer, o comunicar, não requerendo apenas, como no processo mental, por exemplo, um participante consciente. Quando se trata de emprego metafórico do verbo (*A carta diz que o piso do*

banheiro rachou), várias entidades podem assumir o papel do participante Dizente, por isso pode ser considerado um processo simbólico.

Os participantes desses tipos de processos são:

- ✓ O Dizente: participante inerente que diz, que comunica ou aponta algo;
- ✓ O Receptor: participante opcional para quem o processo verbal se dirige;
- ✓ A Verbiagem: participante que corresponde (ou codifica) àquilo que é dito, podendo ser o conteúdo do que é dito ou o nome do que se é dito;
- ✓ Alvo: participante opcional, a entidade que é o alvo do que é dito.

12

<i>(Eu)Solicito</i>	<i>o reparo</i>	<i>em uma colméia do teto</i>
Processo	Verbiagem	Circunstância

CAREC4

Processos Existenciais

Entre os processos relacionais e os materiais, insere-se o processo existencial. Esse tipo de processo expressa a existência de uma entidade, sendo realizado, tipicamente, pelos verbos *haver* (*ter*, no mesmo sentido) e *existir*. Há contextos, todavia, em que outros verbos, tais como *emergir*, *surgir* e *ocorrer* podem ser concebidos como tal. Apenas um participante está presente nas orações que são engendradas com processo existencial: o Existente.

13

<i>tem</i>	<i>dias que o vazamento aumenta</i>
Processo Existencial	Existente

(exemplo criado)

14

<i>Haverá</i>	<i>assembleia geral</i>
Processo Existencial	Existente

CAREC6

Este capítulo procura não só situar o paradigma funcionalista no campo dos estudos da linguagem e a inserção nesse panorama da Linguística Sistêmico-Funcional, cujos pressupostos teóricos foram evidenciados, como também busca demonstrar algumas categorias-chave para a análise da língua a partir dessa perspectiva. Assim, conceitos relacionados às Metafunções, em particular a Metafunção Ideacional, ao Sistema da Transitividade, que expressa a construção do mundo interno e externo do indivíduo em termos de processos materiais, relacionais, verbais, mentais, comportamentais e existenciais, presentes na modelagem da oração, foram abordados como forma de dar suporte a algumas das discussões que surgirão posteriormente.

Na seção seguinte, para finalizar o capítulo, são apresentadas considerações acerca da Linguística de *Corpus* nos estudos da linguagem e sua utilização pela Linguística Sistêmico-Funcional em particular. Procura-se ressaltar, assim, a importância da abordagem de *corpus* nos estudos linguísticos que concebem a linguagem como um fato social, cujos padrões de ocorrência são passíveis de apreensão, indicando interpretações que ampliam o conhecimento acerca desse objeto.

4.3.

A Linguística de *Corpus* nos estudos sobre a linguagem

Conforme já se assinalou, o panorama dos estudos da linguagem, no século passado, se mostrou bastante diverso. Um breve olhar para esse cenário no qual o paradigma formalista e o paradigma funcionalista disputam território é, no mínimo, revelador de uma série de enfoques por meio dos quais se podem estudar a linguagem verbal. A esse contexto diversificado, associa-se a Linguística de *Corpus* (LC), área de pesquisa que vem acrescentar ao estudo da linguagem recursos singulares provenientes de um campo também inovador: a tecnologia computacional. Teuber (1996, p. vi) já apontava ser esta área “a face moderna da linguística empírica”. Desse modo, a associação entre informática e estudos da linguagem, embora possa parecer improvável ao senso comum, dada a tendência a entender o estudo das Letras como menos tecnológico, tem se expandido pelo menos desde os anos 60, de acordo com Sardinha (2004).

A LC, como mostram Thompson e Hunston (2006), “envolve a investigação de grandes quantidades de ocorrência de texto natural, as quais são armazenadas eletronicamente” (p.8). Já *corpus*, em Sardinha (2004, p. 18), é definido como:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.

Em Santos (2006, *apud* Santos 2008, p. 45), lê-se essa outra definição: “Um corpo é uma colecção classificada de objectos linguísticos para uso em Processamento de Linguagem Natural/Linguística Computacional/Linguística”

Tendo em vista as definições acima, pode-se afirmar que a Linguística de *Corpus* tem como escopo desde a compilação de *corpora* ao desenvolvimento de pesquisas empíricas. Estuda fenômenos como léxico, sintaxe, textura, por meio da utilização de ferramentas computacionais, aplicando várias metodologias que possibilitam a análise tanto quantitativa quanto qualitativa de grandes volumes de língua natural. Como o estudo de uma língua requer análise de textos orais ou escritos, lugar em que operam os mecanismos que caracterizam a linguagem verbal como tal, a possibilidade de coleta e organização com base em critérios específicos mostra-se bastante útil. Ao se organizarem tais dados em computador e torná-los disponíveis aos pesquisadores da linguagem, amplia-se a visão do objeto.

Cabe mencionar que, para a Linguística de *Corpus*, a linguagem é um fenômeno social. Ser um fenômeno social implica ser algo a que se tem acesso, já que se mostra à observação. Conseqüentemente, “porções de linguagem” digitalizadas e organizadas em *corpora* que são disponibilizados em bancos de dados e podem ser analisados por meio de programas computacionais possibilitam ao pesquisador a apreensão de padrões de ocorrência, objetivo precípuo da LC. A face qualitativa da pesquisa encontra, pois, na tecnologia computacional um poderoso aliado. A utilização de *parsers* (analisadores sintáticos), *concordancers* (concordanciadores), *tags* (etiquetadores), dentre outras ferramentas disponíveis como programas computacionais, auxilia o linguista em análises que envolvem um grande número de palavras. A LC tem se mostrado

aliada de peso em trabalhos na área da lexicografia, de tradução e de gêneros da mesma forma que se mostrou poderosa aliada nesta pesquisa, para a qual foi selecionado um *corpus* de aproximadamente 73.000 palavras (somado o total dos dois *corpora*), que foi analisado com o auxílio de um concordanciador.

Dessa forma, a Linguística de *Corpus*, assim como a Linguística Sistêmico-Funcional, considera o uso efetivo da língua em textos reais. Além disso, entende ser a dinamicidade do objeto língua passível de variação e de mudança. Nesse sentido, se enquadra na denominada teoria exofórica que, conforme menciona Hasan (1999), *apud* Oliveira (2009), se define por não isolar o objeto de estudo, relacionando-o, ao contrário, aos demais universos da experiência humana.

Historicamente, os *corpora* computadorizados têm sua forma inspirada no *Survey of English Usage* (SEU). Esse material, que abrangeu 1 milhão de palavras, foi organizado em Londres, no final da década de 50 (1959), tendo como responsáveis por sua compilação Randolph Quirk e sua equipe. Pode-se supor a dificuldade de se operar com essa quantidade de palavras manualmente, além do aspecto propriamente físico de armazenamento dos dados.

Esse quadro, todavia, sofreu significativa mudança com o advento da tecnologia computacional. Se o processamento manual de *corpora* era alvo de críticas, em razão da não confiabilidade na capacidade do homem de lidar com tamanha quantidade de material, a invenção do computador pôde promover alteração nesse quadro. O surgimento do *Brown University Standard Corpus of Present-day American English*, primeiro *corpus* eletrônico de linguagem escrita, alavancou o desenvolvimento da Linguística de *Corpus*. O *corpus* Brown, lançado em 1964, abrangeu 1 milhão de palavras e representou um marco para a LC.

À medida que a tecnologia computacional se desenvolveu, inclusive com a proliferação dos microcomputadores, tornou-se maior o alcance da LC. *Corporae* ferramentas de processamento passaram a ser mais conhecidos, integrando definitivamente o panorama dos estudos da linguagem. Importa assinalar, nesse contexto, que não apenas centros acadêmicos demonstram interesse em estudo de *corpus*, mas também empresas têm interesse na área, dada a sua aplicação comercial. Na Europa, utilizam-se bastante os seus recursos nos centros de pesquisa mais conceituados. Nos Estados Unidos, a despeito do seu forte potencial em desenvolvimento tecnológico computacional, a área é menos desenvolvida. No Brasil, a LC encontra-se em expansão, sendo mais utilizada nos

centros voltados ao Processamento de Linguagem Natural, à Lexicografia e à Linguística Computacional, como aponta Sardinha (2004, p.6). Em Sardinha (2008, p.23-33), é possível se verificar o crescimento de estudos envolvendo a LC. O autor realizou uma pesquisa, buscando evidências empíricas da expansão e do interesse em *Corpus* bem como da aplicação de seus pressupostos e constatou que entre 1999 e 2008 ocorreu

um crescimento qualitativo e quantitativo das pesquisas realizadas e a crescente formação em recursos humanos, uma vez que já se observa a realização de iniciações científicas, mestrados e doutorados que têm como tema central a Linguística de *Corpus* [...]

Convém ilustrar que há um grande número de *corpora* de língua inglesa, em face do maior desenvolvimento da LC em países europeus, conseqüentemente em maior oferta de recursos para compilação de *corpus* dessa língua. Nesse sentido, *corpus* como *Bank of English*, *British National Corpus*, *Brown* (já citado), *Birmingham corpus*, dentre outros, figuram na galeria dos inúmeros *corpora* dessa língua.

Ainda que em menor quantidade, tendo em vista a fase de desenvolvimento em que se encontra, há *corpora* de língua portuguesa representativos de extratos dessa língua. Sardinha (2004) apresenta um quadro-resumo dos *corpora* localizados em São Paulo, Minas Gerais Rio Grande do Sul, Rio de Janeiro, Paraná e mesmo exemplos como o Borba-Ramsey *Corpus of Brazilian Portuguese* ou o Modern Newspapers, ambos da Brigham Young University. Assim, *corpus* como NURC (UFRJ), VARSUL (UFSC, UFRGS, UFPR), Banco de Português (PUC-SP), dentre outros, compõem o cenário dos *corpora* de língua portuguesa, variando de 1 milhão a 152 milhões de palavras.

Oliveira & Dias (2009) apontam também iniciativas bem-sucedidas no tocante à compilação de *corpora* em português, em diferentes regiões do país. Dentre elas, o *corpus* do NILC – Núcleo Interinstitucional de Linguística Computacional (USP/UFScar/UNESP), o portal da Linguateca (2009), cuja composição inicial se limitava a uma coleção de textos do português europeu, mas ao qual vem sendo incorporado os textos do NILC e de outros projetos (COMET), o *Corpus Histórico do Português Tycho Brahe* (2008), que reúne a escrita de autores nascidos entre 1380 e 1845. A esse conjunto se agrega o CORPOBRAS PUC-Rio. Trata-se de um *corpus* representativo do português do Brasil, que se

encontra em fase de desenvolvimento. Esse *corpus* pretende fornecer dados e subsídios para uma análise multidimensional da variação entre gêneros discursivos. Como se trata do *corpus* no qual foram colhidos os dois gêneros textuais em estudo neste trabalho, maiores detalhes sobre o CORPOBRAS estarão presentes no capítulo relativo à metodologia da pesquisa (cf. capítulo 5).

Em Oliveira (2009), evidencia-se que algumas pesquisas no Brasil vêm se valendo de dados de uso real da língua, o que mostra a necessidade da compilação, organização e sistematicidade no trabalho com *corpus*. Oliveira enumera autores que se voltam para o *corpus* como subsídio para descrição do português em uso. Nessa esteira, encontram-se, por exemplo, Castilho (1999), Neves (1999), Azeredo (2008).

4.3.1.

A Linguística Sistêmico Funcional e A Linguística de *Corpus*

Um dos aspectos que confirma a possibilidade de complementaridade entre LSF e LC pode ser identificado em Thompson e Hunston (2006). A abordagem empirista e a visão da linguagem como sistema probabilístico, assim como para a LSF, são aspectos de fundamental relevância na LC. Conforme já foi apontado anteriormente, a oposição entre uma visão empirista e uma racionalista estabelece duas perspectivas de pesquisa bastante diferentes. Se em uma abordagem os dados que emergem de um *corpus* serão um caminho revelador de evidências linguísticas, em outra, de base racionalista, os dados necessários a análise estão na mente do pesquisador. A assunção de que a linguagem é um sistema de probabilidades implica o entendimento, por sua vez, de uma não equidade na frequência com que muitos traços linguísticos ocorrem. Essas frequências podem ser medidas em termos quantitativos pela LC e interpretada pela LSF em termos qualitativos.

Um pesquisador pretende descrever o uso do português do Brasil a fim de verificar a variação sincrônica de determinados gêneros escritos da língua portuguesa, com abordagem multidimensional. Outro estudioso investiga um gênero específico, a redação do vestibular, com abordagem discursiva. Um terceiro pesquisador quer analisar a variação intercultural em um *corpus* de

redações de alunos universitários, elaboradas em dois contextos culturais: Brasil e Estados Unidos.

Trata-se, evidentemente, de estudos que têm uma visão de linguagem enquanto fenômeno social requerendo, portanto, uma análise que considere os atos concretos de comunicação envolvidos na elaboração dos textos, o que faz deles produtos reais, possibilitando-se, destarte, analisar o significado onde ele é construído, ou seja, no discurso.

As três propostas de trabalho mencionadas foram efetivamente realizadas. Trata-se, respectivamente, dos trabalhos denominados *Variação de gêneros discursivos: a explicitação do contexto em um corpus do português escrito* (LANZIOTTI, 2002), *A Redação do vestibular como gênero: configuração textual e processo social* (CALDEIRA, 2006) — dissertações de mestrado e da tese de doutorado *Variação intercultural na escrita: contrastes multidimensionais em inglês e português* (OLIVEIRA, 1997).

Lanziotti trabalhou com um *corpus* de 176 textos dos gêneros: e-mail, carta pessoal, carta profissional, redação de aluno, artigo científico, editorial, notícia, circular, discurso político, romance, crônica, com aproximadamente 76.000 palavras. Os textos fazem parte do CORPOBRAS PUC-Rio. Caldeira, por sua vez, trabalhou com 30.000 palavras em um *corpus* composto por 135 redações de quatro instituições, compiladas entre 2004 e 2005. Já Oliveira (1997) trabalhou com 270 redações de alunos universitários, do Brasil e dos Estados Unidos, divididas em três grupos: inglês (L1), português (L1) e inglês como língua estrangeira (L2).

Os trabalhos utilizaram concomitantemente os pressupostos teórico-metodológicos da Linguística Sistêmico-Funcional e a abordagem da Linguística de *Corpus*, mostrando tratar-se de união possível e, por vezes, necessária. Não tem sido raro encontrar opiniões acerca da complementaridade de ambas as abordagens.

Em Thompson & Hunston (2006), argumenta-se que a Linguística Sistêmico Funcional entende a totalidade da língua em termos de potencial de significado da linguagem. A Linguística de *Corpus* substitui esse conceito pelo de ocorrência, ou seja, foca-se menos no que pode ocorrer e mais no que está ocorrendo, perceptível pelas porções de linguagem que se têm agrupadas a partir de critérios pré-estabelecidos, sendo, portanto, representativas da linguagem como

um todo. Todavia, o que está ocorrendo corrobora a noção de potencial de significado proposta pela LSF.

No que diz respeito à necessidade da união das duas abordagens, é importante ressaltar que, quando se lida com um número considerável de palavras, o uso de ferramentas computacionais (listas de frequência, por exemplo) que possam ser mais precisas que o homem na contabilização de dados contribui para maior precisão na sua quantificação e confirmação de resultados.

Para além do aspecto meramente matemático em questão, estudos de *corpus* apontam evidências empíricas que auxiliam o linguista na produção de novas informações teóricas ou aplicadas com base no *corpus*, já que na interpretação dos dados há “que levar também em conta o co-texto e os aspectos socioculturais que estão ligados aos textos” (OLIVEIRA, 2009, p. 50).