

## 5 Metodologia

A fim de descrever possíveis diferenças quantitativas e qualitativas relativas ao funcionamento do Sistema de Transitividade nas duas modalidades, este trabalho compara esta unidade semântica na oralidade e na escrita. Parte-se do pressuposto de que não há total equidade na distribuição de recursos lexicogramaticais que os indivíduos usam para dar forma, via linguagem verbal, à sua experiência interna e externa de mundo. Supõe-se que os processos presentes nos *corpora* analisados, usados para expressar tanto eventos, elementos, acontecimentos quanto crenças, pensamentos e sentimentos são diversificados qualitativa e quantitativamente, a depender da modalidade de língua, o que pode levar a efeitos de sentido diversos.

Com apoio da abordagem baseada em *corpus*, buscaram-se, por meio da observação e análise dos dados, interpretações que pudessem não apenas contribuir para ratificar peculiaridades da fala e da escrita, mas também para ampliar a descrição lexicogramatical desse sistema, cujos estudos, em língua portuguesa, ainda são escassos, necessitando-se, assim, de mais pesquisas que se direcionem para tal questão. Conforme aponta Lima-Lopes (2008), pouco se sabe sobre as particularidades dos processos e sobre a forma como os seus participantes e as circunstâncias se relacionam.

Dessa forma, a visão sociossemiótica desse sistema, por envolver questões de ordem não apenas linguística, mas também social, e suas relações com outras semioses, parece prover o seu estudo no âmbito acadêmico de maior utilidade, por estarem implicadas questões relacionadas ao uso da língua enquanto meio de construção da realidade. Tal perspectiva busca respostas para questões que envolvem configurações semânticas diversificadas, oriundas tanto do contexto quanto provenientes também das relações estabelecidas entre os interlocutores e deles com o tema sobre o qual discorrem.

## 5.1. O *corpus*

Os textos selecionados para a análise integram o CORPOBRAS PUC-Rio, *corpus* representativo do português brasileiro (Oliveira, 2009). O projeto é coordenado pela Professora Doutora Lúcia Pacheco de Oliveira e se desenvolveu com o financiamento do CNPq e FAPERJ e a partir do trabalho de coleta de dados e da contribuição de alunos e professores do Departamento de Letras da PUC-Rio. Segundo Oliveira (2009), sua montagem teve início em 2002, surgindo a partir da constatação de que no Brasil havia carência de *corpora* cujas dimensões fossem abrangentes, configurando-se, portanto, como gerais e representativos da língua usada no país. Tem sido evidenciado na literatura sobre o assunto o crescimento dos *corpora* desde a confecção, há 40 anos, do *Brown Corpus*, da Universidade de Brown, nos Estados Unidos. O advento da informática, no entanto, fez com que a compilação de *corpus* se desenvolvesse ainda mais em função do uso dos computadores não apenas para armazenar informações, mas também para a criação de programas que operassem com grande quantidade de dados linguísticos.

Na fase atual de organização e sistematização dos dados que integram o CORPOBRAS, podem ser contabilizados aproximadamente 1.400 textos e 1.200.000 palavras. *Corpora* dessa proporção, de acordo com Sardinha (2004), são considerados médio-grandes, já que há padrões de extensão a serem seguidos e a quantidade de textos e de palavras acaba por classificar um *corpus*, de acordo com o autor, em pequeno, pequeno-médio, médio, médio-grande e grande. O pequeno é aquele com menos de 80 mil palavras. O pequeno-médio é composto de 80 a 250 mil. Já o médio contém 250 mil a 1 milhão de palavras. O médio-grande, por sua vez, tem de 1 milhão a 10 milhões de palavras e, por fim, o grande é constituído de 10 milhões de palavras ou mais.

Os gêneros discursivos que compõem o CORPOBRAS são, atualmente, 28 (vinte e oito), distribuídos em 21 (vinte e um) gêneros de modalidade escrita, 05 (cinco) da modalidade oral e 02 (dois) gêneros do discurso escrito para ser falado. Dessa forma, artigos científicos, cartas ao editor, cartas de reclamação, cartas de recomendação, cartas pessoais, cartas profissionais, cartas profissionais acadêmicas, circulares, contos, crônicas, dissertações, editoriais, *e-mails*

acadêmicos, *e-mails* pessoais, notícias de jornal, redações de alunos do Ensino Médio, redações de alunos do Ensino Superior, memoriais, redações de vestibular, romances, teses, conversas cariocas, conversas de criança, entrevistas acadêmicas, grupos de enfoque, atendimento ao cliente, discursos políticos e roteiros cinematográficos são os gêneros que compõem, até o presente momento, o *corpus*.

Para este estudo, foram selecionados, desse conjunto, tendo em vista a sua finalidade, um gênero textual escrito e um o gênero textual oral, a saber, a carta de reclamação e o atendimento em central telefônica, formando-se, assim, um sub-*corpus* do CORPOBRAS. Dessa forma, foram analisadas 126 cartas, com 21.800 palavras e 100 atendimentos ao cliente, com 51. 273 palavras. Trata-se de, portanto, de *corpora* autênticos e produzidos espontaneamente, que correspondem à língua portuguesa em uso.

O critério de escolha dos dois gêneros obedeceu ao princípio de se buscar correspondência na oralidade e na escrita a partir de propósitos comunicativos semelhantes, ou seja, fazer algum tipo de solicitação. Além disso, ambos os gêneros selecionados pressupõem relação formal entre empresas e clientes bem como demandam solicitação e resposta. Nesse sentido, o Campo e as Relações são de natureza similar, diferenciando-os, porém, o Modo (cf. capítulo 3) e o meio como fazem uso da linguagem verbal.

As cartas de reclamação foram enviadas por clientes insatisfeitos, em virtude de constatarem nos imóveis adquiridos problemas de diversas ordens (rachaduras, vazamentos, infiltrações, entre outros), a uma grande empresa do ramo da construção civil. A empresa constrói prédios residenciais cujos apartamentos são adquiridos por membros da classe média alta. Os textos, segundo Cerbino (2000), são acolhidos pelo setor de atendimento ao cliente. Lá, um profissional específico (gerente de apoio técnico) e sua equipe de funcionários entram em contato com o conteúdo do material para estabelecerem as diretrizes de atendimento às solicitações.

O gênero atendimento em central telefônica, por sua vez, corresponde a textos transcritos de conversas entre atendentes de uma companhia de plano de saúde e clientes dessa operadora, que entram em contato telefônico com diversas finalidades: solicitar informação acerca de médicos credenciados em sua cidade, sobre juros cobrados em dado mês, sobre datas de pagamentos; podem também

requerer senha para atendimento, alteração cadastral, reenvio de carteira, entre outros.

Na análise, para identificação dos textos, foi mantido, em parte, o padrão de codificação criado pelo CORPOBRAS, como, por exemplo, ATEND1PORT1, em que o gênero atendimento ao cliente é identificado (ATEND), o número dado ao texto (1) e a língua em que o texto foi produzido (PORT1). No caso das cartas de reclamação, o CORPOBRAS adotou a seguinte anotação: CAREC1PORT1, em que o gênero carta de reclamação é identificado CAREC, 1 corresponde ao número dado ao texto e PORT1 evidencia a língua. Foi retirada, entretanto, a identificação da língua PORT, já que não se analisou nesta pesquisa outra língua.

## 5.2.

### **Procedimentos de descrição**

Pressupõe-se que descrever ações, percepções e relações humanas que estão em jogo em uma atividade comunicativa oral ou escrita, bem como apreender padrões linguísticos que se configuram nessas atividades sociais não se revela tarefa simples. Uma vez que os processos organizam em torno de si, possibilitando sua expressão, participantes e circunstâncias, acabam sendo determinantes desses outros papéis que itens lexicais assumem no discurso, por isso, nesta pesquisa, concentrou-se maior atenção em torno deles. Os processos podem ser considerados como o elemento central do sistema semântico da Transitividade, sendo os participantes e as circunstâncias, os elementos que os circundam ou complementam. Da mesma forma, nesta pesquisa, escolhemos os processos como nucleares para a análise dos dois gêneros enfocados na Dissertação, e incluímos os outros elementos – participantes e circunstâncias – apenas quando necessário para esclarecer ou complementar a análise dos processos.

Para se chegar à identificação e descrição dos processos propriamente ditos, dada a quantidade de textos e conseqüentemente de palavras, optou-se, primeiramente, pela identificação da configuração contextual dos gêneros a fim de se levantarem as suas etapas bem como os elementos obrigatórios e os opcionais (cf. capítulo 3), buscando-se, assim, observar o que ocorria nesses elementos em termos de funcionamento do Sistema da Transitividade.

Esse procedimento foi realizado manualmente, por meio da leitura dos textos. As sucessivas leituras acabaram por identificar etapas que se repetiam com margem a algumas variações. Para ter a quantificação precisa das ocorrências, elaborou-se um quadro no Excel e classificaram-se as etapas das cartas de Inicial, Intermediária e Final e nos atendimentos de Abertura, Desdobramento e Desfecho, partes que equivalem à tradicional taxonomia *introdução, desenvolvimento e conclusão*.

Observou-se que em cada etapa identificada ocorriam (ou não) elementos que foram categorizados como obrigatórios ou opcionais (cf. capítulo 6, seção. 6.2.1).

Percentuais de ocorrência foram obtidos com base na identificação da ocorrência de cada tipo de elemento e posterior contabilização dos totais, sendo considerados obrigatórios aqueles que mostravam percentuais acima de 95%.

O levantamento da Configuração Contextual (cf. capítulo 6) foi, portanto, fundamental para que se pudesse dar encaminhamento à análise com base nessas ocorrências já que elas indicaram um caminho possível para a análise dos processos, principal objetivo da pesquisa.

Foi necessário, após essa etapa, gerar as listas de frequência de palavras no MonoConc Pro (Barlow, 1990) — conforme figura 2 abaixo — para se identificarem nelas os verbos e se chegar aos processos presentes, de um modo geral, tanto nas cartas quanto nos atendimentos. Essas listas, porém, têm duas limitações: a primeira diz respeito a não identificação da classe da palavra, o que leva o analista a precisar refinar a listagem manualmente, isolando o item que lhe interessa, no caso deste trabalho, os verbos.

Count	Pct	Word
1105	5,0642%	de
701	3,2126%	a
619	2,8368%	que
518	2,3740%	o
478	2,1907%	do
458	2,0990%	e
342	1,5674%	da
247	1,1320%	para
242	1,1091%	no
222	1,0174%	em
194	0,8891%	com
183	0,8387%	na
182	0,8341%	nro
171	0,7837%	por
152	0,6966%	os
135	0,6187%	c
135	0,6187%	uma
125	0,5729%	as
124	0,5683%	-
120	0,5500%	empresa
120	0,5500%	p
120	0,5500%	v
107	0,4904%	apartamento
103	0,4720%	um
97	0,4445%	atenciosamente
94	0,4309%	estô
93	0,4262%	ao
93	0,4262%	jê
92	0,4216%	das
91	0,4170%	se
89	0,4079%	a
87	0,3987%	como
87	0,3987%	mais
84	0,3850%	sr
82	0,3758%	dos

Figura 2- Reprodução de janela com a ferramenta Frequency List do MonoConc Pro

A segunda, consequência da primeira, é que, uma vez isolados os verbos, torna-se necessário reordená-los em ordem alfabética, já que o programa não os identifica na sua forma infinitiva. Tal identificação é fundamental para que se tenha o quantitativo exato que diz respeito aos verbos presentes nos textos. Por fim, foi necessário, de posse dos quantitativos, analisar os verbos nas linhas de concordância para se verificar o sentido assumido pela forma verbal no co-texto e poder, assim, categorizar o tipo de processo.

Realizados esses procedimentos, pôde-se voltar aos contextos mais amplos dos textos e partir para a análise qualitativa do Sistema de Transitividade, identificando processos mais expressivos, participantes e circunstâncias dos significados ideacionais criados a partir de escolhas lexicogramaticais.

Conforme já se assinalou anteriormente, este estudo combinou a abordagem baseada em *corpus* aos pressupostos da Gramática Sistemática Funcional (GSF). Acredita-se na necessidade de análise de dados empíricos para fundamentar as posteriores interpretações discursivas daquilo que é feito com a linguagem verbal. Em outras palavras, assume-se com Halliday (1994) que, para fazer análise do discurso, é preciso fazer descrição gramatical.

Tendo em vista a grande quantidade de palavras constituintes do *corpus* analisado, as ferramentas computacionais, nesse contexto, foram poderosas aliadas do pesquisador dada sua precisão, por exemplo, na contagem de palavras, na sua ordenação alfabética, na criação de listas de concordância:

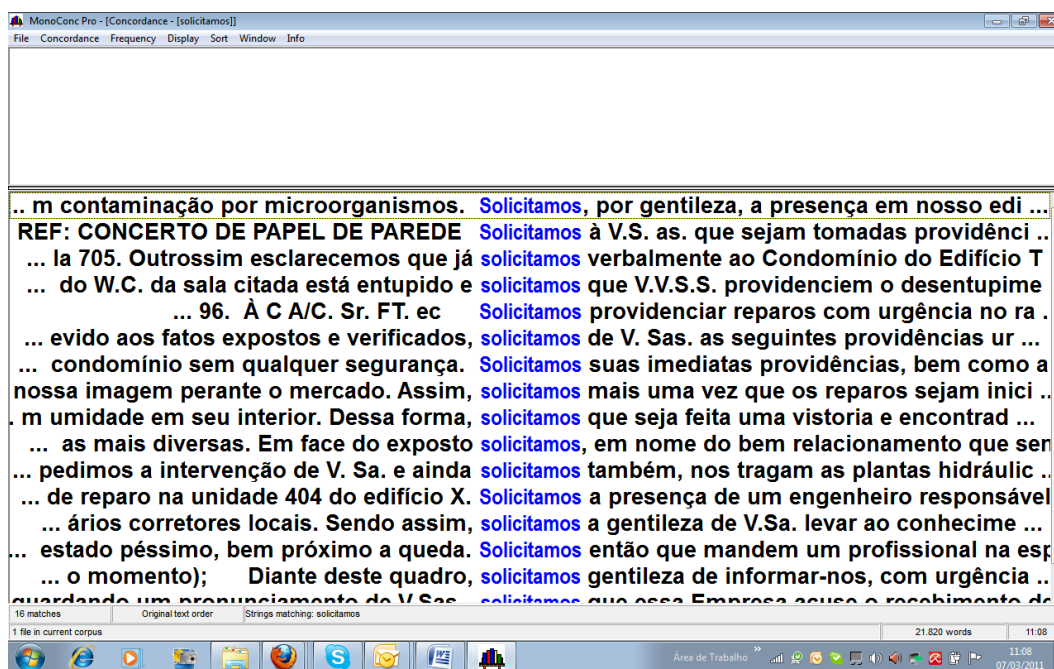


Figura 3 - Reprodução de janela com linhas de concordância geradas pela ferramenta *Concordance*

Trata-se de ferramenta por meio da qual se busca a palavra no co-texto em que o item ocorre, o que gera uma lista com todos os co-textos nos quais o termo, destacado dos demais, se localiza. Essas ferramentas foram também fundamentais no que se refere às informações percentuais e aos gráficos de distribuição de palavras pelos números de linhas, que possibilitaram, entre outros aspectos, importantes interpretações e posteriores categorizações:

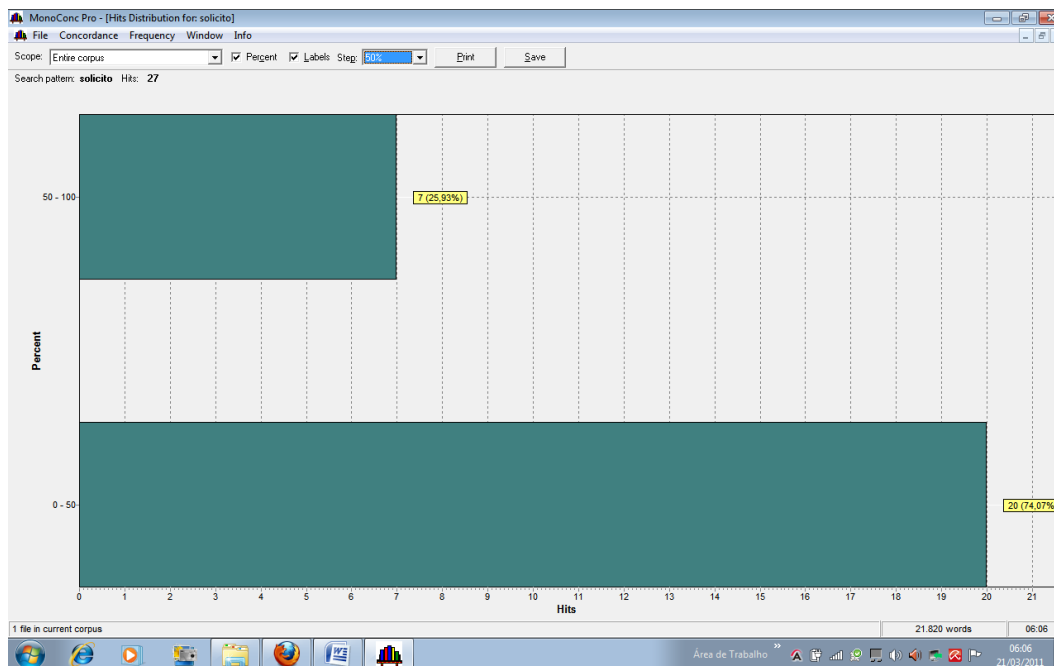


Figura 4- Reprodução da janela com um gráfico de distribuição gerado com o Processo Material “solicito”

Sardinha (2004) aponta ainda outra vantagem do uso do computador na análise linguística, quando assegura que fatos novos podem ser descobertos com a análise de grande quantidade de dados e ainda, por meio da observação e estudo desses fatos, a refutação de opiniões e crenças estabelecidas. Nesse contexto, o programa MonoCon Pro foi utilizado para auxiliar na descrição da Transitividade dos *corpora* em questão.

A ferramenta *Frequency*, por exemplo, auxiliou na identificação dos verbos e respectiva quantidade e o *Concordance*, na possibilidade de agrupar, por meio de listagens das concordâncias em que a palavra de busca ou nóculo é destacada e se pode verificar todo o entorno, ou seja, o co-texto de sua realização. Como pontua Halliday (1994, p. 176),

O processo é o elemento mais central na configuração. Os participantes estão próximos ao centro; eles estão diretamente envolvidos no processo, fazendo com que ele ocorra ou sendo, de alguma forma, afetado por ele.

No caso específico da análise dos participantes que “gravitam” em torno dos verbos, as linhas de concordância auxiliam em sua identificação. Além disso, pelo fato de os processos serem realizados via verbos e corresponderem a um aspecto semântico da organização do sistema, a análise do co-texto possibilitada pela ferramenta orientou para a apreensão do sentido assumido, logo para o tipo de processo em jogo. As listas de frequência, como o próprio nome assinala,



permitiram ao analista verificar a frequência com que cada termo apareceu no *corpus*. A partir delas, foram identificados todos os verbos, que, por sua vez, precisaram ser digitados separadamente em planilhas do Excel, para se contabilizar o total.

O CORPOBRAS já disponibiliza uma versão dos arquivos que contêm os textos que compõem o *corpus* em txt., formato necessário para que possam ser lidos pelo MonoConc, gerando, assim, as listas de palavras e as de frequência em ordem alfabética ou em ordem numérica de quantidade de ocorrências, conforme solicitação do analista.

No caso dos verbos, pelo fato de, nos textos, aparecerem flexionados e mesmo em locuções verbais, a identificação da forma não flexionada por meio do isolamento do lexema pareceu favorecer a compreensão do processo que estava em jogo. O procedimento de normatização de frequências foi aplicado, já que a quantidade de palavras não era a mesma nas duas modalidades analisadas. Como apontam Biber & Reppen (1998, p. 263):

A “Normatização” é uma forma de ajustar contagens de frequência bruta de textos de tamanhos diferentes de forma que possam ser comparados com precisão. O número total de palavras em cada texto deve ser levado em consideração quando se normatiza contagens de frequência. Especificamente, a contagem de frequência bruta deve ser dividida pelo número de palavras no texto e em seguida multiplicada pela base escolhida para normatização.

Como o objetivo da pesquisa foi comparar o funcionamento do Sistema de Transitividade na oralidade e na escrita, a análise das escolhas processuais feitas em uma e outra modalidade e a sua comparação com outras disponíveis foi um caminho para a apreensão das motivações (interpretação qualitativa) dos redatores das cartas ou usuários do sistema de atendimento por telefone da empresa X-Saúde para a produção de seus textos.