

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE COMUNICAÇÃO E EXPRESSÃO  
CURSO DE PÓS-GRADUAÇÃO EM LINGÜÍSTICA**

**REJANE CROHARÉ DANIA**

**A CONSTRUÇÃO DO TEXTO ARGUMENTATIVO: UMA  
PESQUISA BASEADA EM *CORPUS***

Dissertação apresentada ao Programa de Pós-graduação  
em Lingüística da Universidade Federal de Santa  
Catarina, como requisito parcial para a obtenção do  
título de Mestre em Lingüística

Orientador: Prof. Dr. Marco Antonio Rocha

Florianópolis, 2002

**A CONSTRUÇÃO DO TEXTO ARGUMENTATIVO : UMA PESQUISA  
BASEADA EM CORPUS**

REJANE CROHARÉ DANIA

BANCA EXAMINADORA:

---

Prof. Dr. Marco Antonio Rocha

Orientador

---

Prof. Dr. A. P. B. Sardinha

---

Profa. Dra. Nilcéa Lemos Pelandré

---

Profa. Dra. Rosângela Hammes Rodrigues

(suplente)

## **AGRADECIMENTOS**

Ao CNPQ , órgão financiador dessa pesquisa;

Aos alunos envolvidos nesse processo de construção;

Aos professores e colegas do curso de Pós-Graduação em Linguística;

Ao Prof. Dr. Marco Antonio Rocha, meu orientador, pelo apoio, paciência, disponibilidade e orientação que tornou possível a realização desse projeto ;

À Julia e Felipe, meus filhos, pela compreensão;

Ao Antonio e Angela pelo apoio durante o percurso.

<b>SUMÁRIO.....</b>	<b>004</b>
<b>RESUMO.....</b>	<b>008</b>
<b>ABSTRACT.....</b>	<b>009</b>
<b>CAPÍTULO 1- INTRODUÇÃO.....</b>	<b>010</b>
1.1 Descrevendo o cenário.....	010
1.2 Metodologia de Corpus e o uso de corpora em sala de aula.....	013
1.2.1 Definição de corpus.....	013
1.2.2 Uso de corpora em sala de aula.....	013
1.3 A pesquisa.....	015
1.4 Descrevendo a organização retórica do texto.....	015
<b>CAPÍTULO 2 - REFERENCIAL TEÓRICO.....</b>	<b>017</b>
2.1. Linguística Aplicada (LA) campos de atuação.....	017
2.1.1 Questões de pesquisa .....	019
2.1.2 Relevância da pesquisa no contexto de LA.....	019
2.2 Linguística de corpus.....	020
2.2.1 Corpus, Corpora – definição.....	020
2.2.2 Corpus - antes e depois do computador.....	021
2.2.3 Linguística de Corpus – definição e características.....	023
2.2.3.1 Desempenho lingüístico, ao invés de competência.....	023
2.2.3.2 Descrição lingüística, ao invés de universais lingüísticos.....	024

2.2.3.3 Foco tanto na quantidade quanto na quanto nos modelos qualitativos de linguagem.....	025
2.2.3.4 Foco na pesquisa empírica, ao invés de uma visão mais racionalista de pesquisa científica .....	027
2.2.4 O paradigma de pesquisa em Lingüística de Corpus Computadorizado (LCC).....	027
2.2.4.1 O paradigma da concordância.....	029
2.2.4.2 O paradigma do modelo log-linear.....	029
2.2.4.3 O paradigma dos modelos probabilísticos da linguagem usando os modelos ocultos de Markov.....	030
2.2.5. Concordâncias – antes e depois do computador.....	031
2.2.5.1 A concordância na sala de aula.....	034
2.2.5.2 Onde conseguir textos.....	038
2.2.5.3 O concordanceador e o estudo de caso realizado.....	040
2.3 Lingüística computacional .....	041
2.3.1 Lingüística Computacional, Lingüística de corpus - uma ligação.....	044
2.3.2 Inserção da metodologia de corpus na sala de aula de língua materna e de língua estrangeira - campo de atuação da lingüística aplicada.....	044
2.4 Abordagem Histórico-cultural de lingugem.....	047
2.5. Argumentação.....	052
2.5.1 A Argumentação: leis internas e operadores argumentativos.....	053

2.5.2 O Operador Argumentativo <i>mas</i> .....	055
<b>CAPÍTULO 3 – METODOLOGIA.....</b>	<b>061</b>
3.1 Concepção da pesquisa.....	061
3.2 Objetivos da pesquisa e hipóteses de trabalho.....	061
3.2.1 Objetivos Gerais.....	062
3.2.2 Objetivos Específicos.....	062
3.2.3 Hipóteses de Trabalho.....	062
3.3 O contexto da pesquisa (sem o corpus).....	062
3.3.1 Os Instrumentos Utilizados.....	063
3.3.2 As Instruções.....	064
3.3.3 Metodologia Pedagógica.....	066
3.3.4 Resultados.....	066
3.3.5 Avaliação.....	067
3.4 A Pesquisa com corpus.....	068
3.4.1. Os Sujeitos da Pesquisa.....	068
3.4.2 A Coleta de Dados.....	069
3.4.3 Corpus na pesquisa .....	069
<b>CAPÍTULO 4 – A PESQUISA.....</b>	<b>071</b>
4.1 Exploração do Corpus.....	073
4.2 Frequência e colocações do “ <i>mas</i> “.....	074
4.3 Análise dos resultados.....	078

4.4 Discussão dos resultados.....080

**CONCLUSÃO.....109**

**REFERÊNCIAS.....113**

## **ANEXOS**

ANEXO 1 - *“Aqui se vive Melhor”*

ANEXO 2 - *“Paisagem e Espaço”*

ANEXO 3 - *“Mea Culpa”*

ANEXO 4 - Versão digitalizada da produções textuais feitas em 1999

ANEXO 5 - Concord para a versão digitalizada produzida em 1999

ANEXO 6 - Versão digitalizada da produções textuais realizadas em 2001

ANEXO 7 - Concord para a versão digitalizada produzida em 2001

## RESUMO

Esta pesquisa objetiva inserir a metodologia da Lingüística de Corpus e a manipulação de corpora em atividades cotidianas de sala de aula. Investiga-se o uso de operadores argumentativos e a metodologia da Lingüística de Corpus foi aplicada. Participaram da pesquisa alunos do terceiro ano do ensino médio de uma escola particular de Florianópolis, S.C. O programa WordSmith foi aplicado e usou-se a manipulação de corpora para verificar a intencionalidade no uso dos operadores argumentativos, focando o **mas**. Como resultado final, as produções textuais foram analisadas, segundo a metodologia da Lingüística de Corpus, e comparadas àquelas produzidas antes do estudo com o corpus. O resultado indica que o caminho da manipulação de corpora em computadores é uma alternativa não apenas viável, mas promissora em contexto de ensino.

## ABSTRACT

This study focuses on the insertion of Corpus Linguistics and corpora manipulation in daily activities in the classroom. The corpus-based study aims at the investigation of the Portuguese argumentative operator **but** and the Corpus Linguistics methodology was applied. Students of the third grade of a private school in Florianópolis, S.C. participated in the research. The WordSmith program and the corpora manipulation were used to verify the intentional use of the Portuguese argumentative operator **but**. Finally, texts were produced by the students, analysed according to the Corpus Linguistics methodology, and compared to those produced before the study. The result indicates that the manipulation of machine readable-corpora is not only a promising alternative for teaching, but a practicable one.

## CAPÍTULO 1 - INTRODUÇÃO

### 1.1. Descrevendo o cenário do estudo de caso

Diferentes estudos relacionados a distintos eventos comunicativos têm sido objeto de pesquisa na área de Lingüística. O raciocínio por trás desta pesquisa pode ser resumido da seguinte forma: primeiro a atividade humana é perpassada pelo uso da linguagem; segundo o objetivo maior da educação é preparar os aprendizes para a inserção na sociedade letrada, integrando-os no mundo do trabalho e procurando desenvolver progresso individual; terceiro a Lingüística Aplicada (LA, doravante) é articuladora de múltiplos domínios do saber que têm preocupação com a linguagem.

A linha de pesquisa da LA na qual o presente trabalho se insere investiga a metodologia da Lingüística de Corpus e o uso de corpora em sala de aula como formas de propor novos encaminhamentos pedagógicos, com base em recursos tecnológicos. A importância de incorporar estas tecnologias de maneira desafiadora e criativa à rotina de sala de aula se realiza pelo viés do trabalho com operadores argumentativos, focando o **mas**, e a construção de uma produção textual como produto final das atividades. A escola pesquisada tem como suporte teórico a abordagem histórico-cultural e em Vygotsky (1978) um norteador de sua prática pedagógica. A teoria da Zona de Desenvolvimento Proximal (ZDP) hipotetizada pelo teórico prevê que a aquisição do conhecimento do indivíduo possa ser facilitada através do par mais competente. Através dessa interação, insere-se a aplicação da metodologia da Lingüística de Corpus, conforme maior detalhamento do item 1.2. deste capítulo.

Os princípios sócio-interacionistas defendem que as funções psicológicas superiores se originam no contexto social, isto é, através da interação, antes de serem transferidas para o próprio indivíduo. Em outras palavras, a Zona de Desenvolvimento Proximal (ZDP) constitui um potencial para a aprendizagem que é criado pelo viés da interação social e abrange todas as atividades e funções que o aprendiz consegue desempenhar, desde que seja orientado por alguém mais competente. Durante a pesquisa, observou-se que os

aprendizes mais capazes, em termos de atenção, concentração e realização da tarefa, puderam prestar assistência ao par menos competente, isto é, auxiliando a simplificar a tarefa, parafraseando o que foi solicitado, reduzindo a pressão durante a execução da tarefa, procurando solucionar problemas, etc. Uma pequena resenha sobre a teoria proposta por Vygotsky faz parte do capítulo 2 e de como a ZDP se manifestou, durante a pesquisa, através do uso de computadores está detalhada no capítulo 3.

A construção do texto argumentativo deve ser tomada aqui como processo de produção textual. O tema remete ao trabalho de O. Ducrot (1981) e pesquisa os elementos lingüísticos que têm como função indicar a força dos argumentos dentro dos enunciados e o sentido para o qual apontam. A pesquisa buscou encontrar, através da metodologia da Lingüística de Corpus, o suporte necessário para promover a leitura e manipulação do corpus digitalizado de textos escritos do Núcleo Interinstitucional de Lingüística computacional (NILC) de onde foram selecionados textos da Folha de São Paulo (1995), investigando o uso do operador argumentativo **mas**, objeto dessa pesquisa. Procurava-se, então, o uso de computadores em sala de aula, através de uma tarefa desafiadora e nova tanto para o professor de língua portuguesa da turma como para os alunos.

Esse “novo” instrumento (computador) nos obriga a um deslocamento de nossas posições clássicas na relação professor-aluno e de ambos com a construção do conhecimento. É interessante observar *como* as relações devem se realizar, o que um deve esperar do outro: o aprendiz deve ser capaz de ver o professor como um problematizador, provocador, mediador de descobertas e não um repassador de conteúdos. O professor, por outro lado, deve ser capaz de confiar na possibilidade de interação entre os aprendizes, bem como na busca por informação que leve a um determinado conhecimento partilhado. Portanto, a concepção de que professor é um detentor de conteúdos e o aluno é um ser dependente daquele deve ser modificada para que ambos sejam parceiros na construção do conhecimento. Essa pesquisa demonstrou que é possível reverter o quadro “antigo” dessas relações, conforme detalhes do capítulo 3 - Metodologia.

Para muitos o termo *metodologia* remete a forma de fazer, ao método. Ampliando essa perspectiva para o contexto de LA, podemos nos aproximar muito das investigações feitas em lingüística, isto é, dentro do contexto escolar, podemos aprender a observar os

fenômenos lingüísticos, experimentar possibilidades, cruzar informações, levantar hipóteses referentes ao uso de língua, extrair resultados dessa manipulação através de atividades na sala de aula. A distância do uso real de língua versus o tratamento de linguagem dado pelos livros didáticos cria, nos alunos, um desejo enorme de querer se manter distantes “dessa língua estrangeira”. Para transpor o ensino de língua com base em exemplos que aparecem em gramáticas, livros antigos, escritores do século passado, propõem-se, através desse trabalho um novo olhar sobre a linguagem.

Esse novo olhar, pelo viés da Lingüística de Corpus, impõe o uso de uma nova tecnologia, um professor habilitado para esse fim e uma pesquisa a ser realizada. No caso dos operadores argumentativos, já mencionados acima, os estudos de O. Ducrot ( 1972, 1980, 1987, 1989) remetem ao valor argumentativo de certos morfemas e certas estruturas lingüísticas. Essas considerações teóricas foram combinadas a uma abordagem a partir de corpus para levar os alunos a manipular a linguagem em situações reais de uso , isto é , como um jornal de circulação nacional apresenta o objeto em estudo; ao invés de o professor inventar uma série de sentenças tentando levar os aprendizes a “pensar” sobre o uso de linguagem. Maiores detalhes estão disponíveis no item 1.2.2. O estudo realizado procura, então, demonstrar a importância da manipulação de corpus e os resultados de aprendizagem advindos desta experiência.

O uso de novas tecnologias, a saber, programas de lingüística computacional, Lingüística de Corpus, uso de corpora computadorizados no ensino de língua portuguesa, dicionários informatizados e letramento eletrônico no processo de formação de professores parece ainda um pouco distante da prática pedagógica da maioria dos profissionais graduados e, talvez, pós-graduados. Sob esse aspecto, o direcionamento ao ensino/aprendizado seria muito beneficiado via as recentes pesquisas realizadas em LA. Além disso, a resistência dos professores em reciclar seus conhecimentos também é um impedimento. Durante a pesquisa, verificou-se que o processo de aprendizagem foi bastante beneficiado através do uso de manipulação de corpora computadorizados.

É importante destacar que de nada adianta diferentes perspectivas em linhas de pesquisa em Lingüística, sem um trabalho de formação de graduandos e pós-graduandos que tenham uma visão crítica sobre o seu saber e sua prática pedagógica, que precisa

sempre ser avaliada, transformada, quando necessária, e possa estar inserida em um contexto social que reflita questões de seu papel de professor-educador de língua materna e estrangeira. A pesquisa, detalhada no capítulo 4, expõe a necessidade do professor de explorar ao máximo as diferentes possibilidades de análise de que a Lingüística de Corpus possibilita, como forma de trazer para o cotidiano de sala de aula um instrumento rico, bastante diferente do uso de computadores com exercícios já conhecidos, esperados e nada interessantes: preencher as lacunas, ordenar coluna A de acordo com a coluna B, dentre outros. A proposta da Lingüística de Corpus é promissora e relevante para os estudos de linguagem, conforme a seção abaixo procura explicitar.

## **1.2. Metodologia de Corpus e o uso de Corpora em sala de aula**

### 1.2.1 Definição de corpus

O termo corpus designa uma quantidade grande de textos que são coletados e, atualmente, estocados em um computador, podendo ser facilmente acessados através de programas específicos de pesquisa. Os textos que fazem parte de um corpus podem ser de diferentes gêneros, tamanhos e podem ser manipulados, selecionados. Por seleção de textos, devemos entender como concentração de gêneros tais como textos de jornais, artigos científicos, romances, etc. A metodologia de análise lingüística denominada Lingüística de Corpus pode ser definida como estudo da linguagem em situações reais de uso, tanto de língua falada quanto de língua escrita. O uso de corpora em sala de aula pode beneficiar o aprendizado porque permite atividades de investigação sobre o uso da língua. Algumas possibilidades de uso de corpora em sala de aula estão descritas na subseção seguinte.

### 1.2.2 Uso de corpora em sala de aula

Os usos de corpora podem ser diversos, dependendo do objetivo e do objeto de pesquisa. Primeiramente, o uso de corpus permite aproximar a língua ensinada na escola àquela usada em situações reais no cotidiano. Em segundo lugar, permite que o aluno possa pensar sobre o código lingüístico que está usando e possa fazer suas pesquisas, além de poder confrontar dados com o padrão apresentado nas gramáticas, especialmente contrastando com o português do Brasil. No caso de ensino de língua estrangeira, um corpus permite que um aprendiz consiga solucionar muitas de suas dúvidas através da consulta de corpus, o que nem sempre é garantido quando se consulta um dicionário bilingüe ou monolíngüe. Essa possibilidade de trabalho impõe o uso de uma metodologia de corpus que permita que os textos sejam pesquisados, em detalhe, ou estatisticamente. Precisa-se, então, de textos digitalizados e, depois, de um programa que possa pesquisar esses textos.

A abordagem com base em corpus é uma metodologia e representa um rompimento com a abordagem de base em regras anteriormente predominantes em contextos de Processamento de Linguagem Natural, no contexto da chamada IA simbólica ou lógica ou de base em regras. Resumidamente, a Lingüística de Corpus é uma metodologia, as concordâncias são o instrumento de análise fundamental e as colocações definem as características das palavras ou sintagmas analisados e o mesmo procedimento será feito em relação aos operadores argumentativos, objeto de estudo dessa pesquisa. O termo concordância não deve ser tomado aqui com o sentido que é atribuído à concordância: flexão de adjetivo ao gênero, número e caso do substantivo, e do verbo ao número e pessoa que estiver se referindo. Trata-se de uma lista de exemplos de uma determinada palavra, expressão ou morfema apresentada como ocorrências em um corpus.

Os benefícios da metodologia da Lingüística de Corpus, para o trabalho em sala de aula, conforme já foi mencionado anteriormente, permitem a aproximação do uso real de linguagem com uma pesquisa detalhada do item lingüístico procurado. É desejável, pois, que quanto mais exemplos pudermos obter de um determinado fenômeno, mais rica será a pesquisa. Por isso a necessidade de dizer que o tamanho de um corpus faz toda a diferença.

Maior detalhamento será dado no capítulo 2 dessa pesquisa e a subseção seguinte aborda a pesquisa executada.

### 1.3 O estudo de caso

Para conceber a presente investigação, o pesquisador optou pelo uso da metodologia de base em corpus porque acreditava que os alunos teriam uma oportunidade de aliar o interesse pelo uso de computadores, uma ferramenta conhecida por todos, com uma pesquisa lingüística e uma nova tecnologia. Além disso, a prática pedagógica da escola, facilitou a interação dos alunos e professores de informática e língua portuguesa. Os postulados de Ducrot foram confrontados com o corpus e com a gramática normativa de língua portuguesa. Era intenção, também, confirmar ou refutar as hipóteses de trabalho levantadas para essa pesquisa, discutidas no capítulo 4. Os alunos foram convidados a participar do projeto e de pronto demonstraram curiosidade e se dispuseram a ajudar. Todos os alunos participaram da manipulação de corpora. Esperava-se que todos os alunos produzissem textos, porém apenas 4 alunos voluntariamente fizeram a produção textual, que foi analisada na seção 4.2. A próxima seção apresenta a organização da dissertação.

### 1.4 Descrevendo a organização teórica do texto

Nesse primeiro capítulo descrevo o cenário da pesquisa, a metodologia de corpus e o uso de corpora em sala de aula. No capítulo seguinte, apresento o referencial teórico e tomo como ponto de partida as atividades de Lingüística Aplicada, isto é, o campo de atuação, de pesquisa e a lacuna existente nas escolas, neste momento, para a abordagem da metodologia de corpus. A diferença entre a Lingüística de Corpus e a Lingüística Computacional faz parte desta seção. Na seção seguinte, abordo os trabalhos de Vygotsky e foco a interação aluno-aluno, aluno-professor. Procurei, também, fazer com que os

aprendizes pudessem refutar ou aceitar os postulados de Ducrot e compará-los às pesquisas com o corpus.

No terceiro capítulo apresento a concepção da pesquisa, os sujeitos, a escola e a coleta de dados. No quarto capítulo, apresento os resultados, a análise, a discussão dos dados e as perspectivas de aplicação e ampliação desta pesquisa. Por fim, a conclusão, a bibliografia e os anexos.

## **CAPÍTULO 2 - REFERENCIAL TEÓRICO**

### **2.1. LINGÜÍSTICA APLICADA – campos de atuação**

A Lingüística Aplicada (LA, doravante), como área de pesquisa científica, é vista hoje como articuladora de saberes relacionados com a linguagem. Esta articulação se dá na medida em que a LA procura vincular ciência e tecnologia. Em outras palavras, aproxima pesquisa e ensino, para além do universo acadêmico. Diferentes linhas de pesquisa são oferecidas nos cursos de pós-graduação, tais como Lingüística Textual, pesquisas com gêneros textuais, Lingüística de Corpus e manipulação de corpora para uso em sala de aula, por exemplo. No que diz respeito a outras áreas de atuação de LA, revelam-se questões como aquisição de linguagem, metodologia de ensino de línguas materna e estrangeira, letramento, pesquisas lingüísticas em comunidades indígenas, distúrbios de aprendizagem, dentre outros. Estão apresentados, sucintamente, nos parágrafos seguintes alguns campos de atuação ainda não mencionados.

A alfabetização, primeiramente, abre portas para o mundo letrado e a leitura viabiliza o acesso a tudo que a humanidade já produziu, podendo encaminhar o aprendiz para o processo de letramento, que busca desenvolver habilidades lingüísticas cada vez mais específicas. A inserção desse aprendiz em práticas de ensino específicas voltadas para um mundo cheio de informações revela a necessidade de se desenvolver aprendizagem da leitura e da escrita, visando a integração desses aprendizes em práticas da sociedade letrada, objetivando o desenvolvimento pessoal e a capacitação de mão-de-obra mais competente e criativa. Poder desenvolver habilidades de interpretar, perceber nuances de uso de linguagem e manipular intencionalmente o código lingüístico para atingir os fins desejados significa poder trilhar o caminho que insira o cidadão para além do aprender a ler e escrever. E, por esse viés, revela-se o processo de pensar sobre sua própria aprendizagem.

A aprendizagem de língua estrangeira enfrenta desafios: elevado número de alunos por turma e a natural discrepância de conhecimento do idioma. Oferecer condições de lidar com

diferentes especificidades da escola, turma - série e conhecimento de língua - além de procurar lidar com os diferentes interesses dos aprendizes é de competência da LA. Recursos didáticos tais como videocassete , gravador, acesso a computadores, revelam a necessidade de novas tecnologias, isto é, procurar motivar e promover aprendizagem num contexto didático mais desafiador e interessante. Novamente, faz-se necessário a aproximação ensino-pesquisa, campo de atuação da LA. Novos papéis para professores e alunos também em L2 surgem como possibilidade de construção de conhecimento, aos moldes do que acontece em Língua materna.

Foram sucintamente apresentados acima alguns campos de atuação de LA. Celani (1995) apresenta as principais linhas de pesquisa que podem ser resumidas assim:

- (i) interação em contextos institucionais e informacionais (negócios, academias, texto literário, sala de aula bilíngüe, bidilateral, monolíngüe- médico-paciente, análise crítica do discurso);
- (ii) interação em aprendizagem (letramento, aprendizagem em L2, interações transculturais e interculturais em contexto pedagógico, com foco sócio-cultural discursivo e psicológico);
- (iii) aquisição e desenvolvimento de linguagem- em relação à língua materna e em relação à língua estrangeira (aquisição e desenvolvimento da escrita, de leitura, de habilidade oral, letramento, alfabetização);
- (iv) formação de docentes para ensino de L1 e L2.

A contribuição de LA hoje se destaca em três aspectos: interação em sala de aula, produção de programas e de materiais de ensino e avaliação de aprendizagem. Sob este aspecto, o direcionamento ao ensino/aprendizagem é muito beneficiado via as recentes pesquisas na área. É importante destacar que de nada adiantam perspectivas em linhas de pesquisa diversas em Lingüística, sem um trabalho de formação de graduandos e pós-graduandos que tenham uma visão crítica sobre o seu saber e sua prática pedagógica, que precisa sempre ser auto-avaliada, transformada, quando necessária, e esteja inserida em um contexto social que reflita questões de seu papel de professor-educador de língua materna

ou estrangeira. A presente pesquisa nasceu da necessidade de mudança e de uma prática pedagógica que respondesse aos meus questionamentos e pudesse ser transformadora para os alunos. A seção seguinte aborda a questão da pesquisa e sua relevância dentro do contexto de LA.

### **2.1.1 Questão da pesquisa**

Conforme foi mencionado no item 1.1, a pesquisa busca, pelo viés do operador *mas*, inserir a Linguística de Corpus no universo escolar, com base em recursos tecnológicos. Procura-se uma alternativa de ensino criativa e desafiadora para a investigação linguística pretendida.

### **2.1.2 Relevância da pesquisa no contexto de L.A.**

Primeiramente, explicita-se a necessidade de uma reflexão sobre a prática pedagógica dos professores de língua materna e estrangeira, seguida da necessidade da consciência de um processo contínuo de formação de professores, para que possamos inserir novas tecnologias no cotidiano de sala de aula. Em segundo lugar, busca-se metodologias e a inserção das mesmas em um projeto de trabalho interessante, novo e desafiador como possibilidade de uma aprendizagem mais enriquecedora. A novidade por si só não se sustenta, porém estando vinculada a objetivos específicos, tais como o uso de corpus na investigação linguística e o uso de uma metodologia específica para esse fim, estabelece vínculos de integração com uma nova abordagem frente à aprendizagem. Os itens 2.2 e 2.3 oferecem maior detalhamento sobre Linguística de Corpus e Linguística Computacional, respectivamente.

Observou-se, durante toda a pesquisa, que o uso de material autêntico, aqui tomado como material que foi produzido em situações reais de uso, em oposição à exemplos inventados, foi essencial para a aprendizagem da linguagem explorada por essa pesquisa, conforme o capítulo *Pesquisa* relata. Além disso, os aprendizes envolvidos com o uso de

corpora no ensino de linguagem reconheceram que o uso de material autêntico foi bastante benéfico na aprendizagem: mais envolvimento, maior motivação, melhor aprendizagem. Os benefícios advindos do uso do concordanceador ( explicado em detalhe na seção 2.2), demonstram como o uso de corpora se mostra mais interessante no processo de pensar a linguagem: poder focar um item de análise, promover comparações, enfim pesquisar sobre o código lingüístico em questão. A próxima seção foca a Lingüística de Corpus e a Lingüística Computacional é abordada na seção 2.3.

## **2.2. LINGÜÍSTICA DE CORPUS**

### **2.2.1 Corpus, Corpora - definição**

**Corpus**, segundo aponta Francis (1982) , é “uma coletânea de textos , considerados representativos de uma certa língua, dialeto ou outro subconjunto de uma língua, para fins de análise lingüística“. Hoje em dia, os textos podem ser facilmente obtidos através da *Internet*, por obtenção de textos já digitalizados, ou ainda, por meio de escanear livros. O tamanho do corpus, sua representatividade, seus usos, aplicações dentro da sala de aula, exemplos de análises de língua materna e língua estrangeira , por exemplo, são aspectos tratados nessa seção, além do recorte histórico do termo em questão : *corpus*.

### **2.2.2 Corpus - antes e depois do computador**

O termo Corpus remonta ao século VI, através de *Corpus Juris Civilis* , usado no direito civil, cuja compilação foi incentivada pelo imperador Justiniano, segundo aponta Francis (1992). Continha leis, casos, e explicações relacionadas a novas leis. Alguns séculos mais tarde, no século VIII, o termo corpus aparece em Corpus Glossary, que consistia em uma listagem de palavras latinas difíceis em ordem alfabética , seguidas por sinônimos latinos mais fáceis ou traduções em anglo-saxão. Rocha (2000) menciona,

ainda, que a finalidade da coletânea já se aproxima da de um corpus lingüístico, embora possa ser melhor caracterizada como uma precursora dos dicionários bilingües.

A pesquisa da lexicografia inglesa feita por Starnes e Noyes (1946) demonstra que o trabalho de coletar ocorrências de palavras em textos escritos para a confecção de dicionários era vista como prática comum. Nathan Bailey, responsável pelo *New Universal Etymological Dictionary* publicado em 1755, não chega propriamente a descrever os métodos, além da menção de ter examinado muitos autores. Samuel Johnson, organizador do *Dictionary of English Language*, discute os métodos de coleta no prefácio da edição de 1755 e também em outro trabalho anterior intitulado *Plan of an English Dictionary*, datado de 1747, descrevendo os métodos de escolha feitas nas definições do sentido das palavras nos verbetes. É preciso explicitar que as citações copiadas eram coladas em seqüência, como numa concordância, para que Johnson pudesse decidir os sentidos a serem incluídos.

Os métodos de coleta de Johnson não são diferentes dos utilizados hoje, naturalmente excluindo o uso de computadores. Johnson fazia a marcação a lápis das palavras que queria incluir nos verbetes e encarregava seus amanuenses, colaboradores, de copiar as citações até chegar a uma conclusão quanto aos diferentes sentidos a serem incluídos nos verbetes. Esta metodologia, com alguma variação em relação às fontes utilizadas, ainda é prática entre os lexicógrafos, de um modo geral. Não sabemos o tamanho do corpus coletado por Johnson, mas imagina-se que fosse grande, considerando que o dicionário inclui 150.000 citações para ilustrar 40.000 verbetes.

A prática de recorte de papel foi utilizada durante muito tempo, incluindo a elaboração do *Oxford English Dictionary*, publicado pela primeira vez em 1884, e do *Webster's New Internatinal Dictionary, Second Edition*, de 1934. Um total de 1.655.000 citações fizeram parte deste último naquela publicação. Para a edição de 1961, foram utilizadas quatro milhões e quinhentas citações, retiradas de livros, jornais, revistas, panfletos, etc. Hoje em dia, há uma tendência de elaboração de dicionários com o auxílio de computadores, permitindo o armazenamento e a manipulação de corpora muito maiores, principalmente no contexto de língua inglesa e por razões de mercado, e não porque seja impossível fazer um dicionário sem corpus. O dicionário de A. Houaiss é um exemplo de um trabalho de dicionários feito sem o corpus.

O método de recortes de papel com citações extraídas de livros, revistas, jornais, etc. também foi utilizado na compilação de corpora para fins de levantamento dialetológico na Inglaterra, França e Alemanha, no final do século dezenove e começo do século vinte. Pouco tempo depois, em torno dos anos trinta, foi feito o estudo de gramática por Otto Jespersen, através da publicação de *Analytical Syntax*, que aponta o uso de mais de mil exemplos compilados através da mesma metodologia. Nota-se a utilização de corpora em pesquisas lingüísticas nos estudos relacionados à aquisição de linguagem, sob a forma de registro de enunciações das crianças, isto é, duas metodologias: o estudo longitudinal (com poucas crianças durante muito tempo) ou a amostragem em larga escala (com muitas crianças da mesma faixa etária).

O lingüista Abercrombie (1965) aponta problemas relacionados ao uso da metodologia de base em corpus, devido à dificuldade em manipular um grande número de informações, de tal forma a superar a problemática da representatividade. Por isso, esse lingüista criou a expressão “pseudoprocedimento” para caracterizar o uso de uma técnica que “embora não seja talvez literalmente impossível, ... seria uma maneira tão árdua e trabalhosa ... que ninguém ... se disporia a utilizá-la.” Na realidade, durante os anos cinquenta, os estruturalistas, mesmo privilegiando o trabalho de corpus como objeto de análise lingüística, não conseguiram levar a termo essa abordagem justamente pela dificuldade, na época, de manipular um grande número de informações. O uso de computadores passa, então, a viabilizar o trabalho da Lingüística de Corpus.

Hoje em dia, o uso de computadores permite buscar, recuperar, selecionar e calcular os conteúdos de corpora com muita rapidez e de forma bastante confiável, já que todas essas tarefas são consideradas elementares em termos de processamento de dados. Os trabalhos de corpora atuais operam com muitos milhões de palavras, sem maiores problemas. Desta forma torna-se imprescindível o uso sistemático de computadores em pesquisa lingüística. Esse instrumento de análise torna possível o empreendimento de pesquisa chamado Lingüística de Corpus e liga a mesma à lingüística computacional. As características desse elo é assunto da seção 2.3. e na seção subsequente abordo a lingüística de corpus, a definição, o paradigma de pesquisa, a abordagem e a questão da representatividade.

### 2.2.3. Lingüística de Corpus

O termo Lingüística de *Corpus* propõe uma nova abordagem relacionada aos estudos de linguagem. A Lingüística de Corpus é uma metodologia, que foi usada em Processamento de Linguagem Natural (PLN) como reação às abordagens de base em regras. A base metodológica é o Corpus, isto é, a linguagem como se mostra em seu uso autêntico, tanto em língua escrita como em língua falada. Portanto, podemos definir lingüística de Corpus como o estudo da língua com base em exemplos reais de uso, segundo McEnery e Wilson. (1996). O termo “corpus linguistics”, de acordo com Leech (1992), aparece pela primeira vez com esse título em um livro publicado por Aarts e Meijs em 1984. Lingüística de Corpus não se refere a um domínio de estudos, mas a uma abordagem metodológica de pesquisa lingüística, pois faz uso de traços característicos, segundo Leech (1992), comparados com outras abordagens em lingüística:

- (1) Focos no desempenho lingüístico, ao invés da competência,
- (2) Foco na descrição lingüística, ao invés dos universais lingüísticos,
- (3) Foco tanto na quantidade, quanto nos modelos qualitativos de linguagem,
- (4) Foco na pesquisa empírica, ao invés de uma visão mais racionalista de pesquisa científica.

Cada um dos traços apontados acima ilustra um contraste entre o paradigma de Lingüística de Corpus Computadorizado (LCC) e o paradigma Chomskiano que dominou muito do pensamento desde 1950. Resumidamente, segundo aporte de Leech (1992), estão detalhados os itens mencionados acima:

#### 2.2.3.1. Desempenho lingüístico, ao invés de competência

A LCC dirige sua atenção para as manifestações naturais de linguagem, tanto na sua forma oral quanto na sua forma escrita. Isto é o que o corpus fornece. A dicotomia Chomskiana está voltada para desempenho, ao invés de competência lingüística e, desta forma, inverte o postulado de Chomsky cujo objeto primeiro de análise é a competência mental de um falante nativo, ao invés de sua manifestação física no uso da linguagem (ver Leech, 1992, para maiores detalhes). É importante ressaltar que desempenho por si só não é um fenômeno isolado, isto é, podemos estudá-lo como processo ou como produto. Ao tomarmos desempenho como processo, tomamos os processos psicológicos envolvidos na produção do texto - falado ou escrito - e suas interpretações. Isto significa que desempenho enquanto processo é parte objeto físico e parte psicológico.

Os estudos de LCC, porém, focam desempenho como produto, isto é, um Corpus se constitui de textos escritos ou falados, a manifestação física da linguagem, independentemente (em princípio) dos processos mentais de seu destinatário ou destinatários, conforme aponta Leech (1992). O termo “em princípio” remete a um limite de objetividade conferida aos textos em um corpus computadorizado. Isto se deve ao fato de que procedimentos de tornar textos legíveis por máquina requerem a transferência de informações de um meio visual ou auditivo para um meio eletrônico, podendo levar a algumas decisões seletivas, no que se refere, principalmente, à transcrição de textos orais. Alguns elementos de subjetividade podem aparecer nesses casos. Por outro lado, na compilação de corpus, a codificação de um texto legível por máquina é geralmente feita através da compreensão que um texto original é recuperável como autêntico tanto quanto possível.

#### 2.2.3.2. Descrição lingüística, ao invés dos universais lingüísticos

Antes da presença do computador na lingüística de corpus, o termo estudo de linguagem baseado em corpus remetia à descrição lingüística. Apesar de conotação um tanto depreciativa, os lingüistas descritivos assumem ser “não-teóricos”, em oposição

àqueles “lingüistas verdadeiramente teóricos”, conforme aponta Leech (1992). A distinção “teórico” / “descritivo” é geralmente usada para a distinção entre as investigações dentro da natureza da linguagem, e descrições da estrutura da linguagem. Por extensão, LCC, alinha-se à lingüística descritiva, neste sentido, como aconteceu com a lingüística de corpus da era pré-computacional. Entretanto, o rótulo “descritivo” não deve ser aplicado com menosprezo à CCL. Segundo palavras de Leech (1992), CCL pode ser tão teórica quanto a “Chomskyan I- linguistics”. Mas a diferença entre ambas reside no domínio ao qual a teoria se aplica.

A descrição da linguagem de um indivíduo tem como seu domínio, a partir do ponto de vista da “E - Language”, isto é, linguagem externalizada, todas as realizações de uma língua (utterances). Trata-se de uma única linguagem. Uma teoria como dos universais lingüísticos aplica-se a linguagem em geral, como uma faculdade humana. O domínio desta teoria remete à descrição de todas as línguas naturais do mundo, cada uma delas passível de estudo específico chamado por Leech (1992) de “lower-order theory” (Leech, 1992). A lingüística descritiva, como entendida por muitos, preocupa-se com uma linguagem isolada, ao passo que a lingüística teórica, trata de uma teoria lingüística de outra natureza, isto é, lingüística de “higher-order”. Ambas as teorias são válidas e devem ser tomadas como mutualmente complementares.

Conforme aponta Leech (1992), a lingüística descritiva, na verdade, está mais sujeita à teoria de construção e testagem em consonância com os princípios de metodologia científica, devido a natureza de suas informações, por exemplo, sentenças de uma linguagem em particular são menos abstratas e mais facilmente observáveis. Ao nomear “descritivo” para LCC, note que as aplicações de lingüística estão quase todas focando o uso e o processamento de linguagens individuais. O termo “lower order” não é uma ordem de menor importância.

### 2.2. 3.3. Foco tanto na quantidade, quanto nos modelos qualitativos de linguagem

Ao transferir o foco da teoria lingüística para a competência, Chomsky não ignorou os métodos quantitativos; apenas tratou as informações de desempenho como secundárias no estudo não-probabilístico de competência. A intenção naquele momento era de criticar os métodos associados ao uso de corpus e, talvez, provocar algum atraso nas pesquisas. Nos anos de 40 e 50, o uso de estatística deu os primeiros passos, através de Shannon e outros, segundo Leech (1992). Porém, esta teoria carecia de informações suficientes, isto é, sem o processamento de corpora feito pelo computador, tornava-se praticamente impossível chegarmos a uma descrição estatística dos dados pesquisados em textos de uma determinada língua. Após esse começo, o foco chomskiano de competência questiona a relevância que as informações quantitativas forneciam.

O renascimento da lingüística de corpus nos anos 80 demonstrou que existe uma conexão entre lingüística de Corpus e lingüística quantitativa. Esta conexão se dá em duas direções. Primeiramente, de posse de um corpus computadorizado bastante grande, o primeiro passo é determinar combinação mais freqüente de uma determinada ocorrência. Por exemplo, podemos calcular a freqüência de *forma irregular* no corpus do NILC: é a combinação mais freqüente de *irregular* (M.Rocha, 2001). Em segundo lugar, se quisermos montar um modelo quantitativo de desempenho de linguagem, a exigência primeira é a de um corpus que permita contagem, contrariando o paradigma chomskiano de confiar na intuição de um falante nativo.

O oposto de quantitativo pode conduzir erroneamente a qualitativo, determinístico “rule driven”, conforme aponta Leech (1992). Ao aplicarmos modelos quantitativos de linguagem fazemos uso de regras e distinções qualitativas. Portanto o único antônimo útil para “quantitativo” passa a ser “não - quantitativo” (“non – quantitative”, nas palavras de Leech, 1992). Isto sugere de maneira bastante apropriada que quantidades podem ser adicionadas a qualquer modelo de linguagem, como um estágio separado de descrição, sem sacrificar nenhum dos traços existentes em cada modelo.

O curso dos estudos lingüísticos negligenciou durante muito tempo os métodos quantitativos, porém sabemos que são usados como rotina em muitas disciplinas. Nesse contexto, como tentativa de descrever qualquer língua, o conceito de freqüente e não-freqüente são indispensáveis. Uma possível explicação para negligenciar estatística pode

ser atribuída à suposição de que corpora contribuiria pouco ou nada nos estudos relacionados a linguagem. Outra razão, apontada por Leech (1992), dirige o olhar para o fato que os modelos quantitativos demandam corpora bastante grandes, tornado possível recentemente. Requerem, também, uma grande quantidade de análise de informações, mesmo para aproximar-se da versão quantitativa de um modelo simples de linguagem, tais como as gramáticas livres de contexto (probabilistic context-free phrase structure grammar).

#### 2.2.3.4. Foco na pesquisa empírica, ao invés de uma visão mais racionalista de pesquisa científica

O termo empírico tornou-se um pouco negativo em lingüística, quando associado ao behaviorismo e a Bloomfield e seus sucessores nos anos 40 e 50. A lingüística de corpus pós- Bloomfield refletia essa tendência. Os trabalhos de Chomsky apontaram para o extremo oposto do racionalismo, isto é, o postulado assegurava que a criança adquire muito de sua linguagem nuclear (core language) através da genética, através de estruturas inatas. Essas estruturas, que determinam o tipo de “ gramática “ que adquirimos para nossa língua nativa, são mentais e portanto não observáveis. A LCC nos leva ao fim do espectro , onde a observação contribui para a teoria mais do que a teoria contribui para a observação. Um exemplo disso é o sucesso do desenvolvimento de reconhecimento de fala , através da modelagem de informações de corpus, onde o modelo de Markov é aplicado.

A LCC não deve apontar para qualquer posição extrema na direção empírica, mantendo ênfase na testabilidade em oposição às informações obtidas independentemente. As informações de um corpus, de maneira mais perfeita do que se esperava em lingüística, são independentes dos princípios de uma teoria . A LCC, porém , não nega os princípios racionalistas que constróem uma teoria e determina uma maneira de categorizar e interpretar nossas informações.

#### 2.2. 4. O paradigma de pesquisa empírica em LCC

Segundo G. Leech (1992), o paradigma de pesquisa em LCC aponta para o seguinte padrão:

- (a) evidências observadas são fornecidas em forma de corpora e tomadas meramente como amostras de linguagem em uso;
- (b) tais informações são usadas exaustivamente: não há seleção prévia das mesmas,
- (c) Alguma teoria ou “ modelo “ é planejada para dar conta das informações de corpus,
- (d) A acuidade do modelo é testada, se desejável , em outros corpora,
- (e) A qualidade de um modelo pode ser medido e comparado com outros modelos.

Na maioria dos casos, a metodologia com base em corpus delineada acima alinha-se aos métodos científicos. (veja Leech, 1992 para maiores detalhes). É importante explicitar os **atributos de um Corpus** para ressaltar os conceitos fundamentais que norteiam a abordagem com base em corpus, a saber, **amostragem, representatividade, tamanho, referência padrão**. Os atributos de um corpus mencionados acima revelam aspectos importantes da lingüística de Corpus, conforme já foi mencionado , o discurso escrito e falado, conforme ocorre naturalmente é objeto de investigação essencial da lingüística. As investigações se baseiam em dados abundantes e observáveis e serão usadas em aplicações lingüísticas.

A **amostragem** pode ser definida como um recorte de textos a ser pesquisados e retratam uma variedade de uma língua viva onde o número de enunciações aumenta constantemente . Essa amostragem pode ser considerada como sendo de máxima **representatividade** quando o **tamanho** do corpus faz a grande diferença, isto é, quanto maior o número de palavras, maior será o número de exemplos e possibilidades de limites de produção. Uma ampla variedade de autores e gêneros deve fazer parte de um corpus considerado representativo de uma língua. A **referência padrão** refere-se aos resultados e

a disponibilização dos mesmos para outros pesquisadores, onde a metodologia aplicada pode ser avaliada. Para que a pesquisa se realize e possa atender aos atributos de um corpus, a seção seguinte, de forma sucinta, traz informações sobre os paradigmas de pesquisa baseada em corpus.

#### 2.2.4.1 O paradigma de pesquisa em Lingüística de Corpus Computadorizado (LCC)

Estão sucintamente abaixo relacionados os três paradigmas de pesquisa baseada em corpus , segundo Leech (1992):

#### 2.2.4.2 O paradigma da concordância

A mais simples e mais usada ferramenta de análise da lingüística de corpus. Nesse contexto, o termo não tem relação com o sentido que é atribuído à concordância: flexão de adjetivo ao gênero, número e caso do substantivo, e do verbo ao número e pessoa que estiver se referindo. Trata-se, no entanto, de uma lista de exemplos de uma determinada palavra, expressão ou morfema, apresentados como ocorrências em um corpus. Podemos, hoje, graças ao computador, fazer uso de um programa concordanceador que mostra a palavra escolhida no centro de um fragmento extraído do corpus. O contexto pode ser especificado, através da escolha do número de palavras à direita e à esquerda da palavra escolhida. A partir desse ponto, podemos saber o número de palavras presentes no corpus, as ocorrências da palavra procurada, possíveis combinações mais freqüentes e possibilidades de maior detalhamento de uma determinada combinação.

Poder construir análise de uma determinada palavra e expandir essas *colocações* , como são chamadas as combinações num contexto da lingüística de corpus, a partir de resultados produzidos por um programa concordanceador , viabiliza o cruzamento de colocados mais freqüentes e de colocados menos freqüentes. O exemplo da palavra

**irregular** (Rocha, 2001) evidencia colocados como *forma irregular, de forma irregular, uma forma irregular*, por exemplo. Essa possibilidade de pesquisa lingüística, integrada a um projeto que visa à aquisição de vocabulário, por exemplo, transpõe a reprodução de exercícios esperados, enfadonhos, tradicionais e insere o uso de tecnologia computacional como alternativa mais desafiadora e interessante para a prática de língua portuguesa na escola. A pesquisa realizada enfoca esse aspecto, em relação aos operadores argumentativos, e o Capítulo 4 aborda detalhes do uso do concordanceador.

#### 2.2.4.3 O paradigma do modelo log-linear

É uma variante mais rigorosa do modelo descrito acima. O modelo é similar ao de Altenberg (1982) no estudo feito com o genitivo de língua inglesa do século 17. Imagine que estamos focando o mesmo objeto: genitivo inglês, porém de inglês atual usando corpora computadorizado. Usando o modelo do concordanceador, podemos chegar a alguns parâmetros e níveis, assumindo que vão atuar na previsão das ocorrências de construções do genitivo: “*X’s Y*”, em contraste com a construção “*of- construction*” de “*Y of X*”. Quatro prováveis fatores (ou parâmetros) são:

- a) O tipo semântico da expressão nominal ocorrida em *X*. Uma escala de “personhood” é postulada ao passo que o *X* no genitivo é mais provável de remeter a uma pessoa do que a um animal. Uma organização, um evento, etc., ao passo que o oposto é verdadeiro na construção “*of-*”.
- b) O tipo de relação semântica entre *X* e *Y*. O possessivo, subjetivo, objetivo, etc. relações podem receber diferentes pesos em favor do genitivo ou de construções “*of- constructions*”.
- c) O comprimento relativo ou complexidade sintática de *X* e *Y*. (medido pelo número de palavras, por exemplo.)
- d) O tipo de texto no qual o genitivo ocorre. Pode ser previsível, é mais provável que ocorra em textos informais do que em textos formais. (Altenberg, 1982:251-254).

Os dados que especificam estes fatores para cada um dos casos da amostra são organizados em tabelas. Um pacote estatístico qualquer é geralmente usado para construir um modelo log-linear com base nestes fatores que influenciam no uso do genitivo. Usando o corpus, podemos analisar cada exemplo, em relação a outros fatores e poderemos estabelecer as singularidades em favor dos genitivos, segundo aponta G. Leech (1992).

#### 2.2.4.4 O paradigma dos modelos probabilísticos da linguagem usando os modelos ocultos de Markov

Este paradigma desenvolveu-se através da influência da teoria da informação de fala e aplicações de processamento de linguagem natural. A proposta deste paradigma conta como uma tentativa para formar e testar um campo de modelos de desempenho de toda uma linguagem. Há duas desvantagens : (a) o modelo no momento é um pouco simplista e (b) operam com um princípio para o qual um modelo de linguagem deve explicar, por exemplo como a seqüência de segmentos de um corpus é prevista através de um modelo estocástico (stochastic) tal qual os modelos ocultos de Markov ou uma gramática probabilística, ou na verdade um complexo de tais modelos ( ver Leech , 1992 para maior detalhamento).

O paradigma do concordanceador foi escolhido por se tratar de um modelo mais simples para os usuários do *Windows* e o *WordSmith* ( Scott, 1998) o programa concordanceador escolhido para o estudo de caso. A seção seguinte aborda as concordâncias antes de depois do computador.

#### 2.2.5 Concordâncias - antes e depois do computador

A metodologia de análise lingüística conhecida atualmente com lingüística de corpus, segundo McEnery e Wilson ( 1996) , pode ser definida como o estudo da língua com base

em exemplos autênticos de uso. A disseminação dos recursos computacionais alterou sobremaneira a situação dos estudos relacionados a corpus. A inserção do computador, como ferramenta de pesquisa ( buscar, recuperar, selecionar e calcular os conteúdos de corpora ) possibilita ao lingüista uma investigação bastante segura e fundamentada nos dados. A idéia, porém, de estudar a língua fazendo uso de corpus é bastante antiga e antecede a existência do computador. Conforme já foi mencionado anteriormente no item 2.2.1., o termo Corpus Linguistics aparece em uma coletânea de artigos, com o mesmo título, organizada por Aarts e Meijs ( 1984 ).

Há registros de tarefas relacionadas a concordâncias pelo menos a partir do século XIII, conforme aponta M. Rocha (2001): “quando Hugo de San Charo mobilizou quinhentos monges para a tarefa de produzir uma concordância completa da Bíblia em latim “. Imagina-se que as palavras e expressões consideradas importantes para o estudo da Bíblia foram listadas em ordem alfabética em todos os contextos em que foram usadas, permitindo, por exemplo, que o leitor analise o uso do equivalente em latim do substantivo *glória*. O responsável pela elaboração da concordância atribuía a cada monge a tarefa de investigar uma determinada palavra. M. Rocha (2001) aponta para o fato de cada monge encarregado de produzir a concordância da palavra *glória*, dava início ao trabalho na primeira página de sua cópia da Bíblia. A cada ocorrência, copiava-a, em seu contexto imediato e repetia o procedimento a cada nova ocorrência ( ver Tribble e Jones (1990) para uma história das concordâncias)

A rotina de encontrar a ocorrência, copiá-la no seu contexto imediato, escolher, por exemplo, cinco palavras à direita e cinco à esquerda, retornar à leitura do texto e repetir o procedimento a cada nova ocorrência era repetida inúmeras vezes. Alguns problemas podem ter acontecido ao longo de todo esse processo, pois desvios de foco de atenção poderiam ocasionar o registro de uma ocorrência com uma grafia diferente da originalmente encontrada, ou alguma ocorrência era deixada de lado, para citar algumas das diferentes possibilidades de erro empregadas ao longo desse método. É provável, portanto, que as concordâncias da Bíblia apresentem alguns “defeitos”, porém não chegam a prejudicar os estudos realizados a partir das concordâncias. Por outro lado, as exigências

de precisão são bem maiores no que diz respeito aos estudos científicos da linguagem com base no tipo de coleta de dados realizada hoje em dia.

Alguns séculos depois, podemos encontrar um colaborador trabalhando para Samuel Johnson na elaboração de seu dicionário de língua inglesa. Lendo publicações previamente selecionadas, esse amanuense buscava a ocorrência que lhe fora solicitada. Por se tratar de um dicionário apenas um pequeno número de autores e gêneros textuais não seriam suficientes para detalhar outros possíveis sentidos atribuídos à palavra procurada. Poder garantir a **representatividade** da amostra passa, então, a ser uma preocupação. O problema da representatividade, conforme foi citado no item 2.2. 1, refletido o pensamento vigente nos anos cinquenta e sessenta em relação às abordagens de base em corpus, levou Abercrombie (1965) a criar a expressão “ pseudoprocedimento “ para uma abordagem que, devido as dificuldades de manipulação de um grande número de informações, deveria superar o problema da representatividade.

Os recursos computacionais modificaram o cenário: a capacidade dos computadores de buscar, recuperar, selecionar, calcular os conteúdos de corpora muito rapidamente, possibilitando ao lingüista o tratamento dos dados pesquisados de forma mais segura e confiável. Muitos milhões de palavras são manipuladas através do uso de computadores e, por vezes, sofisticados através do uso de modelos estatísticos mais complexos. É possível dizer que o computador seja um divisor de águas no que diz respeito ao uso de Corpora e concordâncias em pesquisas lingüísticas. Conseqüentemente, podemos incorporar ao ensino de língua a inserção de concordância extraída de corpora computadorizados como uma metodologia bastante útil e interessante no processo ensino-aprendizagem. A seção seguinte aborda a concordância na sala de aula.

### 2.5.1 A concordância na sala de aula

Conforme já foi detalhado anteriormente, o termo concordância pode ser definido como uma lista de exemplos de uma determinada palavra ou expressão, conforme apresentados no corpus. Podemos obter uma concordância através do programa

concordanceador WordSmith, ou fazendo uso do Linux podemos utilizar comandos *grep* para atingirmos os mesmos fins. O uso do concordanceador é uma solução simples para os usuários do Windows e o WordSmith é uma alternativa mais barata e de fácil execução. A palavra escolhida aparece no centro da tela e o contexto pode ser especificado através do número de palavras à direita e à esquerda da palavra escolhida. Abaixo há um exemplo de dez casos extraídos de uma concordância para a palavra *irregular* (Rocha, 2001):

Folha comprovou a comercialização **irregular** de bônus eleitorais pelo

e bloqueio, a recepção do time anda **irregular** e o saque está prejudicado

ações continuam a ser feitas de forma **irregular**. Só que agora implicam

a Portuguesa apresentou um futebol **irregular**. Mas fez o suficiente para

os algodoais sofreram com o clima **irregular** e o surgimento de alguns

de altitude. A formação geográfica é **irregular**. Não há planícies. O clima

apreendeu anteontem 80 kg de pescado **irregular**, em operação realizada

faias, álamos, distribuindo-se de modo **irregular**. Na região mediterrânea

enormes blocos de pedra de formato **irregular**, toscamente ajustados,

casa: é Berger, um piloto atabalhoado e **irregular**, especialista em pôr

A concordância foi extraída do corpus do Núcleo Interinstitucional de Linguística Computacional (NILC, doravante). Foram encontradas 733 ocorrências para a palavra *irregular*. Fazendo uso de alguns recursos do WordSmith, podemos descobrir que “*forma irregular*” é a combinação mais freqüente de *irregular* com palavra não-gramatical. (50 ocorrências), seguida de *situação irregular* (25 ocorrências), *uso irregular* (21

ocorrências), e *venda irregular* (19 ocorrências). Podemos expandir as colocações para três palavras: *de forma irregular*, *venda irregular de*, *de maneira irregular*, *em situação irregular* dentre outros (ver Rocha, 2001 para detalhes).

Outra possibilidade de uso do concordanceador é apontada por P. Humblé (2001) e ilustrada através do exemplo de dez ocorrências de uma concordância para a expressão ‘*en circulación*’ de língua espanhola utilizando o MicroConcord search SW:

1. d de las acciones **en circulación** . Los usual es que
2. o de las acciones **en circulación** de la mencionada
3. iones actualmente **en circulación** est asegurada
4. reflecto del aire **en circulación** . <p> Se estima que
5. cidas por el aire **en circulación** de la c maras, po
6. de su capital **en circulación**, se experimenta c
7. os de los chistes **en circulación** , se experimenta c
8. er dos compañeros **en circulación** y un solo
9. to de los diarios **en circulación** . Juro que leo
10. de dinero **en circulación**. El que se respo

Humblé ( 2001) aponta que “Como os acentos são codificados de maneira diferente de acordo com cada publicação, o Mconcord, programa utilizado, aconselha trocar acentos por um ponto de interrogação, o que explica que SW - search word - seja

‘*en circulación*’ (esse problema de acentuação inexistente no Wordsmith). O quadro mostra quais substantivos que a expressão *en circulación* em espanhol aceita como acompanhantes: *acciones*, *aire*, *capital*, *dinero*, e verbos : *encontrar*, *entrar* e *poner*. “ Os exemplos foram reduzidos a um número de 10 , por questão de ilustração, apenas. Nota-se porém, o padrão de apresentação do concordanceador, que pode ser aplicado em pesquisas em língua materna ou língua estrangeira. No caso da expressão ‘*en circulación*’, P. Humblé ( 2001) investigava como se diz em espanhol a expressão por em circulação: ‘*poner en circulación*’ ou ‘*colocar en circulación*’? “Por razões provavelmente arbitrárias , se diz em espanhol ‘*poner en circulación*’ e não ‘*colocar*’ . É um dado que não é

mencionado pelos dicionários bilingües ou monolíngues consultados. Com o corpus, no entanto, a dúvida “ é rapidamente sanada “, conforme aponta o autor.

Podemos dizer, então, que corpora computadorizados e concordâncias deles extraídos trazem muitos benefícios para o estudo de língua, pois aspectos morfológicos, sintáticos ou pragmáticos podem ser explorados. Assim, pesquisas relacionadas a vocabulário e estrutura do discurso podem se tornar muito mais atraentes para os alunos. Uma decisão que diz respeito ao corpus a ser pesquisado e qual a dimensão do mesmo pode trazer problemas para o professor que pretender aplicar essa metodologia e , também para as máquinas que serão utilizadas para esse fim. É possível selecionar corpora de menor dimensão, para tornar a investigação facilitada . Uma quantidade muito grande de dados pode dificultar a manipulação de corpora computadorizados, se for feita por usuários com pouca experiência .

Um corpus das dimensões do Núcleo Interinstitucional de Lingüística Computacional, NILC, “é uma amostra bastante representativa do uso de língua portuguesa em textos escritos. Isto não significa que seja o *corpus* ideal para uso no ensino.” , segundo M. Rocha (2001) . A pesquisa relatada no capítulo 4 demonstra com clareza a necessidade da escolha e adequação da dimensão do corpus com os recursos tecnológicos oferecidos pela escola. Por essa razão, então, pré-selecionou-se o corpus , definiu-se objetivos de trabalho, aplicou a metodologia geral baseada em Lingüística de Corpus e fez-se uso de concordâncias, que não revelam apenas aspectos de gramática pelo viés dos operadores argumentativos, mas principalmente modificam o paradigma das relações aluno-professor, aluno - conhecimento e o processo de construção de conhecimento compartilhado.

Parece razoável afirmar que corpus de pequeno porte tende a promover a inserção da pesquisa lingüística, através de assuntos variados e gêneros textuais distintos. O professor pode promover uma pesquisa relacionada às preferências do grupo que está trabalhando e, posteriormente, selecionar textos que sejam representativos de um padrão de excelência dentro do assunto em questão e do gênero escolhido. Podemos ampliar o leque trazido pelo professor, promovendo a extensão do tema através de pesquisas na *Internet* e coletarmos um corpus de algumas mil palavras, para o assunto em questão. É importante salientar que além da pesquisa lingüística, o objetivo maior do uso de corpora no ensino está

diretamente ligado aos objetivos traçados para o aprendizado. O Capítulo 4 dessa pesquisa demonstra como procurou-se promover essa aproximação: corpus – aprendizado.

Um corpus também pode ensinar sobre as diferenças entre a língua falada e a língua escrita. “ Por razões práticas evidentes, as gramáticas e os dicionários descrevem a língua escrita e não a falada. A língua escrita pode ser pesquisada mais facilmente, ela não se mexe”, conforme aponta P. Humblé (2001). Podemos consultá-la quantas vezes quisermos. Alguns poderão argumentar que o mesmo se aplica à língua falada, já que todo pesquisador ‘fala’. É fato, porém, que temos pouca consciência de como usamos a nossa língua. Muitas vezes sabemos menos sobre ela do que sobre uma língua estrangeira que aprendemos de forma mais consciente. Tomando o exemplo citado por P. Humblé (2001), procurou-se esclarecer o exposto acima através da palavra “coisa” aparentemente usada por todos.

Ao investigarmos o NILC ( 31.284.099 palavras ) percebemos o quanto as pessoas usam a palavra “coisa” na língua falada quase da mesma forma que a utilizam na língua escrita. Ao olharmos para as palavras que acompanham “coisa” a direita percebemos que quase todas ocorrem tanto na fala quanto na escrita, isto é, “que”, “mais”, “para”, “não”, “muito”. Segundo aponta P. Humblé, “só há uma exceção, a palavra “assim”, que nem aparece na lista de frequência da *Folha*. “ Reconhecemos expressões “uma coisa assim”, “alguma coisa assim” e “qualquer coisa assim”, que aparecem com bastante frequência na língua falada e constituem usos mais significativos da palavra “coisa”. Entretanto, essa possibilidade não nos passaria pela cabeça, afinal o significado de “coisa” é possivelmente ‘objeto’, ou ‘assunto’.

Estendendo essa pesquisa para a língua inglesa usaríamos a expressão *or so*, nem de perto uma sombra da palavra *thing*. Em francês, “alguma coisa assim” pode ser traduzido por *quelque chose* “em que encontramos a palavra *chose*, mas geralmente será traduzido por *quoi, par exemple* ou *comme ça*, segundo palavras de Humblé. “Qualquer coisa assim” seria traduzido por *n’importe quoi*. No Corpus da *Folha de São Paulo (1995)*, por outro lado, a expressão “coisa assim” só aparece uma vez a cada 1.375.604 palavras, de acordo com o mesmo autor. Ainda assim, na maioria das frases em que “coisa assim” aparece na *Folha*, vem acompanhado de “eu” na mesma frase ou numa frase adjacente. Isso quer dizer

que se trata de uma citação de alguém escrevendo em tom informal. Além disso, os analistas mais tradicionais de língua nem sempre têm consciência deste uso real das palavras. Comprova-se isso pelo tratamento dado aos dicionários. (ver Humblé (2001) para maiores detalhes) .

Uma análise no corpus, portanto, pode levar a conclusões bastante interessantes e surpreendentes: primeiramente poder estender o significado de uma determinada palavra ou expressão para além de seu significado literal, isto é, uma outra possibilidade de uso além da comumente empregada por todos. Em segundo lugar, procurar inserir uma metodologia que investigue a língua falada e escrita, sem o “ranço do certo e errado na linguagem” significa poder fazer do uso prático de um corpus, uma arma reveladora e suficientemente capaz de resolver problemas que os verbetes de dicionários e gramáticas tradicionais se revelam insuficientes como questões de produção textual, por exemplo, em língua materna ou estrangeira. Por vezes, o potencial de exemplos que um corpus fornece resolve o problema em questões como decidir entre *quedarse triste?* ou *ponerse triste?*, em espanhol. Ou ainda em inglês: *tools to learn a foreign language?* ou *tools for learning a foreign language?*. Essas questões nem sempre são resolvidas com o uso de gramática ou dicionários. E a investigação através do uso de corpus faz toda a diferença, por razões já mencionadas. A próxima seção procura orientar os interessados sobre *como e onde* conseguir textos.

#### 2.2.5.2 Onde conseguir textos

Os textos podem ser obtidos de diversas maneiras, dependendo dos objetivos específicos que serão trabalhados. Podemos dizer que quanto maior o corpus, melhor. Isso se deve ao fato que um número maior de exemplos podem evidenciar um determinado fenômeno, uma determinada expressão verbal, um determinado verbo, etc. Além disso, certas expressões aparecem em um número reduzido e quanto maior o corpus, maior será a possibilidade de as encontrarmos. A probabilidade de encontrarmos o verbo “eclipsar” é maior num corpus grande, isto é, um corpus de alguns milhões de palavras. Os meios mais

comuns de conseguir textos são escanear livros, obter textos já digitalizados e copiar textos da *Internet*.

O trabalho de escanear livros talvez esteja ultrapassado, pois muitos textos já se encontram digitalizados ou disponíveis na *Internet*. Uma grande quantidade de autores teve suas obras editadas para a versão digitalizada. Caso contrário, podemos escanear alguma obra específica. Diferentes autores de línguas estrangeiras podem ser encontrados na rede em endereços como o Projeto Gutenberg ( <http://promo.net/pg> ). No caso de literatura brasileira, o NUPILL ( [www.cce.usfc.br/~nupill](http://www.cce.usfc.br/~nupill) ) tem algumas obras escaneadas. A segunda maneira de reunir textos é fazer uso dos *cds* que vários jornais costumam publicar incluindo a produção de um ou mais anos. Na presente pesquisa foi utilizado o corpus do NILC e selecionados os textos da Folha de São Paulo (1995). Nesse caso, os artigos foram extraídos desse *cd* e pesquisados por meio do programa Wordsmith. Podemos utilizar o *Mconcord*, disponível na rede gratuitamente no endereço <http://lexically.net/download-freebies/mconcord.zip>. Existe, ainda, uma terceira possibilidade de conseguir corpus que consiste em escolher textos ou publicações inteiras na *Internet* e fazer o *download* para um computador pessoal. Na Universidade Federal de Santa Catarina foi feito um corpus de textos em língua espanhola dessa maneira. São textos de vários países e de vários tipos, principalmente jornalísticos. O total de textos soma aproximadamente 12 milhões de palavras, o que permite uma busca bastante significativa para resolver a maioria das dúvidas que surjam.

Como já foi mencionado anteriormente, o tamanho de um corpus pode variar, porém um corpus de 10 milhões de palavras parece suficiente para um trabalho confiável. Existem vários corpora no mundo cuja compilação foi financiada por universidades, empresas ou órgãos oficiais. Os corpora de língua inglesa são em número maior. O corpus do dicionário *Cobuild*, por exemplo, conta com 300 milhões de palavras. Ele consiste em diversos sub-*corpora* de literatura, periódicos, língua falada e outros. Dessa forma tornam-se possíveis pesquisas mais precisas sobre determinados tipos de linguagem. Porém, para um trabalho com corpus em sala de aula, não precisamos de um número tão grande de palavras.

Podemos, ainda, promover uma busca na *Internet* quando quisermos saber se uma determinada palavra ou expressão existe ou não. Programas de busca como *Google* ( [www.google.com](http://www.google.com)) procuram uma ou várias palavras em milhões de páginas de *site*. O que é prático no caso do *Google* é que este programa além de dar o endereço, dá a frase em que está a palavra ocorre, evitando que tenhamos que abrir cada endereço para procurar a palavra ou expressão desejada.

### 2.2.5.3 O concordanceador e a pesquisa realizada

O pesquisador escolheu o paradigma do concordanceador por razões bastante práticas e de pouco custo para a escola investigada. Na verdade, o uso disseminado do *Windows* foi determinante, pois o programa *Linux*, outra possibilidade de aplicação do concordanceador, havia “sido posto de lado”, por total falta de alguém que soubesse “p’ra que que serve?”, nas palavras de uma das professoras da escola. Havia também um outro contratempo a ser resolvido que seria localizar o programa, isto é, estava guardado em algum lugar na escola e ninguém sabia onde. Portanto, a escolha óbvia do *Windows* foi o caminho trilhado. Como foi mencionado na seção 2.5.1., o programa concordanceador *Wordsmith* foi escolhido por uma questão de imposição das disponibilidades daquele momento. Foi o primeiro contato dos alunos com o trabalho de corpus e o concordanceador.

Resumindo a seção, pode-se afirmar que a base metodológica serve de apoio a outros campos da lingüística, tais como sociolingüística, fonética, sintaxe e outros aspectos da lingüística utilizando corpora, conforme aponta Leech (1992). Atualmente, referir-se à Lingüística de Corpus implica assumir que o corpus seja legível por máquina, isto é, uma grande quantidade de textos estão estocados num computador e são facilmente acessados através de programas próprios para pesquisas. O uso de computadores permite buscar, recuperar, selecionar e calcular os conteúdos de corpora com muita rapidez e de forma bastante confiável,, possibilitando compreender e calcular o teor de tais corpora de uma forma que antes da era do computador seria impraticável. A próxima seção trata da definição de Lingüística Computacional e a inserção da Lingüística de Corpus nos campos de atuação da Lingüística Aplicada.

### 2.3. LINGÜÍSTICA COMPUTACIONAL

A seção aqui apresentada não pretende ser exaustiva, porém apresenta de maneira sucinta uma definição de lingüística computacional, objetivos, metodologia, maneiras de abordar a língua e enfoca, também, a formação do lingüista computacional. O termo *Lingüística Computacional* pode remeter o leitor a uma relação direta com a palavra computador. De fato, essa relação existe. Porém computador é uma ferramenta de trabalho. Lingüística Computacional é uma metodologia de análise e investigação de fenômenos lingüísticos, quer de natureza fonológica, morfológica, sintática ou semântica, por exemplo. A Lingüística Computacional tem dois objetivos fundamentais: o primeiro de natureza tecnológica e o segundo com objetivos cognitivos. Por natureza tecnológica entende-se o funcionamento do sistema, isto é, tradução de máquina, interfaces de acesso a banco de dados e instrução com a ajuda de computador. Os objetivos cognitivos, por outro lado, focam atenção para melhor compreensão da linguagem humana e uso da linguagem para comunicação.

Estabelece-se uma interdisciplinaridade intrínseca entre as áreas da lingüística, psicologia, ciência da computação e filosofia; psicolingüística, psicologia computacional, interação humano-máquina. A proposta de interdisciplinaridade pode ser bastante frutífera, na medida que existe uma preocupação com o processo de formação do lingüista computacional, isto é, o computador como ferramenta de análise ou especializações dentro da lingüística computacional, focando lingüística e programação ou lingüística e WWW. A possibilidade de inserção de novas tecnologias no ensino de língua materna ou estrangeira encontra na lingüística computacional o suporte para sua implementação, desde que haja uma preocupação real de aproximar ensino e pesquisa.

Podemos abordar a língua de duas maneiras: abordagem baseada em regras e abordagem baseada em corpus. A abordagem baseada em regras foi fortemente influenciada por conceitos de psicologia cognitiva, lógica dedutiva e as teorias lingüísticas influenciadas por N. Chomsky. Nessa concepção, os sistemas de processamento automático de linguagem têm que fazer uso de regras gramaticais e capacidades de inferência lógica

dentre outras coisas, já que o processamento humano de linguagem envolve uma compreensão ativa da linguagem processada.

Se o objetivo é fazer a máquina conseguir um desempenho em termos de capacidade lingüística semelhante ao humano, talvez o melhor caminho seja simular essas capacidades por meio de uma reprodução do funcionamento da própria inteligência humana. Lamentavelmente, a capacidade das máquinas atualmente disponíveis no mercado não se aproxima da inteligência humana, principalmente em tarefas de processamento de linguagens naturais. Conseqüentemente, as pesquisas de base em regras apresentaram uma tendência a orbitar em torno da gramática gerativa, aqui tomada como sistema de regras onde todas e somente as sentenças bem formadas de uma língua são definidas. Apesar da discordância entre pesquisadores em relação à natureza formal das regras gramaticais deste tipo, é consenso, porém o fato de que *parsing* - análise sintagmática - será realizada através da comparação de uma entrada a ser processada com as estruturas estabelecidas em uma gramática, de tal forma que sejam encontradas uma ou mais estruturas que se encaixem, caso contrário a frase não é gramatical.

Se fosse possível produzir uma gramática gerativa capaz de reproduzir toda a variedade de uma língua, não teríamos garantia de que seria possível implementá-la, dada a dificuldade de projetar um algoritmo que pudesse localizar uma determinada estrutura correta contida nessa gramática para lidar com uma determinada entrada a ser processada. Segundo M. Rocha (2000), “ embora o mapeamento a partir das estruturas previstas em gramática para as formas literais possa ser definido..., o mesmo não pode ser dito do mapeamento a partir das formas literais reais de uso cotidiano para as estruturas previstas na gramática.” Na verdade esse mapeamento não constitui problema para a teoria da gramática de base gerativa, mas é fundamental para a análise de entradas produzidas por um usuário de um sistema de PLN.

Outra possibilidade de análise lingüística é a abordagem baseada em corpus. Como o termo corpus já foi definido na seção 2.2., é possível considerar a resistência de lingüistas e pesquisadores face a essa abordagem, pois algumas das noções como frequência e probabilidade, podem demandar técnicas estatísticas mais complexas. Tomando o seguinte exemplo em língua portuguesa, *inteligente*, o usuário escolhe se vai utilizar somente o

adjetivo ou se vai, ainda modificá-lo através do advérbio *muito* ou *pouco*, por exemplo, criando expressões completamente diferentes. Podemos, talvez, afirmar que a probabilidade de utilização de *muito inteligente* ou *pouco inteligente* em relação a *não inteligente* é muito maior. O usuário de linguagem natural observa e seleciona “naturalmente” o que é mais freqüente na língua.

Essas **colocações**, isto é, combinações de palavras como são chamadas no contexto da lingüística de Corpus, atravessam as línguas naturais de maneira muito mais significativa do que é possível verificar através de exemplos inventados. Um escritor leva em consideração este tipo de associação natural ao produzir linguagem. Poder observar de que forma as colocações podem assegurar o uso de tal palavra à direita ou à esquerda do item lingüístico escolhido, insere a metodologia de base em corpus em áreas de estudo que até agora procuram resistir a essa possibilidade de investigação. A aproximação com o uso real da língua, ao invés de exemplos inventados, facilita o trabalho do aluno no que diz respeito à pesquisa, à observação, e à confrontação com a gramática tradicional.

A abordagem com base em corpus possibilita uma série de estudos no campo da lingüística e também, a elaboração de sistemas de PLN, que obtêm melhores resultados do que os sistemas baseados em regras. Há muitos passos a serem dados em relação às interações entre léxico e sintaxe, interações que não são capturadas por formalismos de produção de linguagem que não levam em conta noções como freqüência e probabilidade. As técnicas estatísticas mais complexas possibilitaram a implementação de sistemas de reconhecimento de fala, atualmente todos os sistemas de reconhecimento de fala operacionais fazem uso de informações estatísticas, e os resultados também são promissores.

Podemos, talvez, afirmar que os avanços conquistados devem-se muito mais à capacidade dos computadores de lidar com modelos estatísticos e cálculos matemáticos bastante complexos, do que àquelas capacidades de reconhecer estruturas abstratas de produção de linguagem em situações reais de uso a serem analisadas, para que uma máquina possa chegar a uma representação de sentido. Conseqüentemente, durante os anos noventa, houve um incremento de pesquisas e iniciativas comerciais na área de lingüística computacional que utilizam informações retiradas de um ou mais corpora. Muito desse

trabalho é devido ao barateamento dos computadores e a resultante popularização dos computadores domésticos. Pesquisadores e instituições de ensino e pesquisa passaram a dispor de uma metodologia que antes estava restrita a alguns nichos privilegiados. A próxima seção procura estabelecer uma ligação entre a lingüística computacional e a lingüística de corpus e a inserção dessa metodologia dentro da salas de aula de língua materna (L1) e língua estrangeira (L2).

### **2.3.1 LINGÜÍSTICA COMPUTACIONAL, LINGÜÍSTICA DE CORPUS - uma ligação**

Como foi visto na seção 2.2, o termo corpus pode ser empregado para descrever qualquer coletânea de textos reunida para os mais variados fins. Na seção 2.4, o termo lingüística computacional foi definido como uma metodologia de análise e investigação de fenômenos lingüísticos : fonéticos, morfológicos, sintáticos ou semânticos. A ligação da lingüística computacional com a lingüística de corpus se dá na medida em que ambas são metodologias que investigam fenômenos da linguagem através do uso de computador. A Lingüística de Corpus tem no corpus seu objeto de estudo, ao passo que a lingüística computacional pode abordar os fenômenos lingüísticos quer pelo viés da abordagem com base em regras ou através da abordagem com base em corpus. A seção seguinte procura inserir a metodologia de corpus na sala de aula de língua materna (L1) e de língua estrangeira (L2).

### **2.3.2 INSERÇÃO DA METODOLOGIA DE CORPUS NA SALA DE AULA DE LÍNGUA MATERNA E DE LÍNGUA ESTRANGEIRA- CAMPO DE ATUAÇÃO DA LINGÜÍSTICA APLICADA**

Conforme já foi mencionado anteriormente, a resistência frente à inserção de uma nova metodologia dentro da sala de aula pode ser tomada como medo, insegurança, necessidade de manter a todo custo uma prática pedagógica já vigente, pois o novo promove mudanças, implica ajustes teóricos e pedagógicos e insere uma nova tecnologia no

universo das salas de aula de língua portuguesa e de língua estrangeira. As vantagens relacionadas à inserção do corpus na sala de aula foram mencionadas na seção 2.2., e, nesse momento, a atenção se volta para a necessidade de um redimensionamento dos recursos metodológicos usados atualmente. O primeiro problema está ligado à rejeição dos professores frente o uso das máquinas, talvez pela impossibilidade de integrá-las a um projeto de trabalho mais produtivo. O segundo problema é a dificuldade de estabelecer ligações entre especialistas em computação e educadores: um projeto interdisciplinar seria desejável.

Algumas questões de metodologia de ensino já bastante conhecidas e discutidas por educadores, porém ainda desconhecidas dos especialistas em *software*, precisam ser levadas em conta na preparação de programas voltados à inserção de computadores no planejamento pedagógico. Conforme aponta Mohan (1992), o desenvolvimento de tecnologia computacional não tem sido seguido pelo desenvolvimento de usos educacionalmente adequados do computador. Talvez, por razões meramente comerciais, alguns programas educacionais acabam repetindo o que o livro didático já apresenta: exercícios já conhecidos. A concepção de atividades que estão limitadas ao certo-errado-tente-de-novo reproduz o padrão já conhecido da gramática tradicional, isto é, a ênfase no conhecimento de regras gramaticais e correção de “erros”.

As críticas à gramática tradicional, aliada aos métodos de ensino de língua estrangeira representam uma resposta aos problemas gerados pelo ensino calcado numa base pré-determinada de sentenças corretas. A substituição da noção de correto pela ausência de parâmetros “que possam permitir uma especificação de objetivos a serem atingidos gera também várias dificuldades. A atribuição de uma prioridade para as capacidades comunicativas pode se tornar uma definição excessivamente vaga dos objetivos de ensino”, segundo Rocha (2001). Conforme foi explicitado ao longo da seção 2.2., o uso de corpora computadorizados, especialmente sob a forma de concordâncias, foi apresentado como um recurso interessante, motivador e desafiador para investigar o uso de língua, tanto no seu aspecto de gramática quanto no seu aspecto de eficácia comunicativa.

A utilização sistemática das concordâncias foi apresentada na seção 2.2. como uma possibilidade de aproximar a tecnologia computacional com o ensino de língua (L1 e L2),

sem reproduzir os velhos e conhecidos exercícios tradicionais transpostos para uma tela de computador. A pesquisa como um todo aponta para a possibilidade de investigação na língua, através da inserção da metodologia de corpus, e o ganho maior foi possibilitar aos alunos a possibilidade de pesquisar sobre a língua, (ver Sardinha, 1999 para detalhes).

Pesquisando sobre a língua através do corpus computadorizado, demonstrou, ao longo do experimento, como as atividades de língua portuguesa podem se tornar mais interessantes na escola. A inserção da metodologia descrita acima revela, ainda, um outro problema a ser solucionado, que espero seja a curto prazo, isto é, as máquinas estão presentes na escola, porém seu uso ainda não foi otimizado. Revela-se uma outra faceta da presença do computador na escola: a chegada destas máquinas, muitas vezes, antecede ao planejamento necessário para tornar viável e executável qualquer atividade que transpusesse a falta de capacitação técnica dos envolvidos. Além disso, o “medo” da máquina também existe e muitas vezes encoberto por “eu não entendo de computadores”.

Para que esse problema fosse sanado e a presença dos computadores estivesse inserida num planejamento de atividades rotineiras, precisaríamos contar com a consciência do professor da necessidade de seu processo contínuo de formação e também com a necessidade de interação humano-máquina (IHM). Nesse sentido, o experimento foi duplamente beneficiado: o professor de sala (do ano de 1999) passou a ser o pesquisador (em 2001) que idealizou o projeto de pesquisa. As condições pessoais de letramento eletrônico desse pesquisador não foram impedimento para levar o experimento ao seu término, isto é, não é preciso ser uma especialista no assunto de programação ou computação para não inserir a metodologia de base em corpus na escola. Saltou aos olhos do pesquisador como os alunos se comportaram perante a máquina e de que forma a IHM se manifestou. Maior detalhamento está descrito no capítulo *Pesquisa*.

A seção seguinte procura situar o leitor sobre o contexto escolar em que o experimento aconteceu, por isso uma resenha do trabalho proposto por Vygotsky. A escola em questão tem nesse teórico um norteador de sua prática pedagógica. Procurou-se investigar de que forma é possível promover interação professor-aluno, aluno-aluno, aluno-máquina, dentro da proposta de trabalho dessa escola e da proposta da pesquisa. Como já foi mencionado na *Introdução*, investigava-se o uso dos operadores argumentativos, a metodologia de base

em corpus foi escolhida e aplicada. Foi feita uma comparação do corpus com os exemplos presentes na gramática tradicional. Tendo em vista os aspectos ligados à argumentação marcados pelo operador **mas**, através do trabalho de Ducrot, a pesquisa procura demonstrar a importância e eficácia da inserção da metodologia com base em corpus como possibilidade de aceitar ou refutar as propostas feitas por aquele teórico.

## **2.4. ABORDAGEM HISTÓRICO-CULTURAL DE LINGUAGEM**

Em caráter introdutório, os principais aspectos da teoria desenvolvida por Lev. Semenovich Vygotsky (1984, 1995) estão apresentados abaixo. Esse quadro teórico-filosófico apresenta-se com uma base sócio-histórica (ou histórico-cultural). Segundo aponta os Parâmetros Curriculares de Santa Catarina de Língua Portuguesa para o ensino médio (1999, p. 55), “se a linguagem humana pode ser encarada como um fenômeno psicológico e como mediadora da formação do pensamento em suas funções mais complexas, seu funcionamento social mostra-a antes de tudo como objeto que possibilita a interação humana [...], e este seu caráter é fundamental na constituição do próprio pensamento e da consciência “.

Remeter à relação entre indivíduo e cultura parece um caminho para pensar a questão das relações presentes no trabalho proposto por Vygotsky: a cultura não é pensada como um dado, um sistema estático ao qual o indivíduo se submete, mas é um meio de negociações em que os envolvidos estão em processo contínuo de recriação e reinterpretação de informações, conceitos e significados. É nos estudos desse teórico que devemos a fundação de uma nova psicologia a substituir a psicologia introspectiva da consciência individual de até então. Esta nova orientação devia estabelecer suas bases na filosofia do materialismo histórico. Vygotsky admitia a existência de uma base reflexa no comportamento dos homens e animais, mas desejava encontrar a especificidade dos processos humanos, recusando reduzir o comportamento humano a uma cadeia de reflexos.

A teoria histórico-cultural proposta por Vygotsky procura abordar as funções psicológicas superiores: memória voluntária, imaginação criativa e solução de problemas

abstratos. As pesquisas levaram a formulação central da teoria que se refere aos processos psicológicos mencionados acima como tipicamente humanos e são considerados superiores porque referem-se a intenções, ações controladas conscientemente, processos voluntários que tornam independente o sujeito em relação ao momento e espaço presente. Por isso Vygotsky refutava reduzir o comportamento humano a reflexos, tal como acontecia com os estudiosos do *behaviorismo*. Uma das preocupações de Vygostky era buscar um enfoque adequado para abordar as funções superiores. Pesquisando essas funções, Vygotsky hipotetizou o conceito de **mediação**.

O conceito de **mediação** está voltado para os processos de desenvolvimento mental da criança e , associado à linguagem , ressalta o processo sócio-histórico. Desta forma, o sentido que a linguagem opera sobre o sujeito fornece um retrato da história de sua sociedade. Estendendo os estudos de Vygotsky, Luria ( 1999) mostra, através de pesquisas experimentais , que a estrutura do pensamento depende de como se organizam as formas de atividades dominantes em culturas diferenciadas. Dessa forma, Luria apontava que os processos cognitivos tinham raízes sócio-históricas: a formação da consciência humana se vincula diretamente às práticas e às formas de cultura existentes; o sistema simbólico que estabelece a formação da mente e da consciência é a linguagem verbal. Consequentemente, a importância do papel dos adultos na aprendizagem é bastante grande, isto é, a linguagem usada por eles e eles próprios são elementos mediadores.

Dirigindo à questão das funções psicológicas superiores do homem, Vygotsky trabalha com o conceito de mediação na relação homem/mundo e com o papel fundamental do contexto cultural na construção do funcionamento psicológico dos indivíduos. Em outras palavras, o fator social em si: o homem inserido na cultura e dela decorrente. A mediação simbólica, a linguagem e o papel fundamental do outro como ser social. Desta forma estabelece-se como revelador o papel do professor como agente social porque é através dele que os alunos estabelecem comportamentos, reavaliam o comportamento do outro e o seu próprio, repensam relações entre objetos, desenvolvem novas respostas categoriais e emocionais, aprendem a generalizar.

Os processos de memória, imaginação, planejamento, inteligência originam-se pelo viés das relações sociais, isto é, entre sujeitos e desenvolvidos ao longo do processo de

internalização de formas culturais de comportamento. A complexidade da apropriação de instrumentos culturais e a interiorização de operações sociológicas surgem em dois níveis que não são excludentes: nível interpsicológico (social) e nível intrapsicológico (interior do sujeito). A natureza dos processos psicológicos superiores aponta que a internalização de formas de conduta implica a operação com signos e, revela o epicentro do pensamento vigotskyano. A dicotomia interno/externo é desvendada através da descrição da natureza dos processos de interiorização e, conseqüentemente, a reconstrução interna desses mesmos processos. A linguagem em sua natureza constitui-se no elemento de **mediação** que possui um lugar privilegiado na interiorização dos processos psicológicos superiores.

A partir dos elementos expostos acima: a relação entre indivíduos e cultura; a trajetória de vida de cada indivíduo; a natureza das funções psicológicas superiores, podemos dizer que, para Vygotsky, a estreita associação entre sujeito psicológico e contexto cultural não implica determinismo. Cada indivíduo é único e por processos psicológicos mais sofisticados (que envolvem consciência, vontade e intenção), constrói seus significados e recria sua própria cultura. Vygotsky aponta que o indivíduo interioriza formas de funcionamento psicológico dadas culturalmente, mas ao tomar posse delas, torna-as suas e as utiliza como instrumentos pessoais de pensamento e ação no mundo.

A teoria histórico-cultural marca o papel constitutivo e estruturante da linguagem em relação aos processos cognitivos. Assim, não há possibilidade de conteúdos cognitivos fora da linguagem, nem fora dos processos interativos humanos, segundo Morato (1996). Dessa forma, o conhecimento é concebido como produção simbólica e material que tem lugar na dinâmica interativa, já que é através do outro que o sujeito estabelece relações com o conhecimento, e deve ser entendido no espaço da intersubjetividade. Smolka (1995) aponta que os processos de funcionamento mental são mediados por signos que emergem em um terreno interindividual, em que se atribui à palavra o signo por excelência. O sujeito não se constitui somente pelos processos de maturação orgânica, mas através de interações sociais, troca com seus semelhantes.

Dessa forma, funções psíquicas humanas estão estreitamente ligadas ao aprendizado por intermédio da linguagem, do produto cultural do grupo. Esse material incorpora valores, conhecimentos, diferentes sistemas de representação, formas construídas ao longo

do percurso da humanidade. O conceito de mediação proposto por Vygotsky incide no domínio desse material, cujo papel desenvolve-se através das interações sociais. O desenvolvimento de funções psicológicas superiores e de novas atividades mentais acontece pelo viés de um processo ativo/ interativo, de apropriação. A mediação pelo outro e pelo signo caracteriza a atividade cognitiva. Podemos, talvez, estabelecer relação entre a internalização das funções psicológicas superiores e o desenvolvimento da linguagem. Segundo aponta Smolka ( 1997), a linguagem como signo socializado é mediadora do processo de internalização das funções desenvolvidas socialmente.

A teoria histórico-cultural hipotetizada por Vygotsky aponta para o aspecto da inter-relação entre desenvolvimento e aprendizagem, discutindo a apropriação do conhecimento. Vygotsky define o conceito de zona de desenvolvimento proximal (ZDP) : o nível de desenvolvimento potencial - a capacidade de resolução de um problema via orientação de um adulto mais capaz, para além das etapas de desenvolvimento já alcançadas, isto é , o nível de desenvolvimento real. A partir, então, da existência desses dois níveis de conhecimento, real e potencial , toma forma a zona de desenvolvimento proximal como um domínio psicológico de constante transformação. Através desta concepção, o desenvolvimento surge num *continuum*. Esta pesquisa demonstra como a ZDP se realizou.

O centro da teoria vigotskiana aponta para a atividade instrumental e a interação como unidade de análise. Estabelecendo estreitas relações com instrumentos físicos, a teoria atribui papel de *ferramenta psicológica* aos sistemas de signos, especialmente a linguagem . Mesmo assim, a linguagem tem efeito na formação subjetiva e desenvolvimento cognitivo individual. Em outras palavras, a linguagem está presente na formação e funcionamento de todos os processos psíquicos. Além disso, a linguagem está ligada ao pensamento e juntos constroem a significação da realidade. Coexistindo, então, a linguagem dá forma e direciona o pensamento. A linguagem é, portanto, uma articulação de um sistema de signos mediador da ação humana.

O pensamento verbal, para Vygotsky, constitui o conjunto de pensamento e linguagem onde as relações entre ambos assumem uma complexidade a ser melhor entendida. Para analisar o pensamento verbal, Vygotsky utiliza a palavra como unidade de análise. Acredita que o *significado* da palavra constitui a unidade do pensamento e linguagem e não admite

decomposição: “Encontramos essa unidade, que reflete a união entre pensamento e linguagem, na forma mais simples, no significado da palavra.. é a unidade.. que representa : um fenômeno de linguagem ou de pensamento “. O significado é um aspecto da palavra, pois expressa um conceito ou uma generalização. O significado da palavra é ao mesmo tempo um fenômeno verbal e intelectual.

A postulação vigotskiana aponta a existência de três distintas realizações no desenvolvimento das palavras: pensamento sincrético, pensamento por complexo e pensamento conceitual. Todas as realizações são níveis de desenvolvimento do significado das palavras. Não devem ser confundidas com fases de desenvolvimento do sujeito e sim como recorte metodológico para distinguir a formação de conceitos. Dessa forma, o significado da palavra constitui um processo em desenvolvimento: desenvolvimento de um conceito, de um significado ligado a palavra que apenas começa por esse viés onde a construção do significado ou conceituação mais elementar é denominado pensamento sincrético. Na outra extremidade estão formas de conceituação mais avançadas.

A condição da palavra como meio para formação de conceitos constitui o desenvolvimento do pensamento verbal. Tunes (1995) apresenta a distinção no nível do significado da palavra, isto é, distingue conceitos cotidianos dos conceitos científicos. Os primeiros dizem respeito às relações das palavras com os objetos a que se referem, os conceitos científicos, por outro lado, dizem respeito às relações das palavras com outras palavras. A natureza dos conceitos científicos modifica a estrutura funcional da consciência, permitindo níveis mais altos de discernimento, organização e consciência sobre o ato de pensar. Dessa forma, as linhas de desenvolvimento não evoluem em paralelo, mas em relações mútuas, de tal modo que os conceitos científicos se movem em direção aos fenômenos concretos que representam, e os cotidianos movem-se em direção à abstração.

Dessa forma, a construção do conhecimento assume o pressuposto de que os processos psicológicos estão relacionados aos modos de vida de interação dos sujeitos e a linguagem define-se como um fenômeno de natureza social. Usamos a linguagem para nos expressar, estudar, trabalhar, trocar idéias, dentre outras atividades e, assim, passamos a constituir os processos cognitivos e de conhecimento. Muito mais do que abordagem, a linguagem é constitutiva dos processos cognitivos e de conhecimento. A apropriação social da

linguagem é a condição fundamental do desenvolvimento mental e o que permite concebê-la como condição de cognição. Nos estudos relacionados à linguagem, Smolka (1997, p.38) aponta que “a linguagem não é só meio e modo de (inter/oper) ação, é também produto histórico, objetivado; é constitutiva/constituidora do homem como sujeito. “

O modelo teórico proposto por Vygotsky apresenta, no campo da educação, implicações de natureza prática que podem mudar os processos de construção de conhecimento e a concepção de linguagem. Sob esse aspecto destacam-se os conceitos sobre signos, enunciação e a participação do outro nas relações com a linguagem. ( ver Bahktin, 1990 para maiores detalhes) . No que diz respeito ao experimento, observou-se como a ZDP proposta por Vygotsky e a IHM ocorreram . O papel do aprendiz mais hábil também foi observado. Pode-se apontar um novo encaminhamento de prática pedagógica, baseado na metodologia de base em corpus, que no caso deste experimento, tornou desafiadora a inclusão de atividades relacionadas com o estudo da linguagem : operadores argumentativos . A próxima seção apresenta sucintamente esse tema.

## 2.5. ARGUMENTAÇÃO

A inserção da resenha sobre operadores argumentativos : definição , função, exemplos deve-se ao fato do objeto de pesquisa ser o operador argumentativo **mas**. Porém, a presença desse operador, ou outro qualquer, dentro de uma abordagem com base em corpus, poderia ser conflitante. A concepção de investigação lingüística da abordagem com base em corpus, conforme já foi mencionado na seção 2.2., parte de exemplos autênticos de uso de língua, quer sejam de língua falada ou de língua escrita. Os procedimentos metodológicos de análise obedecem a determinados paradigmas, os dados são exaustivamente trabalhados até que se possam extrair algumas indicações e evidências, por exemplo, qual a colocação mais freqüente a direita de *diferente*, em língua portuguesa? Esse não é o caso dos operadores argumentativos , pelo viés do trabalho de seu criador- O. Ducrot.

A proposta de análise da Semântica Argumentativa é bem outra. Tomando exemplos, como ponto de partida, o direcionamento é outro. Os exemplos são fruto de atividade criativa do autor, professor. A concepção de investigação é direcionada para a indicação da força dos argumentos dentro da sentença. Maiores detalhes estão abaixo explicitados. Resta, porém, confrontar uma possibilidade de confirmação ou refutação dos postulados de Ducrot, fazendo uso da metodologia de base em corpus. É esse o objetivo primeiro desse experimento. O segundo objetivo está ligado à inserção do uso de computadores, conseqüentemente, em sala de aula. Da união desses dois objetivos gerais, o experimento demonstrou que é possível promover este confronto. A seção seguinte resume o postulado de Ducrot (1981).

#### 2.5.1. A argumentação: leis internas e operadores argumentativos

O termo *operadores argumentativos* foi criado por O. Ducrot, teórico francês criador da *Semântica Argumentativa*, para nomear os elementos lingüísticos que têm como função indicar a força dos argumentos dentro dos enunciados e o sentido para o qual apontam. Nesse sentido, Ducrot vale-se de duas noções complementares, isto é, a *noção de escala argumentativa* e *classe argumentativa*.

o valor argumentativo de uma frase não é somente uma conseqüência das informações por ela trazidas, mas a frase pode comportar diversos morfemas, expressões ou termos que além de seu conteúdo informativo, servem para dar uma orientação argumentativa ao enunciado, a conduzir o destinatário para esta ou aquela direção. (Ducrot, 1981, p.178)

Uma classe argumentativa é constituída de um conjunto de enunciados que podem servir igualmente de argumento para uma mesma conclusão (convencionalmente designada por R). Tomando como exemplo a sentença abaixo:

(1) Pedro é o melhor jogador de xadrez. (conclusão R)

argumento 1 - tem boa formação técnica

argumento 2 - tem experiência como jogador dessa modalidade

argumento 3 - respeita os adversários

argumento 4 - tem ganho todas as últimas partidas do campeonato

Os argumentos apresentados acima têm o mesmo peso para levar o alocutário a concluir R, portanto pertencem a mesma *classe argumentativa*. Ao passo que quando dois ou mais enunciados de uma classe se apresentam em gradação de força crescente no sentido da mesma conclusão, tem-se uma *escala argumentativa*, como no exemplo abaixo:

(2) A festa de aniversário foi maravilhosa. ( conclusão R)

argumento 1 - todos os amigos estavam presentes

argumento 2 - todos os presentes se divertiram muito

argumento 3 - meu filho estava muito feliz.

Costuma-se representar graficamente a escala argumentativa da seguinte forma:

R: A festa de aniversário foi maravilhosa

( arg. + forte) p'' - todos os amigos estavam presentes

p' - todos os presentes se divertiram muito

p - meu filho estava muito feliz.

Caso a mesma conclusão ( R) for *negada*, invertem-se os elementos da escala.

R: A festa de aniversário não foi maravilhosa.

( arg. + forte) p” - todos os amigos não estavam presentes

p’ - todos os presentes não se divertiram muito

p - meu filho não estava muito feliz.

Portanto *escala argumentativa* é uma classe que comporta uma relação de ordem. (ver Koch, 1984, para maiores detalhes.)

Tomando como exemplo a sentença abaixo, teremos:

(R) A.B. é um político ambicioso.

Ele quer ser { até mesmo presidente ( p”)

{ pelo menos governador ( p’)

{ no mínimo prefeito ( p)

Conforme os exemplos acima demonstram, os *operadores argumentativos* ( até, pelo menos, no mínimo) estabelecem uma hierarquia dos elementos argumentativos numa escala crescente ou decrescente, no sentido de uma conclusão. Dessa forma os conteúdos p, p’, p” não importam, mas o fato de virem **marcados** pelos *operadores argumentativos* que fornecem tanto a posição na escala, quanto o sentido. Isto é, a orientação argumentativa está marcada, como uma regularidade enunciativa, no enunciado. Dessa forma, quando se têm duas escalas orientadas para a mesma conclusão, o encadeamento dos elementos se faz por operadores como **e, nem, tanto... como, não só... mas, também** etc, conforme está explicitado na seção seguinte.

### 2.5.2. O operador argumentativo **mas**

Cada língua apresenta em sua gramática uma série de morfemas responsáveis pelo encadeamento dos enunciados, determinando sua orientação discursiva. Esses operadores têm como função estruturar um discurso, enquanto texto; esta estruturação se faz pelo valor argumentativo desses operadores de discurso. Ducrot (1972) considera o operador *mas* como argumentativo por excelência. O uso de operadores como *mas*, *porém*, *embora*, *contudo*, *não obstante...*, depende da estratégia empregada pelo locutor. A estratégia do operador *mas* é a de frustrar a expectativa que se criou no destinatário, portanto teremos operadores que contrapõem argumentos orientados para conclusões contrárias: *mas* ( *porém*, *contudo*, *todavia*, *no entanto*, etc. ), *embora* ( *ainda que*, *posto que*, *apesar de* (que) etc.) cujo funcionamento, segundo Ducrot (1972), aponta que o locutor estrutura o discurso da seguinte forma:

- (a) introduz em seu discurso um *argumento possível* para uma conclusão R, logo em seguida
- (b) opõe-lhe um *argumento decisivo* para a conclusão contrária *não-R* (  $\sim R$  ).

Porém, ao invertermos a posição dos argumentos - A *embora* B - , essa antecipação já não ocorre, pois A é o argumento que deverá prevalecer, uma vez que *embora* sempre nega **argumentativamente** o enunciado em que aparece. No caso de A *mas* B, nada previne o locutor de que se pretende apresentar, por meio de B, um argumento mais forte a favor de r, do que aquele introduzido por A, de modo que ele é surpreendido quando isso acontece. Entretanto, essa estratégia pode funcionar para certos enunciados, já no exemplo abaixo:

- (1) Moro longe, *mas* tenho carro
- (p)            (p')

Nesse caso, não há oposição entre as duas proposições ligadas pelo *mas*. A conjunção *mas* funciona como um elemento que não permite o interlocutor de tirar a conclusão R ( é difícil de ir lá, é trabalhoso chegar lá, toma muito tempo chegar lá) a que ele é levado pela proposição (p) “ Moro longe ”. No exemplo abaixo temos uma outra situação:

(2) Venha a minha casa, mas chegue logo.

(p)                      (p’)

Na sentença acima (p) é uma concessão, ao passo que (p’) é uma ordem. “Venha” é um argumento para a concessão; “mas chegue logo” é um argumento para uma exigência.

O. Ducrot (1972) aponta para o fato dessa forma:

“ A expressão “ p mas q ” pressupõe que p pode servir de argumento para uma certa conclusão r, e que a proposição q é um argumento que anula esta conclusão. O movimento de pensamento implicado por uma frase afirmativa do tipo p mas q poderia ser parafraseado assim: sim, p é verdadeiro, tu terias tendências a dela concluir r mas, não deve, porque q pode ser apresentado como um argumento mais forte para não - R do que p para r .”

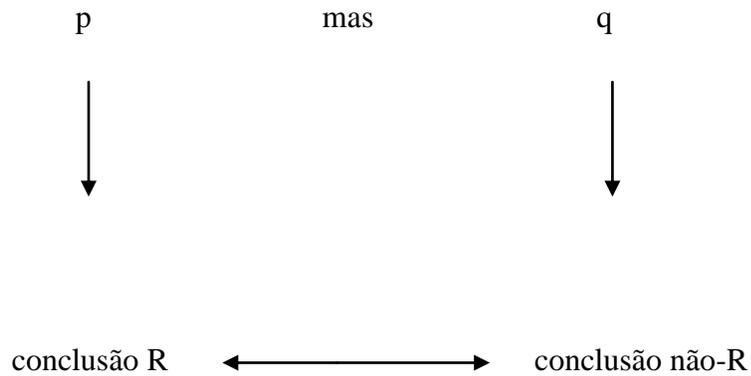
Isso significa que ao unirem-se dois elementos semânticos como **p** e **q** por meio do *mas* temos as seguintes evidências:

- (a) existe uma conclusão R facilmente encontrada pelo destinatário, sugerida por p mas não confirmada por q, isto é, p e q são opostos em relação a R;
- (b) a força de q, contrária a R, é maior que a força de p a seu favor, o que produz um conjunto p mas q seja orientado no sentido de não - R ( $\sim R$ ).

Nesse ponto, O. Ducrot (1989) introduz um aspecto importante na relação argumentativa, isto é, o argumento não-r era considerado **mais forte** do que r, conforme mencionado acima. Segundo o teórico temos:

“ De fato, a única coisa constante é que o locutor declara negligenciar o primeiro (enunciado) da argumentação que está construindo, para apoiar-se apenas no segundo - a força argumentativa superior atribuída a este não passa de uma justificação desta decisão. “( Ducrot, apud Maingueneau, 1989, p. 166-7)

Esquematisando esse funcionamento temos:



onde lemos ser igual a “ ser um argumento menos forte “  
 “ ser um argumento a favor de “  
 “ ser contraditório em relação a “

Esse esquema pode ser percorrido de duas formas, isto é, *mas* introduz q ou não-R.

Portanto, podemos ter as seguintes realizações :

- (1) Ele é político, mas honesto ( p mas q)
- (2) Ele é político, mas é possível confiar nele. ( p mas não-R)

Dessa forma, a conclusão é dada, sendo necessário encontrar o argumento mais forte q. Assim, na sentença (2) temos “ele é político (p), mas é possível confiar nele ( não-R); ele é honesto (q)“, na sentença (1). Vemos que o primeiro percurso de *mas* estabelece uma relação indireta entre p e não-R, enquanto o segundo o faz diretamente.

Ducrot e Vogt ( 1972) remetem à origem da conjunção *mas*, isto é, *mas* não derivou do advérbio latino *sed* , e sim do advérbio *magis*. *Mas* (SN) corresponde ao alemão *sodern* e ao espanhol *sino*, possuindo valor pragmático de refutação, retificação ou justificativa de uma recusa de p, que segue sempre uma proposição negativa: Neg p mas q que pode substituída por ao contrário. O *mas* (PA) não demanda uma proposição precedente negativa. Sua função é introduzir uma proposição que orienta para uma conclusão não-R oposta a uma conclusão R, independente da forma negativa. Os exemplos abaixo ilustram a visão do teórico francês:

- (1) Ele não é inteligente, *mas* apenas esperto. ( *mas* - SN)
- (2) Ele é inteligente, *mas* estuda pouco. ( *mas*- PA)

É importante salientar que a **negação** em semântica argumentativa é diferente do tratamento da **negação lógica**. Os exemplos citados remetem às duas funções do *mas* – a representação do discurso do outro no discurso de quem dirá *mas*, e como traço específico o modo desta representação em cada um dos casos desta forma as duas funções são atravessadas pela intertextualidade. Portanto , com o *mas* (SN) , o discurso do outro se realiza como objeto de recusa por parte de quem o profere , porém com o *mas* (PA), surge como objeto de um acordo a ser desfeito logo em seguida, uma vez que atingida a argumentação contra as conclusões que poderiam ser tiradas do discurso do destinatário. Conforme foi exposto anteriormente, vemos que as duas funções do *mas* realizam dois modos de dizer **não** , e é nesse aspecto que reside o caráter pragmático e ideológico do operador em questão.

Encontramos na gramática tradicional passagens relacionadas ao operador *mas*. Essa investigação, comparação com Ducrot e as atividades relacionadas à abordagem de base em corpus estão detalhadas no capítulo *Pesquisa* . O próximo capítulo intitulado *Metodologia* aborda questões da concepção da pesquisa , os sujeitos da pesquisa e a coleta de dados. Há uma subseção relatando as atividades feita com esse grupo de sujeitos durante o ano de 1999. Os resultados apresentados indicavam a necessidade de uma investigação lingüística

que promovesse aprendizado através de uma nova tecnologia, uma abordagem nova. A lingüística de corpus era desconhecida do professor de sala. Esse mesmo professor, hoje pesquisador e idealizador desse experimento, teve a oportunidade de trabalhar com esse mesmo grupo de alunos no ano de 2001. A descrição do que ocorreu é foco dos capítulos seguintes.

## **CAPÍTULO 3 - METODOLOGIA**

### **3.1. CONCEPÇÃO DA PESQUISA**

A pesquisa foi concebida através da necessidade de inserção de uma nova tecnologia associada à prática e cotidiano na sala de aula. A busca levou à Lingüística de Corpus, à necessidade de transpor metodologias antigas para o meio eletrônico, à opção que as concordâncias representam. Porém, a análise semântica das ocorrências em seus contextos reais evidencia a dificuldade de aplicar isto ao *mas*, já que marca relações textuais lógicas que não podem ser facilmente tratadas no âmbito de uma noção mais simples de colocação em termos de contexto imediato. Buscava-se promover aprendizado através de uma nova tecnologia, uma possibilidade de inserção de uma nova prática pedagógica no universo escolar. Conforme mencionado na *Introdução* é necessário uma revisão de postura que envolve professor e alunos, para além de uma atividade “nova”.

Optou-se por investigar a embricação leitura/escritura focando a construção de textos argumentativos e utilizando a metodologia de corpus, opção de concordâncias para focar o operador argumentativo *mas*. Os alunos trabalharam com o corpus do NILC, de onde foram selecionados textos da *Folha de São Paulo (1995)*, que foi editada na versão texto. Um total de 24 textos foram selecionados que abrangiam tópicos relacionados à notícias sobre Brasil (Br), Cotidiano (Co) e Opinião (Op). Os parênteses referem-se aos nomes de cada arquivo. Posteriormente, os textos selecionados foram gravados em disquete. O concordanceador foi utilizado para direcionar a seleção das ocorrências que seriam trabalhadas em sala. O laboratório de informática da escola foi utilizado e, para o total de 19 sujeitos pesquisados, havia a possibilidade de cada um fazer uso de um computador.

### **3.2. OBJETIVOS DA PESQUISA E HIPÓTESES DE TRABALHO**

Para direcionar essa pesquisa estabeleceram-se os seguintes objetivos:

### 3.2.1. Objetivos gerais

1. propor um ensino de operadores argumentativos , focando o *mas*, baseado na metodologia de lingüística de corpus;
2. verificar se um ensino específico sobre operadores argumentativos resulta num melhor desempenho de habilidades lingüísticas e metalingüísticas: reconhecimento e uso dessa função lingüística.

### 3.2.2. Objetivo específico

1. experienciar a presença da abordagem com base em corpus fundamentada nas operações metalingüísticas, focando o operador argumentativo *mas*

### 3.2.3. Hipóteses de trabalho

Para direcionar a realização dos objetivos , propõe-se investigar as seguintes hipóteses:

- 1) O ensino específico dos operadores argumentativos melhora a cognição quanto a reconhecer e usar as nuances entre os operadores estudados;
- 2) O ensino específico do operador *mas* provoca uma melhoria na capacidade dos aprendizes produzirem sentenças explorando essas diferenças e produz efeitos positivos sobre as habilidades testadas;
- 3) O ensino específico do operador *mas*, utilizando computador e abordagem com base em corpus, melhora a metacognição dos sujeitos a respeito de reconhecer e aplicar determinados usos do operador *mas*

## 3.3. O Contexto da pesquisa ( sem o corpus)

Conforme foi mencionado na página 46, o professor de língua portuguesa (1999) desconhecia a lingüística de corpus. Para situar o leitor sobre o contexto de atividades na qual a presente pesquisa se encaixa, as atividades descritas abaixo ocorreram em 1999 com um grupo de alunos do 1º. ano do ensino médio de uma escola particular de Florianópolis. O grupo era composto por 28 alunos: 15 do sexo feminino e 13 do sexo masculino. Todos os alunos são de classe média e cursaram o ensino fundamental em escolas da rede particular. A idade média do grupo era de 15 anos. As atividades de leitura, que envolviam três gêneros textuais diferentes, deveriam guiar os aprendizes através da embricação leitura/escritura. O professor de língua portuguesa, apesar dos bons resultados apresentados via produções textuais, sentiu que algo mais poderia ter sido feito. Algo novo e desafiador que promovesse a pesquisa sobre operadores argumentativos. Porém não sabia o que. Seguem abaixo os detalhes das tarefas planejadas com o professor de geografia, envolvido no processo, e executadas pelos alunos no ano de 1999 com duração prevista de um bimestre.

### **3.3.1 Os instrumentos utilizados**

Inicialmente, os professores de língua portuguesa e geografia buscaram o fio temático das atividades que dariam suporte para o trabalho de construção da noção de paisagem e espaço, isto é, a definição dos termos, a mão humana agindo sobre a natureza, interferindo no seu equilíbrio, etc. Definiram que o trabalho com textos escritos deveria determinar a percepção da argumentação construída via textos informativos e científicos. Os professores preocupavam-se em fazer com que os alunos soubessem que as atividades subsidiariam a resenha crítica - produto final das tarefas daquele bimestre.

As atividades foram iniciadas com a apresentação de um texto sob o título “Florianópolis , aqui se vive melhor “ publicada pela revista *Veja* da editora Abril em 07 de abril de 1999, seguida do texto “ Paisagem e espaço “ de Milton Santos publicado em

*Metamorfoses do Espaço Habitado* - Hucitec - S.P., 1988 e, finalmente do texto “Mea Culpa” de Eduardo Galeano, publicado na revista *Nossa América* (no. 2) de 1993. Não se tratava de uma atividade interdisciplinar, pois não havia sido planejada para esse fim, mas de uma comunhão de esforços na embricação leitura/escritura e concebia paisagem pelo viés dos sentidos. Em outras palavras, percebida pela aparência, por isso muitas vezes enganosa. Os recursos lógicos e lingüísticos de cada texto deveriam guiar os aprendizes a perceber os procedimentos argumentativos. Portanto, a argumentação foi o ponto de partida.

### **3.3.2 As instruções**

Para realizar o trabalho foram delineadas as seguintes tarefas e dadas as seguintes instruções:

1. Os alunos foram informados de que realizariam tarefas relacionadas à leitura e escritura, e o produto final seria uma resenha crítica dirigida a um leitor universal e publicada num jornal de circulação nacional, e que, portanto, deveria informar o leitor sobre os aspectos discutidos pelos textos lidos e ampliados em sala de aula. A primeira leitura, “Aqui se vive melhor”, publicada pela revista *Veja*, deu início à atividade de fichamento, isto é, fazer uso de palavras-chave para delinear as idéias apresentadas pelo autor. Notaram que o autor apresentava uma visão fragmentada da cidade e não mencionava aspectos negativos. Alguma discussão oral em torno dos prós e contras de uma divulgação desta natureza em uma revista de circulação nacional foram pontuados. Levantaram a hipótese de um fluxo migratório direcionado para Florianópolis: o paraíso perdido. Em seguida, sob orientação do professor de língua portuguesa, utilizaram o fichamento para a construção de um resumo hipoteticamente a ser publicado em um jornal local. Este resumo seria lido por um leitor universal e deveria ser um retrato da visão do jornalista envolvido. Os aprendizes tiveram dificuldade de se limitarem à visão do jornalista.

2. Remetendo ao trabalho já realizado em língua portuguesa, o professor de geografia iniciou as atividades de discussão dos conceitos de paisagem e espaço valendo-se do fluxo de conhecimento do velho para o novo, ampliando essa visão fragmentada de Florianópolis, através do texto científico “Paisagem e espaço” de Milton Santos. Dessa forma, ampliou-se a percepção de espaço urbano e ocupação do solo. Um resumo escrito contendo as idéias principais apresentadas pelo autor foi produzido. Os aprendizes fizeram uso de canetas marcadoras de textos para sublinhar as idéias principais apresentadas pelo autor. Houve alguma dificuldade em relação aos argumentos apresentados pelo escritor, pois o vocabulário técnico e o gênero de construção textual foram itens complicadores da tarefa. Uma saída de campo - passeio de estudos - ao norte da ilha de Florianópolis foi realizada. Os registros fotográficos compuseram um painel sobre o mau uso do solo e a conseqüente urbanização daquelas praias: Ingleses, Santinho e Canasvieiras. A falta de planejamento urbano naquela região estava expressa sob a forma de legenda em cada uma das fotos que foram expostas em sala de aula. Observava-se uma forte influência das idéias apresentadas pelo autor do texto “Paisagem e espaço”. Os aprendizes começaram a manifestar uma maior conscientização sobre os problemas da cidade onde moravam. À medida que os grupos expunham os trabalhos no painel, alguns alunos “tentavam explicar” a escolha do operador argumentativo presente naquela legenda.

3. O terceiro texto “Mea Culpa” de Eduardo Galeano é uma crônica e foi a terceira e última atividade de leitura feita pelo grupo. O professor de geografia utilizou a técnica de perguntas e respostas, isto é, a idéia geral do texto e localização de informações: scanning e skimming, respectivamente. Os alunos começaram de forma espontânea a estabelecer relações com as informações dos textos já lidos. O professor envolvido nessa atividade procurava explicitar as diferentes formas de pensar sobre o tema proposto e de que forma os diferentes autores as escreveram.

4. Ao professor de língua portuguesa coube, então, a atividade de orientação da resenha crítica. Os comandos foram claros, dirigidos ao processo de construção e uso dos operadores argumentativos na produção textual. As idéias de cada autor precisariam ser

respeitadas. A forma de iniciar a resenha crítica, como introduzir opiniões de autores diferentes e fazer uma crítica ao texto da revista *Veja*, subsidiados pelo conhecimento geográfico já ampliado, foi construída em sala com os alunos.

### **3.3.3. Metodologia pedagógica**

Para o primeiro texto “Aqui se vive melhor” foi feita uma leitura silenciosa em sala de aula, seguida de atividade de fichamento, atividade de escritura de resumo e refazer. A atividade descrita como refazer refere-se ao ato de repensar o que foi escrito, reler, promover alterações sugeridas pelo professor, correções e devolução ao professor para avaliação final. Para o segundo texto “Paisagem e espaço” foi aplicada a técnica de iluminação de palavras-chave e posterior chamamento de atenção para a construção de argumentos no texto científico. Os aprendizes produziram um resumo, apresentando as idéias gerais e pontos principais apresentados pelo autor. No terceiro texto, “Mea Culpa”, focou-se a atenção para outro tipo de gênero textual e quais operadores argumentativos foram usados e o porquê.

Os conceitos geográficos envolvidos nos dois últimos textos foram devidamente discutidos e trabalhados pelo professor de geografia. Podemos afirmar que o grupo de atividades propiciou uma progressão de exercícios de escritura, isto é, fichamento, legenda, resumo, resenha crítica.

### **3.3.4. Resultados**

Como resultado desta prática pedagógica, foram obtidas 22 produções textuais. Não cabe, nesse momento, uma análise de cada produção textual, porém saltam aos olhos algumas evidências. Os três textos lidos apresentam apenas a ocorrência do operador *mas*. Por alguma razão, os alunos, ao produzirem texto, remeteram a outros operadores. Observou-se, ainda, a quantidade de 9 operadores argumentativos distribuídos nas 22

escrituras, perfazendo uma média de 2 a 3 operadores por produção textual. Os operadores argumentativos mais utilizados foram *mas* (25 ocorrências), *pois* (8 ocorrências), *portanto* (3 ocorrências), seguidos de *ou seja*, *então*, *por isso*, *portanto*, *apesar de*, *porém*, *porque* (1 ocorrência respectivamente). Observou-se, também, outras ocorrências do *mas*: *mas também*, *mas não*, *mas sim*.

### 3.3.5. Avaliação

Os professores de língua portuguesa e geografia perceberam a complexidade e desafios frente às tarefas, já durante o planejamento. Ambos os professores dispunham de três aulas semanais com duração de quarenta e cinco minutos cada. O prazo de execução estava previsto para um bimestre. Os professores observaram que alguns aprendizes tiveram mais dificuldade de usar os operadores argumentativos de forma intencional, de manipulação. Talvez pela diferença entre os leitores mais e menos hábeis da sala. Não é objeto de análise, nesse momento, as razões para esse comportamento, mas as evidências de falta de coesão, por exemplo, nas produções textuais nos levam a acreditar na necessidade de um outro trabalho de manipulação de linguagem seria necessário para atingir esses aprendizes.

Os professores fizeram uso de uma escala numérica de 0 (zero) a 10 (dez) para avaliarem o processamento das operações de embricação leitura/escritura. Os critérios de avaliação foram determinados pelos professores de geografia e língua portuguesa: índices de coerência, uso de operadores argumentativos e aplicação dos conhecimentos geográficos trabalhados foram considerados como nota entre 8 e 10. A variável entre essas duas dimensões deve-se ao fato de apresentarem mais ou menos os três itens mencionados acima, considerados como desejáveis. Abaixo da nota 8, as mesmas variáveis foram observadas. Os itens descritos acima foram considerados norteadores dos critérios de avaliação, valendo três pontos cada um. Decidiu-se incluir, para efeito total e final de nota, os itens capricho e pontualidade na data de entrega, valendo um total de um ponto.

As escrituras faziam parte da nota do bimestre em língua portuguesa e geografia. Portanto, quanto mais se aproximassem dos itens considerados como relevantes, mais o aprendiz se aproximava da nota dez. As escrituras foram lidas e julgadas pelos professores de língua portuguesa e geografia. Não havia obrigatoriedade de consenso na avaliação, porém, os dois professores se aproximaram muito na nota atribuída a cada produção textual. Os alunos foram informados dos critérios de avaliação, antes de produzirem as mesmas. Por parte dos alunos, verificamos que houve um índice bastante elevado de motivação e interesse ao longo das tarefas solicitadas. Por outro lado, restou ao professor de língua portuguesa a sensação de incompletude, somente ler/escrever parecia pouco. Comparar com a gramática tradicional o uso dos operadores poderia ser um passo a mais. A sensação de incompletude só deixou de existir através do experimento com corpus, seção seguinte.

### **3.4. A pesquisa com corpus**

A pesquisa realizou-se no ano de 2001 com um grupo de alunos do terceiro ano do ensino médio de uma escola particular de Florianópolis. Conforme foi mencionado no item 3.1., concepção da pesquisa, a proposta era atingir os mesmos alunos e focar o mesmo item lingüístico. O professor de língua portuguesa da turma estava trabalhando com eles desde o ano 2000. Segundo relato do professor, a produção textual apresentava aspectos de realização da tarefa bastante melhorados, isto é, postura, responsabilidade, tempo de execução, fluidez de linguagem, observados tanto durante o processo quanto em relação ao produto. As observações feitas pelo professor de sala revelam um amadurecimento do grupo frente as tarefas. A pesquisa com corpus foi algo novo para o grupo e para o professor de língua portuguesa, que desconhecia a Lingüística de Corpus. Seguem abaixo detalhes sobre a pesquisa em si.

#### **3.4.1. Os sujeitos da pesquisa**

O grupo de sujeitos era composto por um total de 19 alunos, 8 alunos do sexo feminino e 11 alunos do sexo masculino. A idade média dos alunos era de 17 anos. Todos os sujeitos dessa pesquisa haviam concluído o primeiro e o segundo ano do ensino médio na escola pesquisada.

### **3.4.2. A coleta de dados**

A decisão por usar textos pré-selecionados deveu-se a necessidade de otimizar os recursos de armazenamento das máquinas disponíveis na escola pesquisada. Para manipular os dados, optou-se pelo WordSmith Tools que é um programa de manipulação de corpus, composto por vários instrumentos ou ferramentas de análise dos dados. O Concord (concordanceador) é uma destas ferramentas. Os textos da *Folha de São Paulo* (1995) foram escolhidos devido a representatividade desse jornal de circulação nacional. Um total de 24 textos fez parte do corpus de referência, subdivididos em três áreas distintas: Brasil, Cotidiano, Opinião.

### **3.4.3. Corpus na pesquisa**

Os alunos decidiram participar da experiência de pesquisa. O pesquisador disponibilizou o corpus, através de disquete, e os alunos copiaram a versão texto do corpus para seus documentos. Os alunos foram informados, através de informações escritas e lidas pelo professor de sala, que a pesquisa se iniciava naquele momento, isto é, as informações que faziam parte da pesquisa estavam disponíveis no computador. Havia um computador para cada dois alunos, em média. Procurou-se concentrar os alunos em apenas uma sala de informática para que o professor estivesse disponível para todos simultaneamente e as instruções fossem lidas para todo o grupo. Evitando dispersão de tempo, atenção e observações que o professor pudesse considerar importantes durante a execução da tarefa. O professor informou aos alunos sobre os objetivos gerais da pesquisa e solicitou que os alunos abrissem a pasta com os textos. Foram informados de que a pesquisa focava o

operador argumentativo *mas*. Maior detalhamento sobre a pesquisa : resultados, análise, discussão e perspectivas de aplicação e ampliação da mesma estão descritos no capítulo seguinte - *Pesquisa*.

## CAPÍTULO 4 - A PESQUISA

Para a realização da pesquisa foram utilizadas quatro aulas de quarenta e cinco minutos cada. A manipulação do corpus, investigação lingüística e o contraste com a gramática normativa fizeram uso de três aulas consecutivas em novembro de 2001. A produção textual foi realizada em dezembro de 2001. Havia uma diferença de trinta dias entre os dois momentos da pesquisa. O tema das produções textuais “*Papéis do homem e da mulher na modernidade*” foi escolhido pela escola.

Tomando como ponto de partida a concepção da pesquisa, mencionado anteriormente, as seções seguintes apresentam os resultados obtidos na pesquisa, bem como a análise e discussão dos mesmos. Por fim, a conclusão da pesquisa.

No dia determinado para a realização da primeira parte da pesquisa, os alunos foram orientados a se dirigirem para o laboratório de informática e deveriam ocupar apenas uma das duas salas existentes para esse fim. Cada sujeito usou um computador. As orientações para o direcionamento e execução do trabalho estavam escritas e foram lidas pelo professor de língua portuguesa:

‘ (1) Este é um projeto de pesquisa em Lingüística de Universidade Federal de Santa Catarina e os textos que vão ser lidos e manipulados por vocês investigam o uso do operador argumentativo *mas*. O arquivo (em disquete) que contém os textos deve ser copiado e transferido para a pasta individual de cada um. ‘

Após essa orientação, os alunos começaram a ler os textos que faziam parte do corpus. Fizeram comentários quanto a diferença de tamanho entre os textos. O professor solicitou aos alunos que procurassem sentenças que mostravam o uso do *mas* e anotassem as sentenças. Depois de 3 minutos aproximadamente, fizeram alguns comentários quanto ao tempo total que seria necessário para completar a tarefa. A preocupação era com o tempo a ser gasto. Mesmo assim, o professor solicitou que todos registrassem três sentenças em que o operador *mas* aparecia. Depois de concluída essa etapa, o professor informou que havia

um programa que realizava a tarefa de processar textos e focar a palavra escolhida. As instruções foram as seguintes:

‘ (2) A investigação lingüística com base em uma coletânea de textos chama-se de Lingüística de Corpus e o programa responsável pela busca de uma palavra escolhida chama-se *concordanceador*, que objetiva processar e apresentar as ocorrências de uma palavra em um corpus. No caso específico dessa pesquisa, o operador argumentativo *mas* é foco da investigação que vocês farão ’.

O programa WordSmith foi usado. Através deste programa, a palavra escolhida (*mas*) apareceu no centro da lista acompanhada de um número de 5 palavras tanto à direita quanto à esquerda da mesma, conforme ocorreu no corpus. As concordâncias foram impressas em papel e utilizadas como atividade de investigação, isto é, observando as palavras que apareceram perto do *mas*, os alunos puderam hipotetizar sobre colocações ou grupo de palavras que podem ocorrer perto do *mas*. Os registros foram feitos em uma folha de papel e cada vez que se fazia necessário um contexto maior, para analisar ou levantar hipóteses de uso, o texto era lido em sua versão ampliada, através do computador. A atividade de investigação procurava revelar os padrões de ocorrências mais frequentes e, nesse sentido, o professor solicitou que os alunos investigassem semelhanças no uso do *mas*. A primeira hipótese remetia à idéia de negação, oposição. O impacto de consultar o corpus, buscar nas concordâncias uma explicação para as implicações do uso do *mas*, revelou ser uma atividade interessante e motivadora para os alunos e, também para o professor: o prazer da curiosidade e da descoberta.

O professor observou que os alunos passaram a trabalhar com mais autonomia de investigação, ao invés de esperar receber tudo pronto. Desta forma, a investigação através do corpus possibilitou o deslocamento do professor como centro de informações e pôs os alunos como parte importante no processo de construção de conhecimento partilhado. Procurando organizar e dar direcionamento às suas descobertas, os alunos procuraram explicar o uso do *mas* e suas colocações, muitas vezes precisando, conforme já mencionado, consultar o corpus para que, num contexto ampliado, pudessem encontrar uma explicação; podendo, dessa forma, testar suas conclusões baseado em evidências. Nas evidências que o corpus manifestava.

Nesse contexto o computador não substitui o professor, ao contrário, foi uma ferramenta, um instrumento utilizado pelos alunos na investigação lingüística. Esse método de investigação, partindo da informação para gerar uma generalização, chama-se de abordagem indutiva. O inverso, portanto, partir da regra para a informação, chama-se abordagem dedutiva. Na maioria das vezes a abordagem dedutiva é privilegiada na prática pedagógica das escolas, quer por razões de praticidade (o professor dá a regra e procura aplicá-la a exemplos inventados) ou por razões de desconhecimento de como fazer de outra forma.

#### 4.1. Exploração do corpus

Investigando o operador *mas*, dois tipos de informação de descrição do corpus foram exploradas: **frequência e colocação**. Conforme já mencionado anteriormente, *WordSmith* é um programa de computador que oferece ferramentas para análise de linguagem em corpus. Para determinar as **colocações**, usamos concordâncias, que são geradas através do Concord Tool do *WordSmith*. O Concord nos permite obter concordâncias de uma maneira mais rápida e mais fácil. É possível, também, acessar a lista de colocações, ou aquelas palavras que ocorrem, perto da palavra escolhida com frequência pré-determinada pelo pesquisador. A distância máxima entre a palavra escolhida e seus colocados é a extensão medida em número de palavras.

Ao escolhermos uma extensão de 3 palavras, podemos determinar se esse número, que não precisa ser o mesmo, vai ser aplicado tanto à direita como à esquerda da palavra escolhida. No caso do experimento, optou-se por uma extensão de 5 palavras, tanto à direita quanto à esquerda. Uma vez que os colocados estejam disponíveis, uma “estrutura” foi gerada, de acordo com Francis e Hunston (1996), isto é, uma generalização sobre o uso da palavra escolhida baseada nos seus colocados. Estruturas foram acompanhadas através de exemplos. Depois disso, um “padrão” foi produzido, buscando uma abstração mais generalizada. Através da comparação desses padrões, as evidências de similaridades e possíveis diferenças começaram a aparecer mais claramente.

A noção fundamental, conforme aponta Sinclair (1991), aplicada à análise é que sentidos distintos são identificáveis como padrões de co-ocorrências distintos. Portanto, se as análises revelarem padrões diferentes para as palavras sob investigação, podem indicar sentidos diferentes. Uma vez que especificados esses padrões, concordâncias foram geradas e impressas. Outro princípio norteador na preparação de materiais foi a frequência dos itens no corpus. Observou-se que esta informação era relevante e devia ser repassada aos alunos porque essa noção - **frequência** - não é observada pelos falantes natos de uma língua e deveria ser obtida através do corpus. Para encaminhamento desta pesquisa, decidiu-se solicitar aos alunos a observação para esta questão, conforme subseção seguinte 4.2.

#### 4.2. Frequência e colocações do “mas “

Solicitou-se aos alunos que observassem com cuidado *como é marcada a presença do operador mas*, isto é, de que forma o *mas* aparece no corpus (nada foi mencionado quanto a aspectos de início, meio ou fim da sentença). Os alunos rapidamente passaram os olhos no corpus e responderam que a vírgula antecedia o *mas*, ou ainda que o *mas* poderia iniciar algumas sentenças. Foram lançadas duas outras perguntas: (1) O que é mais freqüente aparecer à direita ou à esquerda do *mas*, caso inicie uma sentença? ; (2) Caso o *mas* não inicie uma sentença, como o corpus evidencia isso? A segunda pergunta diz respeito às colocações: o que é mais freqüente aparecer ao lado direito e ao lado esquerdo do *mas*?

As perguntas foram lidas pelo professor que acompanhava a turma. Segundo relato do professor, ele mesmo se sentiu tentado a investigar sobre a questão. Os alunos, então, procederam a leitura do corpus, procurando responder a primeira e a segunda pergunta. Como resposta à primeira pergunta, os alunos apontaram para marcações de pontuação, por considerarem evidências, isto é, o *mas* é antecedido de vírgula. Quando inicia uma sentença, o operador *mas* mantém a idéia de oposição: “não acontece nada de diferente”, segundo palavras dos alunos, em relato do professor. Como resposta para a segunda pergunta, os alunos relataram que o *mas* pode ser antecedido de substantivo à esquerda e à direita também. A ocorrência (81) “lhe ocorria a solução. *Mas* ocorreu ao líder” evidencia

o uso de substantivo (solução), *Mas* verbo (acrescentou) que. A ocorrência (73) “forte candidato a vice, *mas* não está decidida” os alunos apontaram o uso de substantivo (vice), *mas* verbo no negativo (não está decidida). Na ocorrência (81) os alunos apontaram para o fato de *mas* aparecer logo após o ponto final da sentença antecedente. Sem maiores explicações, naquele momento.

Apontaram também para a presença de verbos antecedendo e precedendo o *mas* , como no exemplo (77) “eficiente para produzir, *mas* é cego no que..” . Notaram também que no exemplo (78) “continuou falando. *Mas* o coração disparar...”. Nesse momento alguns alunos mencionaram exemplos que incluíam a negação para verbos : no exemplo (51) temos “ não farei nada contra, *mas* também não farei ...” ou , ainda, no exemplo ( 56) temos “não é a de coligar-se, *mas* a de evitar que M..”.

O quadro abaixo resume as conclusões preliminares que os alunos chegaram:

palavra à esquerda	operador	palavra à direita
	<i>mas</i>	
Substantivo	, <i>mas</i>	substantivo
	, <i>mas</i>	verbo
	. <i>Mas</i>	verbo
Verbo (neg.)	, <i>mas</i>	verbo (neg)
Verbo (neg.)	, <i>mas</i>	verbo

Neste momento, os alunos se limitaram a fazer uma análise morfológica das ocorrências evidenciadas pelo corpus. Deixaram de explorar aspectos semânticos de encadeamento lógico provocados pelo operador argumentativo *mas*. Talvez, por se tratar da primeira experiência com o corpus, os alunos apontaram apenas para este nível de análise lingüística.

Posteriormente a esse trabalho, os alunos passaram a comparar os dados do corpus com a gramática normativa de língua portuguesa. Encontraram alguma dificuldade para localizar o lugar dentro da gramática onde poderiam estar os operadores argumentativos e concluíram que as definições dadas pelas gramáticas consultadas não se enquadravam muito bem com a investigação do corpus, conforme sentimento traduzido por uma fala de aluno: “ Pô, parece que não tem nada a ver “, referindo-se a Enéas M.Barros (1982) , focando operador *mas* temos:

a) O *mas* substitui o *e*:

”como também a negação. “Mas ia ele distribuindo bênçãos e carinhos necessitados... “( G. Barros);

“mas estaria ele contente com tudo isso?” (F. Sabino);

b) Indica oposição com valor adverbial:

“ Mas eu é que não vou em teu auxilio, depois do pouco valor que destes aos meus conselhos.” (H. Campos);

c)Indicativo de causa:

“Vocês reclamam de mim por tê-lo esquecido, mas fazendo-o, nenhum outro ato de justiça mais legítimo me faz crescer aos olhos de meus pares. “ (A. Garret)

A fala do aluno refere-se à dificuldade de estabelecer relações diretas com os exemplos autênticos que o corpus aborda. Talvez o aluno tentasse expressar que é bastante difícil para ele, baseado em apenas um exemplo de cada possibilidade de uso, extrair padrões uso de língua. O emprego da palavra *nada* em “Pô, parece que não tem *nada* a ver” pode levar o leitor a tomar *nada* como um distanciamento, uma dificuldade de análise lingüística percebida no contraste gramática tradicional e abordagem de base em corpus. A investigação lingüística proposta pelo experimento tinha, ainda, uma comparação a ser feita, isto é, aceitar ou refutar os postulados de Ducrot para o operador argumentativo em questão: *mas*.

No capítulo 2, *Referencial Teórico*, foram mencionados os postulados de Ducrot (1972), relacionados ao operador argumentativo *mas*. O *mas* (SN) possui valor de refutação, retificação ou justificativa de recusa de *p*, que segue sempre uma proposição negativa. Os exemplos examinados pelos alunos confirmam essa postulação e no caso de *mas* (PA) não demanda uma proposição precedente negativa. Os exemplos investigados do corpus não refutaram nenhuma das proposições de Ducrot (1972). Observou-se, porém, um descontentamento em relação às atividades de pesquisa já conhecidas.

Observou-se que, apesar de promover a investigação lingüística e a busca por palavras mais freqüentes no corpus, não é possível sustentar um ganho significativo na melhora das produções textuais realizadas em 2001, em relação àquelas produzidas em 1999. Houve um natural amadurecimento lingüístico nas produções evidenciado, segundo relato do professor de língua portuguesa, pela fluidez de linguagem ao escrever, pela riqueza de vocabulário e possibilidades de construção textual diversas. Apesar disso, as produções textuais de 2001 revelam que os alunos, usando operador argumentativo *mas*, apontam outras realizações não evidenciadas pelo corpus: “mas será que”, “Mas que ainda não foi” ou ainda “mas sim”. Apesar da manipulação de corpus ter provocado interesse dos sujeitos investigados, apenas quatro alunos produziram textos como atividade de fechamento do experimento. As contingências que envolveram essa produção textual foram: o ano letivo estava quase no fim, os alunos preparavam-se para o vestibular e, muitos deles faziam cursinho pré-vestibular. O desgaste físico e emocional, talvez, tenha contribuído para o número final reduzido de participantes e a qualidade dos textos produzidos.

O professor relatou, em conversa informal, que, para a produção textual específica, alguns alunos dispenderam pouco tempo para a execução da tarefa. Porém, nota-se evidências de escolha dos operadores argumentativos em questão, sobretudo uma organização lógica de caráter global nos textos. O comentário final do grupo, em relação a pesquisa, estava voltado para a inserção de computadores nas atividades de língua portuguesa como algo novo, interessante e que eles gostariam de repetir. O professor, por sua vez, estava deslumbrado com as inúmeras possibilidades de atividades cotidianas que poderiam ser beneficiadas através da manipulação de corpora computadorizados. A subseção seguinte procura explorar ao máximo os resultados em todos os aspectos, especialmente a organização lógica de caráter global dos textos.

### **4.3. Análise dos resultados**

Mesmo com um número tão limitado de produções textuais (quatro), as evidências do uso de técnicas de manipulação de corpora computadorizados transcenderam ao novo, à inserção de computadores na escola. Unem-se a investigação lingüística feita pelo aluno às novas tecnologias e possibilidades de práticas pedagógicas mais interessantes. Observou-se, primeiramente, *como* os alunos *interagiram* com as máquinas e em segundo lugar como a ZDP hipotetizada por Vygotsky se realizou. No caso do experimento, decidiu-se tratar a relação aluno-máquina como uma forma de interação humano-máquina, não foram observadas quaisquer dificuldades com relação ao uso do teclado, mouse, programas, rotinas bastante conhecidas de todos. A escola de forma curricular insere as aulas de informática desde a pré-escola, portanto, é esperado que o aluno ao concluir o ensino médio esteja apto para fazer do computador um instrumento de trabalho e um meio para a realização de tarefas.

Procurando analisar semelhanças no processo de aquisição de linguagem tanto oral quanto escrita, parece razoável esperar que a aquisição do letramento eletrônico também se realize de forma mediada. Conforme já foi mencionado no capítulo 2 - *Referencial Teórico* - a ZDP constitui um potencial para a aprendizagem que é criado na interação entre indivíduos com diferentes níveis de desenvolvimento cognitivo e abrange todas as

funções e atividades que o novato consegue desempenhar, desde que orientado por alguém mais competente. Foi exatamente durante a execução da tarefa baseada em corpus ( busca, frequência e colocações) que a ZDP se manifestou. Os alunos recorreram ao professor ou ao colega (quem estivesse mais próximo, ou disponível naquele momento), na busca de auxílio, buscando resposta para algumas das perguntas relatadas pelo professor: “como é mesmo que eu acho aquele texto ? “ , “o que é que estamos procurando”, “ eu não entendi o quer dizer palavra à direita e à esquerda? “.

As respostas foram “olha o nome do arquivo “ , “as paradas do *mas*”, “quer dizer que palavras estão à esquerda e à direita do *mas*. ”, dadas ou pelo par mais competente ou pelo professor. Nota-se que para a primeira situação podemos, talvez, admitir que o aluno estava com dificuldade de localizar o arquivo que continha o texto desejado. No caso da segunda situação, talvez pudéssemos atribuir uma distração. A última situação, porém, ilustra uma possível ansiedade na resolução do problema. Por se tratar da primeira experiência com Corpus, talvez seja mais esperado que o novo trouxesse alguma insegurança ou ansiedade. O professor procurou manter os alunos na direção da tarefa solicitada, por essa razão as situações descritas acima estavam relacionadas a possibilidade de baixar a ansiedade, focar a atenção e inclusive marcar na tela os elementos que facilitaram a execução da tarefa.

As questões de análise de corpus, contrastando com a gramática normativa, promoveram um grande estranhamento e a impressão de se tratar de línguas diferentes, isto é, os exemplos nem sempre se aproximavam do que o corpus mostrava como uso de língua. Porém, com cuidado, os alunos conseguiram contrastar abordagens distintas e chegar a seguinte conclusão: “ o *mas* expressa negação e tanto faz se introduz ou não uma nova sentença. Podemos colocar uma vírgula antes ou ponto final e começar uma nova sentença com o *mas*. Tem uma ligação aí.” Essa ‘ligação’ a que se referem os alunos remete à análise semântica das ocorrências em seus contextos reais e a dificuldade de perceber como o *mas* marca as relações textuais lógicas que não podem ser facilmente tratadas no âmbito de uma noção mais simples de colocação em termos de contexto imediato.

Concluindo essa seção, podemos apontar para a reação positiva que a experiência possibilitou ao grupo e ao professor, que sentiu “vontade de aprender“ sobre manipulação

de corpora computadorizados. A investigação lingüística proporcionada pela manipulação de corpus provou ser uma atividade interessante e bastante promissora. Foi evidenciada a curiosidade pela descoberta e o corpus como objeto de estudo autêntico em relação aos exemplos de uso de linguagem. A produção textual aconteceu num momento posterior à manipulação de corpus e, como era uma atividade voluntária, apenas quatro alunos se dispuseram a executá-la. O tema das produções textuais foi determinado pelo professor e versava sobre a identidade do homem e da mulher na modernidade. A próxima seção busca uma discussão dos resultados apresentados .

#### **4.4. Discussão dos resultados**

Os resultados apresentados nessa pesquisa apontam para alguns aspectos bastante interessantes. Em uma escola particular com as características de Projeto Político-pedagógico (PPP) como esta, não surpreende que, ao término do ensino médio, os alunos tenham desenvolvido uma consciência crítica relacionada ao seu papel como cidadãos inseridos num mundo de tantos desafios. A situação econômica privilegiada dos alunos, aliada à prática pedagógica dessa escola, viabilizam um trabalho voltado para a prática de projetos, muitos deles engatinhando na interdisciplinaridade. A inserção do uso de computadores, desde a pré-escola, evidencia como a interação humano-máquina pode começar desde muito cedo, no contexto escolar.

Se compararmos os recursos tecnológicos e a proposta de trabalho desta escola com outras da rede particular e pública, certamente os resultados seriam outros. Houve abertura da escola para receber e permitir que a pesquisa se realizasse naquele universo. Os alunos interagiram com os computadores de maneira tranqüila e executaram as tarefas solicitadas de forma contributiva. A inserção de uma nova tecnologia baseada na manipulação de corpora computadorizados não causou estranhamento ou surpresa. Pelo contrário, foi uma experiência bastante rica porque os alunos puderam entrar em contato com uma nova abordagem de investigação lingüística e perceber como o uso de exemplos autênticos de língua pode ser desafiador e interessante.

Em relação às produções textuais, os quatro alunos tiveram as seguintes preocupações: se manterem ligados ao tema proposto, propor um encadeamento de idéias que sustentasse seu ponto de vista, fazer uso de uma argumentação razoavelmente convincente. Os quatro alunos fizeram uso de operadores argumentativos: *pois* ( uma ocorrência) e *mas* (cinco ocorrências). Observou-se, então que o padrão encontrado foi quatro ocorrências de *mas* introduzindo as sentenças, uma ocorrência de *mas* antecedido de vírgula e uma ocorrência de *pois* antecedido de vírgula. O quadro abaixo resume o uso dos operadores , conforme usado pelos alunos :

---

papel do homem,	<i>mas</i> sim de mais destaque
lavam a louça.	<i>Mas</i> isto está mudando
suas próprias casas.	<i>Mas</i> será que chegamos
tão desejada igualdade.	<i>Mas</i> que ainda não foi
que os homens.	<i>Mas</i> será que podemos
ao ser humano,	<i>pois</i> com a quebra de

---

Observa-se que focando o uso do operador *mas* nas produções textuais, as ocorrências de “*mas será que*” são em número de 2 , seguidas de “*mas sim*”, “*Mas isto está*” e, ainda, “*Mas que ainda*”, com uma ocorrência cada.

Neste momento podemos, talvez, afirmar que o número de produções textuais (4) é pequeno e que as ocorrências usadas pelos alunos não aparecem no corpus investigado. A investigação do operador argumentativo *mas*, através do corpus do NILC, pode ampliar os horizontes de uso e fornecer dados que possibilitem uma tentativa de

explicação para as ocorrências de “mas” nas produções textuais produzidas pelos alunos. A busca, então, foi organizada da seguinte forma: corpus corrigido e corpus não-corrigido.

O corpus corrigido inclui textos produzidos ou revisados por profissionais dos quais fazem parte constituição, didático, empresa, enciclopédia, esportes, futebol, história, jornal, jurídicos, literário, livro, livros usp, revista. O corpus não-corrigido inclui textos de inst. Publ. Monografias, orais bbs, publicidade, redação e universitários. Para cada um deles buscou-se o operador *mas*, as ocorrências de *mas sim*, *mas isto e mas será que*. Foi também incluída a busca de *será que*, como tentativa de aproximação de campo semântico com *mas será que*. As subseções seguintes explicitam o percurso de investigação.

#### 4.4.1. O operador *mas* no corpus do NILC

Investigando o corpus corrigido do NILC, focando o operador argumentativo *mas*, foi encontrado um total de 16.001 ocorrências, observando um padrão de 5 palavras tanto à direita quanto à esquerda. Ativando o “show collocates” do programa *WordSmith* é fácil visualizar as colocações do *mas*. A lista a seguir - **MAS 16.001 entries** - mostra as primeiras 45 colocações de *mas* (recorte para fins de limitação desta pesquisa).

## MAS: 16001 entries (sort: 5L,5L)

N	Concordance	Set	TagWord No.	File	%
1	cela!..." Tomou a subir, mas desta vez pela esc		63.268	br~41.txt	58
2	cela!..." Tornou a subir, mas desta vez pela esc		69.664	br~41.txt	64
3	peessoas anualmente. Mas a sociedade que a		10.061	4ag18.txt	99
4	10% a 20% da corrida (mas em viagens peque		8.668	4ab21.txt	92
5	ros usavam camisinha, mas preferiu omitir o te		4.108	4fe14.txt	41
6	10% a 15% da corrida (mas nunca menos de		8.774	4ab21.txt	93
7	\$ 100 o m2, em média. Mas não há mais espaç		2.126	ma22.txt	56
8	4, o mesmo se decidiu, mas não se fica sabend		35.856	cpc10.txt	80
9	Plano derruba inflação, mas não preços EDIT		1.218	4ju30.txt	6
10	e STF - RT 601/262). Mas "a ordem de prisão		26.793	cpc10.txt	59
11	Bol. AASP 1.367/49). Mas, neste caso, até a		26.743	cpc11.txt	70
12	m quatro montanhas. Mas lá também o fim d		1.118	4ab14.txt	10
13	. 14 da Lei n. 6.368/76. Mas não a todas as fig		17.821	cpp03.txt	42
14	dias. Não possui carro, mas mora em casa pró		1.503	ma12.txt	16
15	mil pelo kit-passaporte, mas evita filas e recebe		6.819	4ja20.txt	98
16	4. Cf. capítulo 3. 15. Mas isso não está explí		8.984	ater07.txt	78
17	Jammer 1974, p. 471). Mas, por outro lado, foi		10.145	ater06.txt	69
18	5 de março de 1882. Mas não é só literária a		5.496	acoes.txt	42
19	5 de março de 1882. Mas não é só literária a		15.581	inh02.txt	39
20	15 anos sobre o MTBE, mas até agora ninguém		8.767	ma01.txt	31
21	5 de março de 1882. Mas não é só literária a		5.431	inh03.txt	41
22	mil km sem problemas. Mas conheço casos em		2.281	4ou16.txt	77
23	69 e morrido em 1580, mas tudo é incerto na s		5.994	4ag11.txt	46
24	16 horas e 25 minutos. Mas nada tão distante d		391	4de01.txt	3
25	s, 16, não é tão radical, mas mesmo assim ama		251	4mr21.txt	3
26	/99, RJTAMG 18/212). Mas daí não se segue (		13.285	cpc06.txt	31
27	/99, RJTAMG 18/212). Mas daí não se segue (		13.285	cpc05.txt	31
28	s arts. 17 e 18 do CPC, mas não subtrai à parte		43.056	cpc01.txt	92
29	, sou magra (49 quilos) mas tenho seios muito		3.381	4ag29.txt	33
30	17/355). V. tb. nota 4a. Mas, se o juiz não dete		19.675	cpc04.txt	45
31	fevereiro, na Espanha, mas chega às lojas no i		111	94ja9.txt	4
32	1957) é ídolo nacional, mas os finlandeses têm		4.738	4de22.txt	40
33	878, já valeria a visita. Mas não há como mini		4.554	ma12.txt	51
34	1886. O salto é grande, mas o tempo é um teci		14.722	br~23.txt	20
35	il). Não inclui refeições, mas os apartamentos tê		6.641	ma05.txt	67
36	h, segundo a fábrica. Mas nem mesmo o geni		195	4se25.txt	4
37	a rev ista Minotaure. Mas o jovem argentino		20.657	abato.txt	43
38	razões bem parecidas. Mas como ordenar é ta		6.114	ater09.txt	66
39	á palestras em navios. Mas esta foi a primeira		329	4mr17.txt	3
40	posições semelhantes. Mas, é claro, havia out		5.365	ignolo.txt	65
41	posições semelhantes. Mas, é claro, havia out		5.491	this02.txt	13
42	posições semelhantes. Mas, é claro, havia out		5.496	this04.txt	13
43	propósito da fotografia, mas se comporta realm		1.939	em06.txt	16
44	propósito da fotografia, mas se comporta realm		39.962	em01.txt	37
45	nn & Mainprice, 1987), mas na escala maior de		4.909	ario1.txt	29

A lista seguinte – **collocates (total)** - possibilita observar as palavras mais frequentes tanto à direita( R1, R2, R3,R4,R5) quanto à esquerda (L1, L2, L3, L4, L5).

## collocates (total)

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	MAS	16306	293	30	53	49	46	46	99	83	3	4	7	8	8
2	DE	5837	2969	2868	542	580	744	02	1	0	368	633	616	619	632
3	A	4828	2108	2720	571	566	420	533	18	0	549	533	549	547	542
4	QUE	4193	1262	2931	431	411	284	129	7	0	652	635	627	525	492
5	O	3852	1601	2251	439	440	288	431	3	0	448	411	437	509	446
6	NÃO	3248	1669	1579	558	507	412	152	40	0	397	389	279	294	220
7	SE	2953	1018	1935	364	275	167	164	48	0	344	284	389	421	497
8	E	2722	1434	1288	354	312	355	411	2	0	103	221	309	319	336
9	É	2190	950	1240	216	213	214	293	14	0	277	372	231	194	166
10	DO	1829	1045	784	183	159	224	479	0	0	62	189	170	161	202
11	UM	1798	890	908	198	200	277	212	3	0	235	141	179	166	187
12	DA	1792	976	816	178	186	200	412	0	0	84	193	166	196	177
13	EM	1607	799	808	133	164	226	266	10	0	138	153	201	176	140
14	UMA	1447	646	801	145	160	198	142	1	0	210	150	166	139	136
15	PARA	1185	547	638	149	133	154	110	1	0	108	115	145	134	136
16	COM	1136	544	592	134	141	180	89	0	0	115	108	121	123	125
17	OS	1127	488	639	129	114	96	146	3	0	131	125	121	136	126
18	ME	924	351	573	93	93	79	65	21	0	113	111	112	115	122
19	POR	905	402	503	76	86	112	128	0	0	124	120	91	88	80
20	NO	843	448	395	110	90	81	167	0	0	77	70	88	70	90
21	AS	834	351	483	93	94	84	79	1	0	94	89	96	108	96
22	NA	829	410	419	72	95	97	146	0	0	73	71	93	83	99
23	MAIS	810	368	442	66	90	59	148	5	0	70	90	112	88	82
24	LHE	774	289	485	86	88	72	36	7	0	72	65	107	116	125
25	COMO	762	332	430	83	88	100	61	0	0	75	77	92	99	87
26	SER	647	304	343	70	68	56	99	11	0	68	63	68	69	75
27	AO	607	298	309	50	73	44	131	0	0	49	60	65	59	76
28	DOS	603	340	263	49	74	81	136	0	0	28	59	57	57	62
29	OU	536	349	187	67	60	106	116	0	0	14	23	43	45	62
30	ERA	525	261	264	55	58	59	81	8	0	53	73	47	59	32
31	LO	521	316	205	90	74	58	49	45	0	7	28	46	55	69
32	SUA	511	238	273	40	44	49	102	3	0	59	49	56	51	58
33	LA	507	347	160	86	78	70	65	48	0	3	16	33	48	60
34	À	497	251	246	60	45	42	104	0	0	36	46	50	60	54
35	SEU	462	227	235	32	30	63	95	7	0	44	37	63	58	33
36	SÃO	440	227	213	48	70	42	66	1	0	53	54	41	42	23
37	TAMBÉM	428	116	312	29	35	27	13	12	0	128	79	43	34	28
38	MUITO	422	215	207	35	44	16	94	26	0	39	54	45	36	33
39	DAS	421	225	196	45	49	40	90	1	0	28	49	52	38	29
40	ELE	418	155	263	33	47	24	12	39	0	37	58	49	65	54
41	SÓ	407	221	186	54	59	69	33	6	0	49	46	36	28	27
42	PODE	396	167	229	54	46	44	18	5	0	55	60	43	39	32
43	FOI	371	133	238	34	26	26	40	7	0	61	54	46	39	38
44	EU	329	142	187	40	37	33	19	13	0	34	52	36	35	30
45	NOS	329	135	194	31	26	39	38	1	0	37	31	33	45	48

No caso específico das colocações de *mas* com duas palavras no corpus corrigido, observa-se que a frequência em L1 aponta para as combinações mais frequentes. *Mas se* é a combinação mais frequente de *mas*. São 50 ocorrências, seguidas de 42 ocorrências de *mas não*. Em seguida aparecem *mas ele*, com 39 ocorrências, *mas isso* com 28 ocorrências e *mas muito* com 26 ocorrências. Somando estas cinco colocações, o total é de 185 ocorrências. Procurando expandir as colocações de modo a incluir combinações de três palavras, foi produzida a lista a seguir- **cluster**.

(F)

N	cluster	Freq.
1	mas o que	121
2	mas não é	115
3	mas não se	71
4	ao mesmo tempo	47
5	mas também de	45
6	mas de uma	42
7	mas isso não	41
8	mas que não	39
9	mas também a	37
10	mas é preciso	36
11	mas não há	35
12	mas eu não	31
13	por outro lado	31
14	mas também o	30
15	mas com o	29
16	mas que se	29
17	mas se a	29
18	mas se o	29
19	nem por isso	28
20	mas não o	27
21	mas por que	27
22	mas ainda não	25
23	mas também não	25
24	mas com a	23
25	mas não a	22
26	mas pode ser	22
27	mas é um	22
28	mas é uma	22
29	é claro que	22
30	mas o seu	21
31	mas também para	21
32	mas é que	21
33	não pode ser	21
34	o que é	21
35	mas acho que	20
36	mas nem por	20
37	mas é a	20
38	mas é claro	20
39	o que me	20
40	mas a sua	19
41	mas não me	19
42	que não é	19
43	mas ele não	18
44	mas não são	18
45	mas que a	18

Nota-se que a colocação de *mas o que* aparece como a mais freqüente com um total de 121 ocorrências, seguida de *mas não é* com 115 ocorrências, e por *mas não se* que ocorre 71 vezes. A colocação *mas ao mesmo tempo* ocorre 47 vezes, e *mas também de*, com 45 ocorrências. Somando estas cinco colocações, o total é de 399 ocorrências. Ao cruzar estes dados com as informações relacionadas às colocações compostas por duas palavras, é possível observar que dentre as ocorrências de *mas não*, a sugestão de negação é reforçada.

Investigando o corpus não corrigido do NILC, focando o operador *mas*, foram encontradas 3799 ocorrências, conforme a lista seguinte- **MAS: 3799 entries.**

## MAS: 3799 entries (sort: 5L,5L)

N	Concordance	Set	TaçWord No.	File	%
1	ocais de 0.3 a 2.4 mm , mas o erro inerente pod		2.401	e-e02.txt	89
2	ção tem algum sentido, mas a 2( não tem fund		106	oga22.txt	51
3	mo adversária natural, mas como referência d		834	o-h37.txt	31
4	menina se embelezara, mas Sionésio sequer tin		4.073	o-h06.txt	70
5	ensimesmo." (p. 138) Mas sua admiração pel		1.413	o-h06.txt	25
6	osa, p. 142) "Assim; mas era também o exat		3.439	o-h06.txt	60
7	DAMS & WATT, 1993), mas os mecanismos en		649	e-b01.txt	6
8	(1º grau) é muito alto, mas não tão alto quanto		79	licen2.txt	28
9	ambiente é fornecido, mas o limite do softwar		3.682	e-e46.txt	46
10	ambiente é fornecido, mas o limite do softwar		3.683	e-e47.txt	46
11	fornecido pelo catálogo mas que não foi medido		2.530	e-e03.txt	98
12	as com verbo simples). Mas essa diferença pod		22.185	di-h03.txt	79
13	suporte de hardware, mas utilizando computa		12.113	di-e12.txt	52
14	) A droga é clandestina mas na realidade não é		121	oga47.txt	80
15	5.8, para o modelo do MAS. Por exemplo, a tr		4.640	e-e43.txt	64
16	E HECTARES ESTAO MAS MAOS DE 501 P		77	olitica.txt	18
17	oláres e parece pouco, mas não é - é pouquíss		22	ivre92.txt	9
18	ram sigilo de imprensa, Mas infelizmente ainda		93	v1386.txt	30
19	as ociosas nos cursos. Mas, em 1990, a 358 v		6.221	-arti06.txt	75
20	), temperatura (18oC), mas refinando agora o		19.590	di-e26.txt	90
21	alimento água salgada, mas com algumas espé		2.547	u-b10.txt	98
22	de sua importância. Mas a TV possui um o		178	uv341.txt	47
23	la de um cinescópio. Mas nem tudo é um m		85	lfuv33.txt	46
24	entação de um objeto, mas que seja mantida a		10.892	u-e04.txt	69
25	ostar, comprar e pagar, mas sim de mostrar a s		165	v1214.txt	66
26	norme do divertimento. Mas este aspecto tamb		180	v1256.txt	68
27	MP tem outro critério? Mas todos os critérios q		3.627	-arti06.txt	45
28	r a história de um país. Mas ainda existe a tão		25	lfuv56.txt	8
29	determinado assunto, mas o problema é escre		68	ific11.txt	57
30	o a televisão e o rádio. Mas muitos preferem a		16	uv703.txt	7
31	a frase e o parágrafo, mas não conseguem es		5.256	-arti06.txt	64
32	ira notícia do telejornal. Mas ela enche nossos c		189	v1175.txt	71
33	ito a rede do plim-plim, mas tenho que admitir		164	uv461.txt	68
34	a realidade como ela é. Mas a revelação da ver		73	uv386.txt	27
35	ova não é eliminatória, mas classificatória. Mes		6.363	-arti06.txt	77
36	programas educativos, mas como eles geralme		174	v1190.txt	70
37	uto A é definido em P, mas não em M. Redef		11.746	u-e04.txt	75
38	ia domina a população, mas esta de certa form		228	uv328.txt	64
39	V. perde sua função. Mas, hoje tudo tende a		76	uv408.txt	42
40	rocura de um príncipe. Mas não conseguiu, ela		22.326	di-h02.txt	81
41	a fumando um cigarro, mas não mostra o verd		104	v1191.txt	75
42	oração em seu público. Mas, será que isto é be		51	uv521.txt	25
43	televisão e a imprensa, mas, as pessoas, essas		60	v1463.txt	37
44	speito do que falarem, mas não sabem como e		172	crit24.txt	86
45	a receber informações, mas precisa aprender a		99	uv508.txt	48

## collocates (total)

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	MAS	3972	36	137	12	9	12	1	2	99	43	3	17	29	45
2	DE	1411	813	598	143	171	199	298	2	0	48	81	134	170	165
3	A	1335	539	796	159	127	105	142	6	0	191	147	153	134	171
4	O	1082	427	655	118	94	94	94	27	0	165	143	126	104	117
5	QUE	1037	326	711	118	89	66	51	2	0	134	134	137	142	164
6	NÃO	984	429	555	151	156	91	23	8	0	242	127	90	53	43
7	É	737	269	468	57	72	49	83	8	0	102	86	127	95	58
8	E	658	428	230	97	92	95	144	0	0	17	17	46	59	91
9	DO	448	302	146	32	35	57	119	59	0	11	15	39	39	42
10	UM	435	224	211	51	46	69	56	2	0	20	48	40	46	57
11	SE	424	132	292	41	37	23	29	2	0	51	104	56	39	42
12	PARA	422	195	227	55	46	56	38	0	0	52	38	32	56	49
13	EM	402	198	204	28	60	56	54	0	0	43	55	29	43	34
14	DA	392	250	142	41	50	47	112	0	0	12	14	34	42	40
15	COM	381	144	237	31	42	49	21	1	0	70	62	35	27	43
16	UMA	361	168	193	37	35	58	36	2	0	22	57	49	25	40
17	TAMBÉM	347	22	325	5	2	6	2	7	0	244	44	19	10	8
18	OS	298	135	163	30	33	30	40	2	0	35	34	30	30	34
19	COMO	288	103	185	22	27	33	21	0	0	70	39	18	31	27
20	AS	268	120	148	33	26	23	38	0	0	28	27	34	25	34
21	SÃO	263	119	144	25	31	25	36	2	0	39	39	27	18	21
22	NO	229	147	82	20	26	24	48	29	0	22	19	11	18	12
23	SER	222	105	117	21	29	15	38	2	0	0	31	29	36	21
24	POR	218	96	122	21	15	31	29	0	0	40	29	11	21	21
25	NA	204	93	111	15	22	25	31	0	0	26	26	16	22	21
26	AO	170	91	79	26	16	16	28	5	0	35	14	10	10	10
27	MAIS	170	98	72	26	24	15	30	3	0	1	14	12	23	22
28	PODE	169	46	123	16	10	9	11	0	0	41	26	33	9	14
29	SÓ	163	101	62	19	25	43	14	0	0	19	12	10	9	12
30	MUITO	143	75	68	10	9	4	42	10	0	11	17	17	13	10
31	SIM	141	11	130	2	0	1	1	7	0	126	2	1	0	1
32	ISSO	127	26	101	7	4	5	2	8	0	43	23	11	12	12
33	MUNDO	126	96	30	13	16	16	19	32	0	0	14	4	9	3
34	APENAS	125	76	49	13	31	17	15	0	0	16	3	11	7	12
35	DOS	125	81	44	13	16	15	37	0	0	1	0	7	12	24
36	OU	116	89	27	23	18	22	26	0	0	1	2	3	13	8
37	SUA	111	58	53	11	7	18	22	0	0	14	16	5	8	10
38	ELA	110	46	64	13	13	6	4	10	0	23	9	8	11	13
39	DAS	108	59	49	13	17	7	22	0	0	0	4	16	13	16
40	MESMO	108	22	86	5	6	2	6	3	0	21	31	14	14	6
41	NOS	104	55	49	11	15	14	14	1	0	3	13	12	11	10
42	TEM	104	40	64	10	8	12	9	1	0	22	11	8	14	9
43	PELO	102	46	56	6	6	9	25	0	0	9	9	11	7	20
44	TELEVISÃO	98	57	41	10	10	8	3	26	0	0	12	9	6	14
45	AINDA	90	15	75	5	2	3	5	0	0	38	18	9	6	4

Utilizando o recurso “collocates” do WordSmith é possível apontar que para *mas* a colocação mais freqüente com duas palavras é *mas também*, totalizando em L1, 244 ocorrências. Em seguida aparecem *mas não*, com 242 ocorrências, seguida de *mas a* com 191 ocorrências, *mas o* com 165 ocorrências e *mas que* com 134 ocorrências. Somando estas colocações, o total é de 876 ocorrências. Ao expandir as colocações de modo a incluir combinações de três palavras, foi produzida a lista abaixo - **clusters**.

(F)

N	cluster	Freq.
1	meios de comunicação	31
2	mas não é	26
3	mas o que	26
4	mas será que	25
5	ao mesmo tempo	20
6	mas isso não	20
7	os meios de	20
8	mas ao mesmo	17
9	mas com o	17
10	mas que não	16
11	por outro lado	16
12	é corrigida pelo	16
13	mas mesmo assim	15
14	tendência é corrigida	15
15	mas também o	14
16	mas também com	13
17	essa tendência é	12
18	isso não é	12
19	mas a televisão	12
20	mas essa tendência	12
21	mas não se	12
22	mas por outro	12
23	não é o	12
24	para o mas	12
25	mas com a	11
26	mas não são	11
27	mas o mundo	11
28	mas pode ser	11
29	mas sim a	11
30	mas também para	11
31	mas também pode	11
32	mas que são	10
33	ciclo de vida	9
34	hoje em dia	9
35	mas isto não	9
36	mas não o	9
37	mas para o	9
38	mas sim de	9
39	mas também não	9
40	modelo de objetos	9
41	métodos do mas	9
42	o modelo de	9
43	mas apesar de	8
44	mas muitas vezes	8
45	mas não a	8

Observa-se que a colocação *mas não e* aparece como a mais freqüente, com 26 ocorrências, acompanhada por *mas o que* com 26 ocorrências. Em seguida aparecem *mas será que* com 25 ocorrências, em quarto lugar foi encontrado *ao mesmo tempo* com 20 ocorrências e a quinta colocação encontrada foi *mas isso não* com 20 ocorrências, totalizando 177 ocorrências.

As listas do corpus corrigido e do corpus não corrigido, focando o *mas*, apontam para o mesmo sentido, isto é, a idéia de negação se mantém. Foi necessária uma busca

no corpus do NILC para *mas será que*, pois havia aparecido nas produções textuais dos alunos realizadas em 2001. A subseção seguinte (4.4.2) investiga, portanto, *mas será que*.

#### 4.4.2. “*Mas será que*” no corpus do NILC

Decidiu-se incluir “*mas será que*” na investigação do operador *mas* devido ao fato de estar presente nas produções textuais dos alunos. Notou-se que *mas será que* não fazia parte do corpus de manipulação a que os alunos tiveram acesso, levantaram hipóteses de colocações, frequência, e sentido. Procurava-se uma explicação para tal.

Um total de 61 ocorrências de *mas será que* foi encontrado no corpus corrigido do NILC, conforme lista a seguir: **MASxSERAxQUE**. Optou se por manter o mesmo numero de 45 ocorrências por questões de limitação de apresentação desta pesquisa. Portanto, a lista de produzida para *mas será que* encontra-se na página a seguir.

## MASxSERAxQUE: 61 entries (sort: 5L,5L)

N	Concordance	Set	TagWord No.	File	%
1	ranco. Coisa simples. Mas será que uma pros	880	94fe2.txt	10	
2	ou seja, o ecossistema. Mas será que a casa o	1.531	cien3.txt	17	
3	universal e inevitável. Mas será que os seus	532	4se24.txt	19	
4	so aqui e no exterior. Mas será que podemos	3.926	4ju06.txt	38	
5	nsa prega firmemente, mas será que anda cum	2.859	4mr27.txt	15	
6	Eu sei que o Muro caiu mas será que o Muro c	11.416	4ab28.txt	91	
7	prerrogativa da vítima). Mas será que invocar a	12.444	4ja16.txt	63	
8	binária de "O Ingênuo". Mas será que seu senti	9.229	4no20.txt	33	
9	s primeiros namoros. Mas será que uma mulh	1.362	94jl17.txt	25	
10	dos os outros também. Mas será que vai jogar?	9.137	4ju20.txt	45	
11	ue dá para contar tudo. mas será que eu consig	619	6ab21.txt	68	
12	ssalvar esse aspecto. Mas será que o celular	2.399	4se22.txt	89	
13	ilhete da Lufthansa. – Mas será que não vai s	2.532	4ab10.txt	10	
14	ioso das privatizações. Mas será que elas conti	3.285	4se26.txt	27	
15	squete. Pode até ser, mas será que os pontos	2.772	ma17.txt	26	
16	o país e o contribuinte. Mas será que a concorr	2.211	4ja13.txt	85	
17	salvador de São Paulo, mas será que o segund	6.516	4ou22.txt	71	
18	e excelência desejado. Mas será que existe me	201	4ab10.txt	1	
19	UNK O punk retorna. Mas será que ele estev	9.169	4mr17.txt	70	
20	se expôs ao eleitorado, mas será que os outros	5.678	ma06.txt	81	
21	andam lá essas coisas, mas será que não dava	2.160	4ju10.txt	22	
22	, assalta. É verdade. Mas será que esse anú	9.436	4ja05.txt	82	
23	em agilidade mental. Mas será que a crítica	5.899	4no02.txt	77	
24	s do escritor argentino. Mas será que é igualme	3.503	4ja19.txt	34	
25	smo, um velho safado, mas será que a gente	4.135	4ab03.txt	50	
26	o isso à força do rádio. Mas será que os Raimu	6.280	ma30.txt	82	
27	retudo nossos serviços. Mas será que os trabalh	2.101	94jl07.txt	20	
28	números são números, mas será que Adriana e	9.399	94jl11.txt	87	
29	aldo Jabor condena-a. Mas será que o filme é	4.157	4ou12.txt	75	
30	struidor, é anti-sistema. Mas será que ainda se	3.932	4mr13.txt	45	
31	uma gracinha, não é"? Mas será que ele é "th	3.022	4fe10.txt	25	
32	dada das barraquinhas. Mas será que cortar ba	85	6ju23.txt	2	
33	deste final de século. Mas será que ainda pod	56.010	em01.txt	52	
34	deste final de século. Mas será que ainda pod	5.902	em08.txt	81	
35	limites como crítico. Mas será que fui limitad	9.799	4ja14.txt	75	
36	mo estrelas e galáxias. Mas será que somos re	61	cien1.txt	1	
37	les fazem mal à saúde. Mas será que esse valo	23.481	ma08.txt	96	
38	ão e o Painei do Leitor, mas será que todos os	2.094	4fe14.txt	21	
39	e não, cru e cozido. Mas será que tudo "é li	12.025	4ju24.txt	83	
40	uende, o sujo, o pagão. Mas será que não esta	2.703	4se07.txt	57	
41	es do inverno europeu. Mas será que eles real	21.905	4ju19.txt	97	
42	consagrado americano. Mas será que uns repre	8.468	4ju01.txt	70	
43	buição, o que for. Ah, mas será que tem públi	5.447	4ou17.txt	70	
44	buição, o que for. Ah, mas será que tem públi	5.447	4ou17.txt	70	
45	riscar-se à vontade. Mas será que, quando a	7.777	4de09.txt	60	

Ao produzirmos a lista de colocações para *mas será que*, encontramos a lista de **collocates** abaixo:

**collocates (total)**

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	QUE	68	6	62	2	2	2	0	0	0	0	60	1	0	1
2	MAS	63	1	1	0	1	0	0	0	61	1	0	0	0	0
3	SERÁ	61	0	61	0	0	0	0	0	0	60	1	0	0	0
4	O	21	15	6	4	5	2	4	0	0	0	0	4	0	2
5	É	17	6	11	2	0	1	2	1	0	0	0	5	3	3
6	A	15	6	9	2	2	1	0	1	0	0	0	6	1	2
7	E	13	12	1	2	2	3	5	0	0	0	0	0	0	1
8	DE	12	12	0	2	2	3	5	0	0	0	0	0	0	0
9	NÃO	10	5	5	3	1	0	1	0	0	0	0	3	0	2
10	OS	7	1	6	0	0	1	0	0	0	0	0	5	1	0
11	SE	7	3	4	1	1	1	0	0	0	0	0	0	2	2
12	DA	6	5	1	1	1	0	3	0	0	0	0	0	0	1
13	DO	6	6	0	1	0	3	2	0	0	0	0	0	0	0
14	UMA	5	2	3	1	1	0	0	0	0	0	0	3	0	0

Destas ocorrências a colocação mais freqüente foi *mas será que a* com um total de 6 ocorrências, seguida de *mas será que os* com 5 ocorrências, e por *mas será que é* com 5 ocorrências. A combinação *mas será que ele* é a quarta combinação mais freqüente, com 4 ocorrências e a quinta combinação mais freqüente é *será que o*, com 4 ocorrências, totalizando 24 ocorrências mais freqüentes com *mas será que*.

(F)

N	cluster	Freq.
1	mas será que	61
2	será que a	6
3	será que os	5
4	será que é	5
5	será que ele	4
6	será que o	4
7	que é só	3
8	será que ainda	3
9	será que esse	3
10	será que não	3
11	será que uma	3

No corpus não-corrído do NILC foram encontradas 25 ocorrências de *mas será que*, conforme a lista abaixo: **MASxSERAxQUE**.

## MASxSERAxQUE: 25 entries (sort: 5L,5L)

N	Concordance	Set	TagWord No.	File	%
1	ulação é se horrorizar. Mas será que esta mes		76	oga24.txt	47
2	sam melhor que uma", mas será que nossos d		147	bis13.txt	67
3	a nossa vida atual. Mas será que a TV é s		163	uv403.txt	82
4	do conto" - o leitor dirá. Mas será que estará me		874	o-h41.txt	44
5	cial do próprio homem. Mas será que os nossos		39	uv430.txt	12
6	ndo e de todo o Brasil, mas será que todas as		69	uv152.txt	22
7	ue a televisão chegue. Mas será que isso é bo		48	uv297.txt	27
8	orização do presidente. Mas será que o preside		96	lfuv31.txt	24
9	o ninguém tem dúvida, mas será que vale a pe	1.942	-arti11.txt		65
10	de informação e lazer. Mas será que esses mei		29	uv956.txt	15
11	ela vejo o mundo todo. Mas será que vemos tu		40	uv748.txt	15
12	cuos lugares da Terra. Mas será que com essa		31	v1267.txt	11
13	manipulam as pessoas, mas será que seria pos		137	uv460.txt	57
14	do manipulada como é. Mas será que a TV, se		35	uv560.txt	15
15	próprias identidades. Mas será que a T.V. no		113	v1280.txt	40
16	u não do novo regime. Mas será que o povo, c		52	lebis5.txt	18
17	pior. E há muito mais... Mas será que vale a pe		26	v1352.txt	13
18	as da nossa sociedade. Mas será que elas pode		14	oga60.txt	11
19	Parlamentarista. Mas será que um siste		57	bis21.txt	45
20	sária em nossas vidas, mas será que uma TV		264	uv416.txt	96
21	m uma antena de T.V., mas será que o dono d		137	v1264.txt	57
22	nato ou roubo também. Mas será que podemo		126	ivre94.txt	54
23	de invento do homem; Mas será que esse inve		14	uv936.txt	7
24	mais sem a televisão, mas será que isto é bo		33	uv416.txt	12
25	mundo de sonhos. Mas será que esta vida		101	uv313.txt	45

A lista de **collocates** a seguir para *mas será que* procura evidenciar quais são os colocações mais frequentes:

## collocates (total)

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	QUE	28	2	26	0	1	0	1	0	0	0	25	0	0	1
2	MAS	25	0	0	0	0	0	0	0	25	0	0	0	0	0
3	SERÁ	25	0	25	0	0	0	0	0	0	25	0	0	0	0
4	A	10	4	6	1	1	1	1	0	0	0	0	3	2	1
5	O	8	3	5	0	0	2	1	0	0	0	0	3	0	2
6	E	6	5	1	2	2	0	1	0	0	0	0	0	0	1
7	É	6	3	3	1	0	1	0	1	0	0	0	0	2	1
8	DE	5	4	1	0	2	0	2	0	0	0	0	0	0	1
9	DO	5	5	0	1	0	2	2	0	0	0	0	0	0	0

Destas ocorrências a colocação mais freqüente foi *mas será que a*, com 3 ocorrências, seguida de *mas será que o*, com 3 ocorrências, totalizando 6 ocorrências mais freqüentes. A lista abaixo- **cluster**- demonstra esta freqüência:

## (F)

N	cluster	Freq.
1	mas será que	25
2	será que a	3
3	será que o	3

A investigação feita no corpus corrigido e no corpus não-corrigido para *mas será que* procura evidenciar uma ampliação de sentido no uso de *mas-* objeto desta pesquisa. Os exemplos abaixo procuram explicitar esta possibilidade de leitura:

#### Corpus corrigido

1. “... por “casa” o lugar onde vive o ser vivo, ou seja, o ecossistema. Mas será que a casa onde você mora também não pode ser considerada um pequeno ecossistema?” -file 6cien3.

A introdução de questionamento proposta pela sentença “Mas será que a casa onde você mora também não pode ser considerada um pequeno ecossistema?” provoca no leitor uma possível dúvida, seguida ou não de uma hipótese: caso minha casa não possa ser considerada um pequeno ecossistema, poderia ser um grande ecossistema? Poderia ser um ecossistema complexo, seres vivos e não-vivos? As possibilidades de levantar hipóteses para responder a pergunta podem levar o leitor a diferentes respostas.

2. “ ...água com sais minerais, ou outras substancias. Mas, será que não existe água pura na natureza?” – file 5cien3.

O questionamento sobre a existência ou não de água pura leva o leitor a pensar em uma resposta e, talvez levantar uma hipótese para ambos os casos de resposta afirmativa ou negativa.

3. “ 120 laudas acho que dá para contar tudo. Mas será que eu consigo escrever tanto? – file cl96ab21

O questionamento “ Mas será que eu consigo escrever tanto?” pode introduzir um efeito de dúvida relacionada a produção de 120 laudas. O autor (eu) parece duvidar de sua habilidade de escrita ao referir-se ao número de laudas , reforçada pela palavra “tanto”.

4. “ Saiu a lista dos intelectuais que apóiam o FHC! Eu sei que o Muro caiu mas será que o Muro caiu justo em cima dos intelectuais?” – file 94ab28

A introdução de questionamento proposta por “ mas será que o Muro caiu justo em cima dos intelectuais?” pode levar o leitor a deduzir que Muro caiu , isto é, deixou o cargo que ocupava. Assim sendo, Muro estaria fazendo parte da lista de intelectuais que apóiam FHC, ou não ? Para a hipótese afirmativa saberíamos que Muro faz parte da lista ; para a hipótese negativa Muro “caiu” em outro lugar . Pode-se , talvez, afirmar que o sentido de “mas será que o Muro caiu justo em cima dos intelectuais?” além de um questionamento pode-se inserir uma proposta de levantamento de hipóteses. É difícil afirmar qual sentido prevalece sobre o outro.

#### Corpus não-corrigido

1. “A resposta está naquele parágrafo inicial do conto” – o leitor dirá. Mas será que estará mesmo? – file n-mo-h41

O questionamento proposto por “Mas será que estará mesmo” pode indicar uma dúvida sobre a localização da resposta. Em se tratando de um texto com atividades de compreensão baseadas em perguntas e respostas, precisamos retornar ao texto para confirmar ou refutar a possível localização. Neste contexto, a possibilidade de questionamento parece ser mais apropriada.

2. “Hoje em dia não há canto no mundo em que a televisão chegue. Mas será que isso é bom?” – file fuv297

A introdução de questionamento em “ Mas será que isso é bom?” pode expressar uma dúvida relacionada aos prós e contras de não haver “canto no mundo em que a televisão chegue”. A presença disseminada da televisão pode realmente beneficiar as pessoas? Parece que o autor preocupa-se com isso. A dúvida se mantém.

3. “usada corretamente não sendo manipulada como é. Mas será que a TV, sem essa manipulação que sofre atualmente seria tão importante como é ? –file fuv560

O questionamento proposto pelo autor, isto é, a importância da TV, seria a mesma sem essa manipulação atual ? Segundo o autor, “usada corretamente” o efeito poderia ser outro.

4. “drogas são consideradas um dos principais problemas da nossa sociedade. Mas será que elas podem ser consideradas um problema conseqüente da atual sociedade capitalista consumista? - File-oga60

O autor afirma que drogas são um ” problema da nossa sociedade”. Ao introduzir um questionamento sobre uma possibilidade deste problema poder ser considerado uma conseqüência da sociedade capitalista consumista, o autor procura resposta para sua pergunta. Nesse contexto, talvez seja possível afirmar, que independentemente da resposta ser afirmativa ou negativa, as drogas são um problema e o palco de discussão se instaura.

Observa-se que os exemplos acima apontam para uma introdução de indagação ou hipótese, dependendo do contexto imediato. Muitas vezes é difícil classificar uma determinada ocorrência de *mas será que* como apenas uma indagação ou hipótese.

Procurando, nesse momento, uma aproximação de sentido do corpus do NILC para *mas será que* com as produções textuais produzidas pelos alunos no ano de 2000, as duas ocorrências encontradas estão listadas abaixo.

(1) Aluna A.B.:

“conquistando seu espaço; no voto, no trabalho, em suas próprias casas. Mas será que chegamos a ter hoje direitos iguais para homens e mulheres.”

A introdução de questionamento proposta pela autora, isto é, os direitos são iguais para homens e mulheres pode levar o leitor a diferentes respostas. Parece que o objetivo da autora é exatamente este: promover uma reflexão.

(2) Aluna T. :

“muitas mulheres atuando em muitas profissões, exercendo seu trabalho como , ou melhor que, os homens. Mas será que podemos dizer que o preconceito ao sexo frágil chegou ao fim?”

A autora questiona o fim do preconceito ao sexo frágil . Ao fazer isso, a introdução se faz por meio de “ Mas será que podemos dizer que o preconceito ao sexo frágil chegou ao fim?” Desta forma, a autora procura descobrir se o preconceito ainda existe.

As outras duas produções textuais e as ocorrências de *mas* mencionadas no item 4.4., isto é, *mas sim, mas isto, mas que* , diferentemente das ocorrências de *mas será que*, procuram reiterar a idéia de negação.

Foi incluída, também , uma busca a *será que* tanto no corpus corrigido como no corpus não-corrigido para procurar uma aproximação, talvez, de sentido entre *mas será que* e *será que*. Observou-se um total de 1022 ocorrências no corpus corrigido de *será que*, conforme lista abaixo: **SERAxQUE: 1022 entries**

## SERAxQUE: 1022 entries (sort: 5L,5L)

N	Concordance	Set	TagWord No.	File	%
1	rk. Ai, eu me pergunto: será que é de interesse		2.458	ma17.txt	23
2	u ter que ficar sozinho? Será que tenho que des		5.288	4ju24.txt	37
3	MODELO-PADRÃO Será que já cobri todo		38.548	filcien.txt	84
4	itarrista da Sepultura. Será que é justo conde		3.063	4ja31.txt	33
5	a emissora). Ah, sim... será que dava para as		5.933	l94ja8.txt	50
6	ão viu o lobo –e bang! Será que o bicho morre		184	4ag05.txt	2
7	preciso do transplante. Será que eu vou suport		9.739	94jl15.txt	79
8	preciso do transplante. Será que eu vou suport		1.546	94jl15.txt	13
9	, em várias interações. Será que existem outra		3.218	imi05.txt	31
10	esquema dominante. Será que assim "Fera F		20.489	4ag21.txt	85
11	lquer tico-tico. Agora, será que de lá para cá		8.478	4ab26.txt	61
12	ante da grave questão: será que vou ter que li		4.612	4de11.txt	97
13	, fico olhando para ele. Será que tento de novo		19.051	ma08.txt	65
14	z não acabe em pizza! Será que nem o Boris f		7.891	94jl15.txt	50
15	o, estou sendo racista. Será que todo mundo n		6.034	4no02.txt	79
16	er o que vai acontecer. Será que a operação da		1.189	jor~43.txt	45
17	com isso. Eu pensava: será que vou ter que fi		5.271	4ju24.txt	37
18	er o que vai acontecer. Será que a operação da		1.189	6mr05.txt	45
19	e ninguém diz nada? Será que a imprensa es		2.096	4se14.txt	21
20	Ao Galvão Bueno? Ou será que você é parent		734	94jl31.txt	9
21	nter aquecidos? Deus, será que alguém está p		13.142	ma15.txt	52
22	bém não está careca. Será que ela rasparia o		3.079	4se11.txt	38
23	ito bem nas pesquisas. Será que isto é precon		21.946	4se11.txt	79
24	acontecer com a moda. Será que amanhã não h		3.304	4ab04.txt	31
25	acontecer com a moda. Será que amanhã não h		3.304	4ab04.txt	31
26	MACHADO Por que será que o Lobo Mau n		1.790	4ag27.txt	41
27	r exemplo, esbravejou: "Será que os estilistas		6.390	4ja20.txt	54
28	e pergunta: "Em quem será que eu votei?". 1		8.367	4no16.txt	34
29	e pausa e indagou:"Ou será que muita coisa ac		3.217	4mr27.txt	20
30	o principal até simples: será que os bancos dep		405	4ju18.txt	15
31	quisa dessas no Brasil. Será que os brasileiros		1.777	4ag07.txt	17
32	gracinha, não é"? Mas será que ele é "the one		3.023	4fe10.txt	25
33	ivamente sem sentido: Será que existe alguma		3.506	capb7.txt	49
34	e pausa e indagou:"Ou será que muita coisa ac		3.217	4mr27.txt	29
35	ivamente sem sentido: Será que existe alguma		9.336	etafisi.txt	46
36	na moderna. Porque será que sua afirmação		18.055	4ju05.txt	74
37	ão das mais difíceis. Será que os exemplos		2.616	4no13.txt	96
38	como outra qualquer? Será que o Palácio do		3.559	4fe14.txt	63
39	ção clara: "Para existir, será que eu precisava		6.731	4no09.txt	63
40	aralelo com a seleção. Será que César Sampai		5.203	4fe14.txt	66
41	ser inominável: "O que será que seria o ser da		11.141	demo.txt	25
42	rosto alegre. Por que será que existem pesso		4.898	lant14.txt	68
43	de trambiqueiros? Ou será que os verdadeiros		1.997	4ag03.txt	16
44	o"? Mais amplamente, será que existe alguma		3.517	capb7.txt	49
45	renesi no meio político. Será que o PL é o únic		4.805	4ag05.txt	23

A lista de colocações para *será que* está a seguir: **collocates**

A lista de colocações para *será que* está a seguir: **collocates**

**collocates (total)**

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	QUE	1264	163	1101	42	33	19	10	59	0	944	78	6	32	41
2	SERÁ	1134	19	95	9	3	0	0	7	20	78	1	4	3	9
3	O	307	117	190	34	26	21	36	0	0	0	120	20	18	32
4	A	280	127	153	34	27	28	38	0	0	0	78	26	25	24
5	DE	256	182	74	51	32	42	57	0	0	0	2	7	34	31
6	NÃO	201	54	147	21	13	11	8	1	0	0	60	36	34	17
7	É	154	48	106	11	8	12	14	3	0	0	41	29	20	16
8	E	138	120	18	31	29	27	20	13	0	0	0	3	5	10
9	OS	98	28	70	12	8	4	4	0	0	0	52	11	1	6
10	DO	96	71	25	9	17	18	27	0	0	0	0	0	11	14
11	OU	94	77	17	3	6	3	3	62	0	0	0	1	10	6
12	MAS	85	83	2	6	5	7	1	64	0	0	0	0	0	2
13	SE	82	35	47	3	9	6	17	0	0	0	15	11	16	5
14	DA	81	58	23	13	12	18	14	1	0	0	0	1	12	10
15	POR	76	59	17	5	6	11	37	0	0	0	0	7	5	5
16	UM	76	43	33	18	10	7	8	0	0	0	4	6	8	15
17	EM	70	49	21	10	8	16	15	0	0	0	1	4	4	12
18	NO	68	49	19	14	9	10	16	0	0	0	3	7	5	4
19	PARA	67	39	28	8	15	13	3	0	0	0	0	4	9	15
20	VAI	67	10	57	3	2	2	3	0	0	0	18	19	14	6
21	COM	65	49	16	17	4	13	15	0	0	0	0	4	7	5
22	UMA	65	39	26	11	11	6	11	0	0	0	9	3	4	10
23	ELE	59	17	42	3	2	6	3	3	0	0	32	8	0	2
24	ESTÁ	58	20	38	6	3	8	2	1	0	0	3	16	10	9
25	TEM	54	8	46	1	4	2	1	0	0	0	11	17	10	8
26	AS	51	17	34	7	5	1	4	0	0	0	18	6	4	6
27	ELA	49	10	39	2	0	3	5	0	0	0	31	5	1	2
28	NA	45	28	17	4	11	3	10	0	0	0	1	2	6	8
29	MAIS	44	28	16	4	12	3	7	2	0	0	0	1	4	11
30	VOCÊ	42	7	35	1	2	0	1	3	0	0	23	2	2	8
31	EU	41	19	22	9	3	1	6	0	0	0	17	3	2	0
32	DOS	37	26	11	6	9	5	6	0	0	0	0	2	2	7
33	ISSO	35	11	24	2	4	3	1	1	0	0	11	5	6	2
34	SER	34	18	16	3	1	3	10	1	0	0	1	2	12	1
35	ALGUÉM	32	1	31	0	1	0	0	0	0	0	24	6	1	0
36	ESSE	31	7	24	0	2	4	1	0	0	0	11	7	3	3
37	AO	28	22	6	7	3	7	5	0	0	0	4	0	0	2
38	FOI	28	15	13	5	2	3	4	1	0	0	7	4	1	1
39	SEU	28	16	12	2	6	4	4	0	0	0	4	0	2	6
40	AINDA	27	9	18	1	2	3	2	1	0	0	5	6	4	3
41	ELES	27	4	23	1	1	1	0	1	0	0	15	4	2	2
42	TAMBÉM	27	15	12	1	4	2	6	2	0	0	2	4	5	1
43	ME	26	14	12	0	3	3	8	0	0	0	0	1	10	1
44	SÓ	26	6	20	1	2	2	1	0	0	0	9	5	2	4
45	COMO	25	16	9	2	4	6	2	2	0	0	0	0	5	4

Procurando ampliar estas combinações para 3 palavras, a lista de **cluster** evidencia esses dados:

(F)

<b>N</b>	<b>cluster</b>	<b>Freq.</b>
1	será que o	127
2	será que a	82
3	ou será que	66
4	mas será que	61
5	será que não	61
6	que será que	59
7	será que os	56
8	será que é	47
9	será que ele	40
10	será que ela	37
11	por que será	34
12	será que alguém	27
13	será que você	25
14	será que eu	19
15	será que vai	19
16	o que será	18
17	será que as	18
18	será que eles	17
19	será que esse	16
20	e será que	13
21	será que isso	13
22	será que se	13
23	será que tem	13
24	será que existe	12
25	será que ninguém	11
26	será que já	10
27	que você não	9
28	será que nós	9
29	será que só	9
30	será que uma	9
31	que não tem	8
32	será que podemos	8
33	que ele não	7
34	será que foi	7
35	será que sou	7
36	será que vale	7
37	que ela vai	6
38	que não é	6
39	que vale a	6
40	que você está	6
41	será que esta	6
42	será que estou	6
43	será que existem	6
44	vale a pena	6
45	folha será que	5

Deste total as colocações mais freqüentes foram *será que o*, com 127 ocorrências, seguida de *será que a*, com 82 ocorrências. A terceira colocação mais freqüente foi *ou será que*, com 66 ocorrências, seguida de *mas será que*, com 61 ocorrências e a quinta colocação mais freqüente foi *será que não*, com 61 ocorrências, totalizando 397 ocorrências.

No corpus não corrigido foram encontradas 166 ocorrências de *será que*, conforme lista abaixo- **SERAxQUE: 166 entries**

SERAxQUE: 166 entries (sort: 5L,5L)

N	Concordance	Set	TagWord No.	File	%
1	rem de fome? Enfim, será que, parafraseand		336	v1419.txt	95
2	ido ao progresso. Mas, será que essas transfor		28	uv916.txt	11
3	pulam as pessoas, mas será que seria possível		138	uv460.txt	57
4	escravos da telinha. Será que a televisão te		194	uv396.txt	53
5	ejo o mundo todo. Mas será que vemos tudo o		41	uv748.txt	15
6	a antena de T.V., mas será que o dono deste		138	v1264.txt	58
7	vida para ser pensada: será que o bonito mostr		87	uv486.txt	35
8	ouco assustada: o que será que está acontece	3.460	-arti06.txt		43
9	em dirige a imprensa? Será que esse mentiros		82	uv956.txt	41
10	smite são verdadeiras? Será que ela pode bene		52	uv399.txt	13
11	está voltado para eles. Será que um dia cairão		189	v1352.txt	91
12	is união numa família? Será que é preciso uma		215	lfuv39.txt	75
13	fontes são confiáveis? Será que não estamos		77	uv152.txt	25
14	acabar com eles, então será que o Brasil vai ac		138	ser33.txt	84
15	pergunta poderá surgir: Será que os meios de		71	v1113.txt	23
16	o benéfica para todos ? Será que aqueles que c		78	uv983.txt	28
17	rapontos e indagações. Será que tudo que nos		108	tfuv2.txt	51
18	ira ainda esperanças. Será que sempre terem		122	int171.txt	74
19	o próprio homem. Mas será que os nossos mei		40	uv430.txt	12
20	SERÁ QUE TEMOS TA		0	v1468.txt	0
21	HÃ DE AMANHÃ Será que podemos proj		3	uv302.txt	3
22	PERANÇA. SERÁ ? Será que aTV é a chav		6	uv388.txt	4
23	TELEVISIONADA Será que realmente a te		2	uv926.txt	2
24	er e agir da sociedade? Será que tal sociedade		18	uv926.txt	8
25	geração teleguiada. Ou será que precisaremos	209	uv913.txt		92
26	anipulada como é. Mas será que a TV, sem es		36	uv560.txt	15
27	ios de comunicações ? Será que iríamos sobre		97	lfuv24.txt	44
28	E alguns perguntavam: Será que esta caixa nã		46	uv264.txt	21
29	nesse sentido. Porque? Será que se sentiram c		328	int153.txt	74
30	s no fazendo chorar. Será que hoje saberiam		132	v1478.txt	46
31	m presidente corrupto. Será que sem uma mas		140	v1433.txt	72
32	o ela seria malfeitora ? Será que o lucro que te		262	uv854.txt	87
33	bastível irá aumentar, será que ele não repass		119	sser4.txt	70
34	e é sem volta mesmo? Será que a maconha va		54	oga28.txt	44
35	a temperatura local. Será que o dia-a-dia re		21	uv372.txt	12
36	acontece no mundo? E será que interessa que t		134	uv589.txt	50
37	SER RESPONDIDA Será que estamos send		8	v1194.txt	4
38	televisão chegue. Mas será que isso é bom? T		49	uv297.txt	28
39	nossa sociedade. Mas será que elas podem se		15	oga60.txt	11
40	luência governamental. Será que as grandes e		136	uv490.txt	36
41	amentarista. Mas será que um sistema (r		58	bis21.txt	46
42	ção do presidente. Mas será que o presidente c		97	lfuv31.txt	24
43	convocados pela TV, será que sem a TV hav		161	uv190.txt	50
44	ores e injustiças. Será que tantas mentira		102	enac2.txt	58
45	er tamanha selvageria. Será que poderá deposi		177	dicina.txt	79

A lista abaixo procura mostrar as colocações de *será que*: **collocates**.

collocates (total)															
N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	QUE	202	22	180	7	5	4	1	5	0	162	4	5	4	5
2	SERÁ	186	14	6	4	3	3	2	2	166	2	0	0	2	2
3	A	67	26	41	8	3	8	7	0	0	0	23	7	7	4
4	O	44	20	24	3	8	1	8	0	0	0	16	0	3	5
5	MAS	34	34	0	1	0	1	0	32	0	0	0	0	0	0
6	É	32	11	21	2	4	3	2	0	0	0	5	7	6	3
7	DE	31	21	10	6	3	5	7	0	0	0	0	1	7	2
8	E	26	21	5	7	3	8	2	1	0	0	0	0	2	3
9	UMA	20	12	8	2	5	2	3	0	0	0	2	2	2	2
10	NÃO	19	5	14	3	2	0	0	0	0	0	7	1	3	3
11	DO	18	12	6	1	4	3	4	0	0	0	0	0	2	4
12	OS	17	8	9	1	3	3	1	0	0	0	5	1	2	1
13	AS	13	7	6	2	2	3	0	0	0	0	2	4	0	0
14	DA	13	11	2	1	1	4	5	0	0	0	0	0	1	1
15	COM	12	6	6	1	2	3	0	0	0	0	3	0	2	1
16	EM	12	9	3	4	2	1	2	0	0	0	1	0	0	2
17	OU	12	10	2	0	2	4	2	2	0	0	0	0	1	1
18	SE	12	6	6	2	0	2	2	0	0	0	2	1	1	2
19	TV	12	4	8	2	1	0	0	1	0	0	0	6	1	1
20	UM	12	5	7	1	1	3	0	0	0	0	3	0	3	1
21	NOS	11	4	7	0	1	1	2	0	0	0	0	1	4	2
22	PARA	11	10	1	3	3	2	2	0	0	0	0	0	0	1
23	TELEVISÃO	10	6	4	1	1	1	2	1	0	0	0	2	1	1
24	DIA	9	3	6	0	0	1	0	2	0	0	0	5	0	1
25	MAIS	9	7	2	2	1	3	1	0	0	0	0	1	0	1
26	POR	9	5	4	1	1	1	2	0	0	0	1	0	0	3
27	BANCO	8	8	0	4	0	0	4	0	0	0	0	0	0	0
28	ESTÁ	8	2	6	0	1	0	1	0	0	0	2	1	2	1
29	SEM	8	4	4	1	1	1	1	0	0	0	2	1	1	0
30	TUDO	7	1	6	1	0	0	0	0	0	0	1	1	2	2
31	AO	6	5	1	0	1	2	2	0	0	0	0	1	0	0
32	DIFERENTE	6	1	5	0	0	0	0	1	0	0	0	4	0	1
33	ELA	6	1	5	0	0	1	0	0	0	0	4	1	0	0
34	ESSE	6	1	5	0	0	1	0	0	0	0	3	1	0	1
35	MEIOS	6	1	5	0	0	1	0	0	0	0	0	3	1	1
36	MESMO	6	1	5	0	0	0	0	1	0	0	0	2	0	3
37	SER	6	4	2	2	0	0	2	0	0	0	0	0	2	0
38	VIDA	6	5	1	0	1	2	1	1	0	0	0	1	0	0
39	ESSA	5	0	5	0	0	0	0	0	0	0	1	1	1	2
40	ESSAS	5	0	5	0	0	0	0	0	0	0	4	1	0	0
41	ESTA	5	2	3	0	1	0	1	0	0	0	3	0	0	0
42	ISTO	5	1	4	0	1	0	0	0	0	0	3	0	1	0
43	MUNDO	5	4	1	0	1	1	1	1	0	0	0	0	1	0
44	NA	5	4	1	1	0	2	1	0	0	0	0	0	0	1
45	PESSOAS	5	3	2	0	1	0	1	1	0	0	0	1	0	1

A lista de colocações para *será que* -cluster- segue abaixo:

(F)

N	cluster	Freq.
1	mas será que	25
2	será que a	23
3	será que o	16
4	será que não	7
5	que a tv	5
6	que será que	5
7	será que essas	5
8	será que os	5
9	será que é	5
10	banco plano ou	4
11	diferente posicionamento do	4
12	o diferente posicionamento	4
13	o que será	4
14	ou banco inclinado	4
15	plano ou banco	4
16	que o diferente	4
17	será que ela	4
18	liberdade de expressão	3
19	meios de comunicação	3
20	os meios de	3
21	que a televisão	3
22	que não é	3
23	que todas as	3
24	será que com	3
25	será que esse	3
26	será que esta	3
27	será que este	3
28	será que isto	3
29	será que todas	3
30	será que um	3

No corpus não corrigido foram encontradas 166 ocorrências de *será que*. Deste total as colocações mais frequentes foram *mas será que* com 25 ocorrências, seguida de *será que a* com 23 ocorrências. A terceira colocação mais frequente foi *será que o* com 16 ocorrências seguida de *será que não* com 7 ocorrências e a quinta colocação mais frequente foi *que será que* com 5 ocorrências, totalizando 76 ocorrências

Começa a aparecer, neste momento, informação relevante no que diz respeito ao uso da expressão *mas será que* em contexto, a partir de resultados produzidos por um programa concordanceador, no caso desta pesquisa o programa *WordSmith*. A estratégia de usar estas informações numéricas possibilitam o cruzamento de resultados relativos às colocações de duas palavras mais frequentes com os dados das colocações de três palavras mais frequentes. Focando as cinco primeiras ocorrências de *mas* no corpus corrigido, nota-se que não aparece a ocorrência de *mas será que*. Observando as ocorrências de *mas será que*, podemos apontar para uma recorrência de determinantes (a, os) e pronomes (ele, esse) à direita de *mas será que*.

Nos contextos analisados e, alguns exemplos retirados do corpus corrigido e não corrigido, o sentido de indagação ou hipótese se mantém. Assim sendo, baseado nas evidências que os dados do corpus do NILC apresentam, talvez possamos afirmar que *mas será que* produz um efeito de sentido diferente do efeito de adversativa atribuído ao operador argumentativo *mas*. No caso do *será que*, a indagação está presente em todas as ocorrências e as colocações de *será que* encontram em *mas* uma frequência bastante elevada como colocado.

Procurando analisar as produções textuais produzidas pelos alunos, tanto para o ano de 1999 quanto para o ano de 2001, buscou-se as concordâncias, colocações e frequências para o operador argumentativo *mas*. As produções textuais de 1999 evidenciam 25 ocorrências do *mas*, as colocações mais frequentes são *mas também*, com 4 ocorrências, seguida de *mas não*, com 3 ocorrências. A terceira ocorrência mais frequente foi *mas a*, com 3 ocorrências, *mas é*, com 2 ocorrências e *mas uma* com 1 ocorrência, totalizando 13 ocorrências, conforme listas de anexos. Não foram encontrados “clusters”  $\geq 3$  para *mas*.

O mesmo programa concordanceador foi empregado tanto para o corpus selecionado da *Folha de São Paulo* quanto para analisar as produções textuais produzidas pelos alunos no ano de 2001. Tomando como corpus de manipulação os textos da *Folha*, foram encontrados 459 ocorrências de *mas*. As colocações mais freqüentes foram *mas o*, 40 ocorrências, *mas não*, com 39 ocorrências, seguida de *mas a*, com 22 ocorrências. A quarta ocorrência mais freqüente foi *mas é*, com 18 ocorrências e a quinta ocorrência mais freqüente foi *mas os*, com 14 ocorrências, totalizando 133 ocorrências. Para uso em sala de aula, por ser o primeiro contato com Lingüística de Corpus, a manipulação de corpus foi realizada, envolvendo todos os textos previamente selecionados, porém as folhas com o concordanceador de *mas* foram limitadas a um número de 138 ocorrências apenas. Não há nenhuma ocorrência de *mas será que*.

As produções textuais de 2001 apresentam 2 ocorrências de *mas será que*, uma ocorrência de *mas sim*. Também foi encontrada 1 ocorrência de *mas isto* e 1 ocorrência de *mas que*. Ambas as ocorrências aparecem como colocações de *mas*, tanto no corpus corrigido como no corpus não-corrigido. As produções textuais de 1999 caracterizam o uso de oposição, mudança argumentativa, negação. O mesmo não pode ser dito das produções de 2001. A introdução de questionamento proposta por *mas será que* leva a um outro rumo diferente da negação de *mas*. As marcas identificadas em várias ocorrências de *mas será que* produzidas pelos alunos propõe uma leitura outra: não podemos nos ater ao fato de que *mas* é sempre uma forma de negar, de dizer não.

Observou-se, também, que os gêneros textuais mais freqüentes para a investigação com *mas será que* no corpus corrigido foram os textos de jornal. Para o corpus não corrigido, a presença mais marcante foi a de redação de vestibular.

Desta forma, a Lingüística de Corpus proporciona uma investigação abrangente, confiável, pois os atributos do corpus devem ser preenchidos para poderem ser considerados representativos de um recorte de língua: escrita ou falada. Sem a presença do corpus esta pesquisa não teria se realizado e os alunos não poderiam ter experienciado uma investigação lingüística diferente dos padrões já conhecidos. Pode-se, talvez, propor neste momento uma aproximação de campo semântico entre *mas será que* e *será que*. Ambos

introduzem uma indagação, uma hipótese, por isso completamente diferente do sentido de negação de *mas*. A possibilidade de propor estas atividades de pensar sobre a linguagem que usamos, dentro do contexto de sala de aula, é bastante desafiador e pode possibilitar cruzamentos e comparações diversas. Poder mostrar para os alunos como a língua é repleta de recursos, sentidos, intenções deve sinalizar um novo caminho tanto para a prática pedagógica quanto para as atividades de descoberta da língua.

O capítulo seguinte- *Conclusão*- procura identificar se os objetivos gerais e específicos desta pesquisa foram alcançados e também procura apontar sugestões para pesquisas futuras.

## CONCLUSÃO

A justificativa da pesquisa descrita aqui, conforme apontada na *Introdução*, está relacionada à importância da inserção de novas tecnologias de forma a fazer uso criativo dos computadores em sala de aula. O uso de corpora atende a busca por soluções novas com base em recursos tecnológicos.

De acordo com as pesquisas mais recentes em Linguística Aplicada, uma abordagem que leve em conta o uso autêntico de linguagem para fins de investigação linguística, promove simultaneamente uma mudança de paradigma da prática pedagógica ( que precisa estar sempre se atualizando), isto é, o professor deixa de ser o centro das atenções e passa a dividir com o aluno a possibilidade de construção de conhecimento compartilhado. Acredita-se que os novos e promissores rumos de inovações tecnológicas não tragam um sentimento ingênuo de necessidade urgente de “ um novo fazer ” . Além disso, se o uso de técnicas de manipulação de corpora computadorizados pode fazer com que os alunos passem a se tornar ativos e responsáveis por suas descobertas, muito do processo de aprendizagem será beneficiado. Desta forma, o uso de corpus pode ser um recurso poderoso de aprendizado.

Se a perspectiva interativa for trabalhada, estaremos favorecendo uma prática pedagógica mais interessante e o desenvolvimento de alunos mais ativos. Desta forma, o corpus pode ser utilizado como recurso e, conseqüentemente, ferramenta de aprendizado. Assim, o principal objetivo desta pesquisa foi propor um ensino de operadores argumentativos baseado na metodologia de linguística de corpus e verificar se um ensino específico sobre operadores argumentativos resulta num melhor desempenho linguístico e metalinguístico: reconhecimento e uso dessa função linguística. De forma geral, pode-se dizer que esse objetivo foi atingido.

Embora não haja medição objetiva para o objetivo principal desta pesquisa, talvez seja possível afirmar que o ensino de operadores baseado na metodologia da Linguística de Corpus foi bastante beneficiado, isto é, os alunos investigaram as ocorrências em uso autêntico de linguagem e promoveram contraste com a gramática normativa. Neste sentido,

observou-se que a investigação lingüística promovida pelo corpus manteve os alunos motivados e envolvidos durante a tarefa. Desta forma, podemos afirmar que o principal objetivo desta pesquisa foi atingido.

Observou-se que os Parâmetros Curriculares de Santa Catarina adotados pela escola pesquisada que subjazem ao ensino de leitura/produção de textos não fazem qualquer referência à Lingüística de Corpus, técnicas de manipulação de corpora computadorizados. Além disso, é possível perceber que as possibilidades de conteúdo sintetizadas na página 76 dos Parâmetros Curriculares de Santa Catarina evidenciam os eixos de trabalho com língua portuguesa. No que diz respeito à construção de texto e argumentação encontramos “ análise e comparação de argumentos”, apenas.

O segundo objetivo específico trata da possibilidade de experienciar a presença da abordagem com base em corpus fundamentada nas operações metalingüísticas, focando o operador argumentativo *mas*. Pode-se afirmar que o uso de técnicas de manipulação de corpora computadorizados possibilitou a experiência da abordagem com base em corpus viabilizando um trabalho de investigação lingüística bastante interessante e confirmando a primeira hipótese de trabalho levantada para esse experimento, isto é, “o ensino específico dos operadores argumentativos melhora a cognição quanto a reconhecer e usar as nuances entre os operadores estudados”. A segunda hipótese que apontava para “o ensino específico do operador *mas* provoca uma melhoria na capacidade dos aprendizes produzirem sentenças explorando essas diferenças ” também foi confirmada.

A terceira e última hipótese sobre o ensino específico do operador *mas* , utilizando computadores e abordagem com base em corpus, melhora a análise e a metacognição dos sujeitos a respeito de reconhecer e aplicar determinados usos do operador *mas* foi parcialmente confirmada. O termo parcialmente é empregado aqui como um limitador da hipótese de reconhecer e aplicar determinados usos do operador *mas*. Este fato está ligado ao reconhecimento de usos, porém devido ao número reduzido de produções textuais, não é possível afirmar que todos os sujeitos investigados aplicam determinados usos do operador *mas*, pois apenas quatro voluntários fizeram a produção textual.

A manipulação do corpus foi trabalhada buscando agrupar similaridades explicitadas nas sentenças. Também foi solicitado aos alunos que procedessem à comparação daquelas evidências com a gramática tradicional. Foram realizados contrastes com o corpus, para explorar as possibilidades de poder de explicação da gramática tradicional. Também foi solicitado aos alunos que produzissem um texto. Por fim, as produções textuais realizadas foram analisadas e o “ padrão ” do *mas* foi evidenciado. O contexto de produção textual não foi favorável: a atividade foi voluntária e, segundo o professor de língua portuguesa, a qualidade dos textos produzidos foi inferior àquela normalmente produzida por eles. Apontou que pouco tempo foi dedicado à execução da tarefa, possível causa comprometedora da qualidade dos textos.

Ao utilizarmos abordagens mais modernas sobre operadores argumentativos obtém-se ganhos expressivos no ensino e no processos de aprendizagem. Em primeiro lugar, insere-se o uso criativo de computadores em sala de aula em contexto de pesquisa. Desta forma, a investigação lingüística provoca um conseqüente deslocamento dos papéis tradicionais de professor e aluno frente ao objeto de estudo. A inserção desta uma tecnologia traz ganhos, de modo geral, para professores, lingüistas e ensino-aprendizagem. Em outras palavras, poder tratar ensino como pesquisa. Em segundo lugar, tomando o uso do “mas” não somente como um fenômeno lingüístico, mas cognitivo, assume-se que a compreensão do uso do *mas* produz novo conhecimento. Na verdade a idéia de “oposição” do *mas* não é algo novo, porém a descoberta de novas relações entre os operadores argumentativos produz uma nova leitura de mundo.

Assumir, então, que essa visão pode possibilitar uma prática pedagógica diferenciada permite adotar uma outra e importante postura como professor. Para que novas tecnologias sejam inseridas no contexto de ensino e a aceitação de suporte tecnológico em aulas de língua seja prática cotidiana, precisamos rever referenciais teóricos e práticas pedagógicas. De um modo geral, professores, especialistas, lingüistas podem se beneficiar muito com as vantagens de tornar rotina o uso criativo de computadores em sala de aula. O caminho de manipulação de corpora em computadores, como esta pesquisa demonstrou, é uma alternativa possível, barata e muito frutífera. Como sugestão para futuras pesquisas, podemos apontar para a possibilidade de aplicação a outras escolas e séries da rede

particular ou pública. Talvez, também, poder contrastar as ocorrências de *mas, mas será que e será que* com a língua falada. Enfim, fomentar a investigação sobre a linguagem, o que, por extensão, torna o ensino mais interessante. O uso da Linguística de Corpus, em situação de ensino-aprendizagem, se mostrou promissora. Sugere-se, pois, mais pesquisas voltadas para a investigação lingüística com foco na Linguística de Corpus, bem como a inserção de práticas pedagógicas que usem a Linguística de Corpus.

## REFERÊNCIAS

- AARTS, J. e Meijs, L. *Corpus Linguistics*. Amsterdam: Rodopi, 1984.
- ABERCOMBRIE, D. *Studies in phonetics and linguistics*. Londres: Oxford University Press, 1965.
- BIBER, D. , CONRAD, S. e REPPEN, R. *Corpus Linguistics investigating Language structure and use*; Cambridge: U.C. , 1998.
- CABRAL, L. G . *Lingüística e ensino: reflexões para a prática pedagógica da língua materna/* Loni Grimm Cabral, Edair Gorski. Florianópolis: Insular, 1998.
- CELANI, M. A. A. *A transdisciplinaridade da lingüística aplicada no Brasil*. Conferência Plenária proferida pelo IV Congresso Nacional de Lingüística Aplicada. Campinas: Unicamp, 1995.
- CITELLI, A. *O texto argumentativo – 1ª. ed*, São Paulo: Scipione, 1994.
- DUCROT, O. *Princípios de semântica lingüística: Dizer e não dizer*. São Paulo: Cultrix, 1992.
- \_\_\_\_\_. *Provar e dizer. Leis lógicas e Argumentativas*. Trad. Maria Aparecida Barbosa e alii. São Paulo: Pontes, 1980.
- \_\_\_\_\_. *O dizer e o dito*. Trad. Eduardo Guimarães, São Paulo: Pontes, 1987.
- \_\_\_\_\_. *Argumentação e “topoi” argumentativos*. In: Guimarães, Eduardo, org. *História e sentido da linguagem*. Campinas: Pontes, p.13-38, 1989.
- FRANCIS, W.N. Problems of assembling large corpora. In: Johansson (Org.) *Computer Corpus in English Language research*. Bergen: Norwegian Computing Centre for the Humanities, 1982.
- \_\_\_\_\_. Language Corpora B.C. In: Jan Svastvik (Org) *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter, 1992.
- GERALDI, J. W. *Linguagem e ensino: exercício de militância e divulgação- 1ª. ed.*, Campinas: SP, 1998.
- HUMBLÉ, P. O uso de corpora no ensino de línguas. Alguns exemplos do português e do espanhol. In: CABRAL, L., SOUZA, P., VASCONCELLOS, L. R. , PAGOTTO, E. (Org.) *Lingüística e ensino: Novas Tecnologias*. Blumenau: Nova Letra, 2001.

- KLEIMAN, A. *Texto e Leitor: Aspectos Cognitivos da Leitura*. 1ª. ed. , Campinas: Fontes, 1989.
- KENNEDY, G. . *An introduction to Corpus Linguistics*. New York: Longman, 1998.
- KOCK, I.G.V . *A inter-ação pela linguagem*, 4ª. ed. , São Paulo: Contexto, 1998.
- LEECH, G. *Corpora and theories of linguistics performance*. In: Jan Svartvik (ed) *Directions in corpus linguistics*, p.105-22. Berlin, Mouton de Gruyter, 1992.
- MCENERY, T. & WILSON, A. . *Corpus Linguistics* Edinburgh: Edinburgh University Press, 1996.
- MOHAN, B. Models of the role of the computer in second language development. In: PENNINGTON, M. e STEVENS, V. (Org.) *Computers in applied linguistics*. Clevedon, Multilingual Matters, 1992.
- MORATTO, E.M. *Linguagem e cognição : as reflexões de L.S.Vygotsky sobre a ação reguladora da linguagem*. São Paulo: Plexus, 1996.
- ROCHA, M.A. Anotações do curso de pós-graduação em Lingüística Aplicada na disciplina de Lingüística Computacional . Florianópolis, Universidade Federal de Santa Catarina, 2000.
- \_\_\_\_\_. O uso de corpora computadorizados no ensino de língua portuguesa: metodologia e avaliação. In: CABRAL, L. SOUZA, P. , VASCONCELLOS, R. e PAGOTTO, E. (Orgs.). *Lingüística e ensino: novas Tecnologias*. Blumenau: Nova Letra, 2001.
- SARDINHA, A.P.B. . Beggining Portuguese Corpus Linguistics: Exploring a Corpus to teach Portuguese as a foreign language. *D.E.L.T.A.*, vol, 15, no. 2: 289-99, 1999.
- SINCLAIR, J. *Corpus, Concordance, Collocation*. (Describing English Language Series) Oxford: OUP, 1991.
- SMOLKA, A.L.B. Conhecimento e produção de sentidos na escola. A linguagem em foco. A linguagem em foco. In: *Cadernos C.E.D.E.S*. São Paulo: Papyrus, no. 35, 1995.
- TUNES, E. Os conceitos científicos e o desenvolvimento do pensamento verbal. In: *Cadernos C.E.D.E.S*, 1995.
- VYGOTSKY, L.S. *Pensamento e Linguagem*. São Paulo, Martins Fontes, 1989 .

**ANEXOS**

1. Texto “Aqui se vive melhor, publicado pela revista *Veja* em 07 de abril de 1999
2. Texto “Paisagem e espaço” publicado pela *Metamorfoses do Espaço Habitado*-Hucitec- S.P., S.P. (1998)
3. Texto “Mea Culpa” publicado na revista *Nossa América* (no. 2) ano de 1993
4. Concordanceador das produções textuais produzidas no ano de 1999
5. Versão digitalizada das produções textuais produzidas no ano de 1999
6. Concordanceador das produções textuais produzidas no ano de 2001
7. Versão digitalizada das produções textuais produzidas no ano de 2001