

SÍNDIA ROSA BALLEM DA SILVA

**O ITEM LEXICAL ‘QUE NEM’: UM ESTUDO COM
BASE EM CORPUS**

FLORIANÓPOLIS

2006

SÍNDIA ROSA BALLEEN DA SILVA

O ITEM LEXICAL ‘QUE NEM’: UM ESTUDO COM BASE EM CORPUS

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre, pelo Curso de Pós-Graduação em Lingüística – área de Lingüística Aplicada – da Universidade Federal de Santa Catarina.

Orientador: Prof. Dr. Marco Antônio Esteves da Rocha

FLORIANÓPOLIS
2006

TERMO DE APROVAÇÃO

SÍNDIA ROSA BALLEEN DA SILVA

O ITEM LEXICAL ‘QUE NEM’: UM ESTUDO COM BASE EM CORPUS

Dissertação aprovada como requisito parcial para obtenção do grau de MESTRE EM LINGÜÍSTICA, na área de Lingüística Aplicada, Linha de Pesquisa em Lingüística de Corpus, pelo Programa de Pós-Graduação em Lingüística, da Universidade Federal de Santa Catarina, Mestrado Interinstitucional UFSC/UTFPR, pela seguinte banca examinadora:

Prof. Dr. Marco Antônio Esteves da Rocha
Orientador - UFSC

Prof. Dr. Antonio Berber Sardinha
LAEL - PUC/SP

Prof. Dr. Felício Wessling Margotti
UFSC

Prof. Dr. Philippe René Humblé
Suplente - UFSC

Florianópolis, 19 de abril de 2006

AGRADECIMENTOS

Com a alegria de ter contado com o incentivo de muitos na realização deste trabalho, deixo registrados nesta página os meus mais profundos agradecimentos:

- A Deus, fonte de luz e inspiração;
- Ao prof. Marco Rocha, orientador deste trabalho, pela acolhida, compreensão, orientação competente e, principalmente, pela paciência;
- Aos Luiz Eduardo (tanto o esposo, como o filhinho), pelo incentivo e por compreenderem minha ausência. Vocês foram a minha motivação para a continuidade e persistência neste trabalho;
- Aos meus pais: Ana Loiva e Setembrino, pela grande força e incentivo;
- Aos demais familiares, pelo incentivo;
- À Deise, pela dedicação e grande auxílio durante esse período tumultuado;
- À Adriana, pelos valiosos comentários sobre este trabalho e pela grande doação e amizade;
- À Denize, pela amizade e auxílio;
- À Zulméa, pelo companheirismo e pelos livros;
- À Maria Helena, Elizeth, Adelaide e a todos os colegas do Mestrado, pela amizade, apoio, companheirismo;
- Aos colegas de trabalho, pelo incentivo;
- A todos os professores que contribuíram efetivamente para meu crescimento acadêmico;
- E a todos aqueles que, de alguma maneira, contribuíram para a concretização deste trabalho.

SUMÁRIO

| | |
|---|-----------|
| LISTA DE FIGURAS | 6 |
| LISTA DE TABELAS | 7 |
| RESUMO | 8 |
| ABSTRACT | 9 |
| 1 INTRODUÇÃO | 10 |
| 2 REFERENCIAL TEÓRICO | 14 |
| 2.1 Lingüística de Corpus | 14 |
| 2.2 Corpus – definição e características | 15 |
| 2.3 Lingüística de Corpus – breve histórico | 16 |
| 2.4 O uso do computador | 18 |
| 2.5 Lingüística de Corpus: disciplina ou metodologia | 19 |
| 2.6 Tipos de pesquisas privilegiados | 20 |
| 2.7 Padrões de linguagem | 21 |
| 2.8 ‘Que nem’ nas obras de referência | 24 |
| 2.8.1 ‘Que nem’ – de acordo com o dicionário..... | 24 |
| 2.8.2 ‘Que nem’ – de acordo com a Gramática Tradicional | 26 |
| 2.8.3 ‘Nem’ – em outras obras de referência | 30 |
| 2.9 Gêneros textuais | 37 |
| 3 METODOLOGIA | 41 |
| 3.1 Corpus da pesquisa | 41 |
| 3.2 O Programa <i>WordSmith Tools</i> | 43 |
| 3.3 A pesquisa | 47 |
| 4. Resultados e análise | 48 |
| 4.0 Classificação geral do ‘que nem’ | 49 |
| 4.1 1º padrão: que nem = sequer, ao menos | 49 |
| 4.2 2º padrão: que nem = não | 51 |
| 4.3 3º padrão: que nem = como | 52 |
| 4.3.1 Categoria de natureza léxico-gramatical | 52 |
| 4.3.1.1 Categoria 1 – verbo + que nem + substantivo | 53 |
| 4.3.1.2 Categoria 2 – adjetivo + que nem + substantivo | 54 |
| 4.3.1.3 Categoria 3 – substantivo + que nem + substantivo | 55 |
| 4.3.1.4 Categoria 4 – substantivo + que nem + pronome | 55 |

| | |
|---|-----------|
| 4.3.1.5 Categoria 5 – pronome + que nem + substantivo | 55 |
| 4.3.1.6 Categoria 6 – substantivo + que nem + verbo..... | 55 |
| 4.3.1.7 Categoria 7 – verbo + que nem + pronome | 56 |
| 4.3.1.9 Categoria 8 – pronome + que nem + pronome..... | 56 |
| 4.3.1.9 Categoria 9 – adjetivo + que nem + pronome..... | 56 |
| 4.3.1.10 Categoria 10 – adjetivo + que nem + verbo..... | 56 |
| 4.3.1.11 Categoria 11 – outros..... | 57 |
| 4.3.1.12 Conclusões sobre as categorias de natureza léxico-gramatical..... | 57 |
| 4.3.2 Categoria de gêneros textuais | 59 |
| 4.3.2.1 Registros literários | 60 |
| 4.3.2.2 Registros jornalísticos | 61 |
| 4.3.2.3 Entrevistas..... | 62 |
| 4.3.2.4 Livros técnicos..... | 63 |
| 4.3.2.5 Monografia | 64 |
| 4.3.2.6 Dissertação | 64 |
| 4.3.3 Conclusões sobre as categorias de gêneros textuais | 65 |
| 4.4 4º padrão: que nem todo (s), toda (s), tudo | 68 |
| 4.5 5º padrão: que nem sempre | 69 |
| 4.6 6º padrão: que nem ... nem (alternância) | 69 |
| 4.7 7º padrão: que nem mesmo | 69 |
| 4.8 8º padrão: que nem sequer | 70 |
| 4.9 9º padrão: que nem ao menos | 71 |
| 4.10 10º padrão: que nem + advérbio (só, bem, mais, assim)..... | 71 |
| | |
| 5 CONSIDERAÇÕES FINAIS | 73 |
| | |
| 6 REFERÊNCIAS BIBLIOGRÁFICAS | 76 |

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1: Lista de concordância | 45 |
| Figura 2: Agrupamento lexical (<i>cluster</i>)..... | 46 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1: Padrões do que nem : ocorrências e frequência..... | 49 |
| Tabela 2: Coligações 1: ocorrências e frequência | 52 |
| Tabela 3: Coligações 2: ocorrências e frequência | 58 |
| Tabela 4: Gêneros textuais: ocorrências e frequência | 59 |

RESUMO

Este trabalho apresenta um estudo sobre o item lexical **que nem**. Através de instrumentos da Lingüística de Corpus, pretende-se descrever o padrão de uso desse item do português brasileiro, ou seja, o padrão de regularidade nos tipos de associações a que se submetem essas palavras. O objeto principal da pesquisa centra-se no padrão **que nem**, locução conjuntiva comparativa, pois se tenciona verificar se esse item lexical, recorrente na modalidade oral da língua, é também recorrente em textos extraídos da língua escrita e, em sendo, em que gêneros textuais aparece comumente. Para isso, serão extraídas de um corpus eletrônico as ocorrências autênticas do item lexical **que nem**, que serão comparadas com o que é mencionado em obras de referência, como gramáticas e dicionários e outros referenciais que apresentam o assunto, partindo dos padrões de sentido. A percepção inicial de limitações em obras de referência foi confirmada. A maneira pela qual o tema é tratado pelas gramáticas e dicionários nem sempre corresponde ao uso real da língua, porém a utilização de corpora como fonte de informação sobre essa mesma língua, constitui um grande recurso para investigar seu uso, possibilitando novas descobertas. Além disso, quando são utilizados recursos computacionais para a manipulação de dados, os resultados obtidos são mais confiáveis.

Palavras-chave: Lingüística de Corpus; classe de palavras; padrões de linguagem; obras de referência lingüística.

ABSTRACT

This paper presents a study about the lexical item **que nem**. Through the instruments of Corpus Linguistics, we intend to describe the pattern of use of this item of the Brazilian Portuguese, in other words, the pattern of regularity in the types of association that the words submit. The main goal of the research is centered in the pattern **que nem**, comparative conjunctive locution, because we aim to verify if this lexical item that appears to the oral modality of the language happens in texts taken out of written language and, being so, in what textual gender it commonly appears. For this, the authentic occurrences of **que nem** will be extracted from an electronic corpus, which will be compared with what is mentioned in reference works, such as grammars, dictionaries and other reference that present the subject, from patterns of meaning. The initial perception of the limitations in reference works was confirmed. The way the theme is treated by the grammars and dictionaries does not always correspond to the real use of language, but the using of corpora as resource of information about this same language constitutes a great resource to investigate the use, enabling new discoveries. Besides, when computing resources are used for the data manipulation, the results are more reliable.

Key-words: Corpus Linguistics; word class; language patterns; linguistic reference works.

1 INTRODUÇÃO

A Lingüística de Corpus (doravante LC) é uma área da Lingüística que tem como prioridade utilizar um corpus lingüístico como fonte fundamental de informações e de conhecimento sobre as línguas humanas. Para tanto, utiliza o computador como instrumento de análise.

Os estudos relacionados à língua normalmente costumam fundamentar-se em situações artificiais de uso da linguagem, com exemplos inventados e/ou selecionados de obras de autores consagrados, inclusive nas obras de referência como os dicionários e as gramáticas. Já numa abordagem baseada em corpus, procura-se fazer um estudo empírico da língua com a análise sistemática dos dados que se apresentam em determinado corpus e que correspondem a exemplos reais de uso da linguagem.

Assim, com base em uma abordagem denominada de *abordagem com base em corpora lingüísticos* que, no pensamento de Leech (1997, p.1), têm sido comumente designados como um conjunto de textos autênticos para ser usado como base para a pesquisa lingüística, este trabalho apresenta um estudo sobre o item lexical **que nem**.

Estabelece-se o **que nem** como um item lexical, pois assim define Basílio (1998, p.54): “... em princípio, um item lexical é um complexo de propriedades morfológicas, sintáticas e semânticas. Assim, sua pertinência a classes deve ser estabelecida em termos morfológicos, semânticos e sintáticos”.

Através dessa abordagem da Lingüística de Corpus (LC), pretende-se descrever o padrão de uso desse item do português brasileiro, ou seja, o padrão de regularidade nos tipos de associações a que se submetem as palavras, principalmente o padrão **que nem** enquanto locução conjuntiva comparativa.

Um corpus é definido como uma grande quantidade de textos (orais e/ou escritos) produzidos por falantes nativos em situações reais de comunicação. Esses textos são armazenados eletronicamente e devem ser legíveis por meios computacionais que têm a capacidade de buscar, selecionar, ordenar, recuperar, contar e calcular os conteúdos de corpora de textos de grande porte, de forma relativamente rápida. Essas tarefas são básicas em termos de processamento de dados, porém a

investigação manual de um determinado fenômeno da língua ou variedade lingüística, sem os recursos computacionais, seria muito trabalhosa e mais sujeita a erros, principalmente se o corpus for grande.

Nesse sentido, uma pesquisa baseada na abordagem da Lingüística de Corpus, com a utilização de corpora eletrônicos, apresenta-se como uma alternativa viável para descrever a língua, pois será possível, por exemplo, verificar as ocorrências de determinado item lexical, no caso, o **que nem**, através do acesso a um grande número de exemplos de expressões da língua em uso, com a utilização da concordância que, nesse contexto, não tem relação com a concordância no sentido gramatical. Por se tratar de uma metodologia de análise lingüística, a LC utiliza como ferramenta de análise a concordância que é um instrumento que consiste numa lista de ocorrências de uma determinada palavra, expressão ou morfema, apresentados no contexto em que ocorrem em um corpus.

Pode-se obter uma concordância usando um programa concordanciador, tal como o *WordSmith Tools*, que é uma alternativa mais simples para os usuários do Windows, além de ser barata e fácil de usar.

Neste trabalho, será utilizado o programa *WordSmith Tools* a fim de extrair concordâncias de um corpus para investigar qual é a padronização do item lexical **que nem**. Optou-se por esse tipo de pesquisa, pois o uso de corpora computadorizados, sobretudo sob a forma de concordâncias extraídas destes corpora, é um recurso poderoso para investigar o uso da língua, tanto no que diz respeito à gramática quanto à eficácia comunicativa ou mesmo na descoberta de fatos novos da língua.

Quanto ao objeto principal desta pesquisa, parece não restar dúvidas de que, na modalidade oral da língua, o item lexical **que nem** – locução conjuntiva comparativa - forma popular e estigmatizada, conforme Almeida (1982, p. 348) seja recorrente. Por isso, pretende-se verificar se esse item lexical também é recorrente em textos extraídos da língua escrita, reconhecidos como pertencentes à norma culta da língua portuguesa. E, em sendo, pretende-se descrever seu padrão de uso e observar em que gêneros textuais é mais comum a sua presença. Além disso, será possível avaliar se os padrões recorrentes no corpus são os mesmos que aparecem na gramática tradicional e no dicionário.

Essas questões formam a base de investigação deste trabalho, que será sustentado pela análise de uma amostra de um corpus do português do Brasil compilado pelo Núcleo Interinstitucional de Linguística Computacional, doravante corpus NILC, composto por aproximadamente 35 milhões de palavras.

Inicialmente, a pesquisa envolvia o item lexical **nem** comparado com o **não**, pois interessava descobrir qual a diferença de sentido entre um e outro em contexto de negação, conforme fora estudado na disciplina de Pragmática. Porém, ao se fazer a coleta dos dados, percebeu-se que o tema seria muito abrangente.

Desse modo, a pesquisa ficou restrita apenas ao **nem**, com o intuito de descobrir de quais padrões lingüísticos ele fazia parte. Um bom motivo para essa investigação é o fato de que, no *Banco de Português (BP)*, que é um corpus de linguagem geral com 233 milhões de palavras, compilado e mantido na Pontifícia Universidade Católica de São Paulo como parte do projeto *Direct*, dentre as 3.000 mais freqüentes, o **nem** figura em 171º lugar. Portanto, sua ocorrência no português é bem grande. Além disso, a partir da descrição da padronização do **nem**, seria possível fazer outros estudos envolvendo esse item lexical.

Ao se fazer a busca das ocorrências no corpus, despertou a atenção o fato de que os vocábulos que acompanhavam o **nem**, na maioria das vezes, eram: **que** (que nem), **e** (e nem), **sempre**, **mesmo** (nem sempre, nem mesmo).

Novamente fez-se necessário restringir as buscas. Por ser bastante recorrente, optou-se pela descrição do padrão de uso do **que** seguido pelo **nem**.

A grande motivação, porém, centrou-se no interesse de descobrir se o **que nem**, utilizado como locução conjuntiva comparativa (equivalente ao **como**), e até visto com preconceito por ser marca da oralidade, apareceria nos textos extraídos da língua escrita do corpus pesquisado.

Outro fator de interesse centrou-se no fato de que não há estudos sistemáticos sobre esse item lexical, portanto, faz-se necessário descobrir qual é a sua padronização. Na língua, todas as palavras possuem padronização, conforme Berber Sardinha (2004, p. 39), ou seja, há uma regularidade nos tipos de associações a que se submetem essas palavras. Ao se fazer a busca de uma palavra específica, aparecerão as ocorrências da palavra selecionada no corpus de estudo e as palavras que coocorrem

junto com essa forma selecionada. Assim, será possível descobrir qual é o seu padrão de uso e sua colocação.

A análise de concordâncias (listagens das ocorrências de um item específico) e da descrição de padrões (da recorrência de uma colocação, coligação ou estrutura que se repete significativamente) e a informação com base em frequências serão descritas com mais detalhes no corpo do trabalho, visto que fazem parte da metodologia utilizada na pesquisa.

A relevância dessa investigação deve-se, primeiramente, ao fato de que uma pesquisa baseada em corpus pode trazer resultados mais confiáveis frente ao que está sendo analisado, pois, como serão utilizados textos que correspondem a exemplos reais de uso da linguagem, será possível estabelecer qual é o padrão de uso desse item e se um item lexical mais utilizado na oralidade é recorrente em textos de língua escrita.

Além disso, o conhecimento produzido através da pesquisa pode contribuir para o aperfeiçoamento de um sistema capaz de processar línguas humanas e para a reflexão sobre a língua em uso em comparação com o que se menciona sobre ela em obras de referência.

O presente trabalho encontra-se assim estruturado: no capítulo 1, justifica-se a relevância da pesquisa e quais seus objetivos. O capítulo 2 contém o referencial teórico relacionado à Lingüística de Corpus e à maneira como o **que nem** aparece em obras de referências (dicionários e gramáticas). No capítulo 3 encontra-se a metodologia utilizada no trabalho e a descrição das ferramentas utilizadas para a análise lingüística e o capítulo 4 contém os resultados da pesquisa e a análise, seguido pelas considerações finais.

2 REFERENCIAL TEÓRICO

Este trabalho encontra suporte teórico na Lingüística de Corpus (LC). A Lingüística de Corpus é uma área da Lingüística que se dedica à pesquisa através da observação de uma grande quantidade de textos autênticos armazenados eletronicamente (corpus) e analisados através de programas de computação capazes de selecionar, ordenar, contar e calcular. Por textos autênticos estendem-se discursos orais e/ou escritos produzidos por falantes nativos em situações reais de comunicação. (Berber Sardinha, 2004, p.17.)

Neste capítulo serão apresentados os conceitos que envolvem a Lingüística de Corpus, como a noção de corpus nesse contexto, seu histórico, os padrões de linguagem e como o item lexical em análise aparece nas obras de referência.

2.1 Lingüística de Corpus

“A Lingüística de Corpus ocupa-se da coleta e da exploração de corpora, ou conjunto de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por meio do computador”. (Berber Sardinha, 2004, p.3).

Essas evidências empíricas (os dados reais de uso da língua) ganham destaque numa pesquisa em Lingüística de Corpus, conforme McEnery e Wilson (1996, p.1). Para eles, “Corpus linguistics is perhaps best described for the moment in simple terms as the study of language based on examples of ‘real life’ language use”.

Utilizar dados do uso da linguagem da vida real traz resultados mais palpáveis e confere mais credibilidade às pesquisas lingüísticas.

Para se entender melhor a definição de Lingüística de Corpus, faz-se necessário esclarecer o que se entende por corpus nesse contexto, pois nem todo conjunto de dados é um corpus. Há várias definições para esse termo.

2.2 Corpus – definição e características

Todo conjunto de mais de um texto pode ser chamado de corpus. O termo ‘corpus’ vem do Latim e é definido como ‘corpo’. Por esta razão, um corpus pode ser definido como qualquer corpo de texto. Porém, quando o termo corpus é usado no contexto da lingüística moderna, ele terá, freqüentemente, conotações mais específicas do que essa definição mais geral. (McEnery & Wilson, 1996, p. 29).

Para Leech (1997, p.1)

Traditionally, linguistics have used the term **corpus** to designate a body of naturally-occurring (**authentic**) language data which can be used as a basis for linguistics research. This body of data may consist of written texts, spoken discourses, or samples of spoken and/or written language. Often it is designed to represent a particular language or language variety. In the past thirty-five years, the term **corpus** has been increasingly applied to a body of language material which exists in electronic form, and which may be processed by computer for various purposes such as linguistics research and language engineering.

Outra definição, mais completa, é a de Sanchez (1996, p.8-9) apud Berber Sardinha (2004, p.18), segundo a qual um corpus é:

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

A partir dessa definição, depreende-se que um corpus computadorizado deve ser composto de textos autênticos (produzidos por falantes nativos) e ter a finalidade de ser um objeto de estudo lingüístico. Isso significa que ficam excluídos os textos produzidos com fins didáticos ou para análise lingüística, pois não são espontâneos. Devem ser escritos em linguagem natural (e não artificial como a de programação de

computadores), porém devem ser legíveis por computador. O corpus precisa ser criteriosamente escolhido e deve ser vasto para ser representativo da linguagem, de uma variedade lingüística ou de um idioma.

2.3 Lingüística de corpus - breve histórico

De acordo com Kennedy (1998, p.13-19), a pesquisa baseada em corpus já existia antes do surgimento dos computadores. Eram produzidos corpora de citações da Bíblia ou de textos literários. Havia também a produção de dicionários e de gramáticas descritivas. Em boa parte do século XX, houve pesquisadores que se dedicaram à descrição da linguagem por meio de corpora não-eletrônicos (os dados eram coletados, mantidos e analisados manualmente, o que exigia um trabalho intenso e consumia muito tempo). Também eram utilizados corpora no ensino de línguas.

Foi um corpus não-computadorizado que deu feição aos corpora atuais, o SEU (*Survey of English Usage*), compilado por Randolph Quirk e sua equipe, em Londres, em 1959. O SEU, planejado para um milhão de palavras, poderia ser usado como base para descrever a gramática do Inglês Britânico e seu uso por falantes nativos adultos. A transformação completa do *Survey* em corpus eletrônico só foi atingida em 1989, mas a sua parte falada foi computadorizada antes e ficou conhecida como o *London-Lund Corpus*. O SEU foi importante, pois serviu como referência para outros corpora, inclusive o corpus *Brown*, que foi o primeiro corpus lingüístico eletrônico, lançado em 1964, quando houve um impulso no desenvolvimento da Lingüística de Corpus.

Além do corpus *Brown*, do final da década de 1960 em diante, outros corpora eletrônicos foram compilados especialmente para projetos de Lexicografia (ibid., p.33), para fazer dicionários. Também houve corpora para estudar o Inglês falado e para pesquisa sobre aquisição de linguagem. Em 1987, foi publicado o *Cobuild Dictionary*, desenvolvido a partir do Projeto *Cobuild*, que foi o primeiro grande corpus de lexicografia legível por meios eletrônicos.

Outros corpora foram surgindo até a década de 1990 e de acordo com McEnery e Wilson (1996, p.24), “Corpus linguistics is maturing methodologically and

the range of languages addressed by corpus linguists is growing annually.” E eles concluem (ibid., p.195) que “... corpora are challenging our view of what it is possible to study in linguistics and how we should study it. Such challenges should always be welcomed.”

Kennedy (1998, p.294) esclarece essa idéia:

Corpus linguistics has begun to exploit the opportunities for new types of linguistic description provided by the development of computing, with consequences for linguistic description, natural language processing and language pedagogy. The use of computerized corpora as a basis for developing models and descriptions of language and for various applications may prove to be among the most far-reaching achievements of the language sciences in the contribution made to reconciling the confusion of Babel, while at the same time permitting the maintenance of linguistic and cultural diversity within heterogeneous human societies.

Todos esses avanços contribuíram para o desenvolvimento de pesquisas sobre a linguagem, com a utilização de corpora computadorizados, principalmente em países da Europa.

No Brasil, a LC ainda está em estágio inicial, segundo Berber Sardinha (2004, p.6). A pesquisa em corpus se dá em centros mais voltados ao Processamento de Linguagem Natural, à Lexicografia e à Linguística Computacional, já que a história da LC está intimamente ligada à disponibilidade de corpora eletrônicos, que são, na maior parte, em língua inglesa. Na língua portuguesa, há vários corpora eletrônicos de destaque como o Banco de Português, o CETEM, o COMET, o NILC. Para explorar e analisar um corpus eletrônico são necessárias ferramentas computacionais como o *MICROCONCORD* ou o *WORDSMITH TOOLS*, que são usados para se extrair as concordâncias.

Conforme Rocha (2001, p. 144), o uso do concordanceador é a solução mais simples para os usuários do Windows e o WordSmith é uma alternativa barata e fácil de usar. O recurso utilizado resulta em alguma variação no formato da concordância, mas, de um modo geral, a palavra escolhida aparece no centro de um fragmento extraído do centro do corpus. Já o MicroConcord, (Berber Sardinha, 2004, p. 192) é

um programa que roda no Windows, é distribuído gratuitamente na internet, oferece os procedimentos básicos da produção de concordâncias, mas é muito limitado quanto ao número de ocorrências que o programa comporta: cerca de 1.600.

Neste trabalho, foram utilizadas as ferramentas computacionais do *WORDSMITH TOOLS* e o corpus NILC, que serão descritos no capítulo 3.

2.4 Uso do computador

É de conhecimento geral que o computador se faz presente nos mais diversos lugares, principalmente nos ambientes profissionais.

No decorrer da década de noventa, o acesso a recursos computacionais e à *internet* tornou-se suficientemente disseminado para que a questão de sua utilização na educação de um modo geral, passasse a ser uma preocupação de pesquisadores, professores e educadores. Desde então, devido à popularização dos computadores nas universidades, houve um aumento significativo de estudos que se valem de meios eletrônicos como instrumento de análise e armazenamento de dados. Mesmo assim, a parcela da pesquisa lingüística assistida por esses meios ainda é minoritária, pois há falta de conhecimento dos instrumentos disponíveis ou há rejeição ao computador e aos modelos de análise de natureza mais empírica que ele oferece. Além disso, conforme Rocha (2001, p.137), o uso dos computadores nos estabelecimentos de ensino, muitas vezes, sem um planejamento ou *softwares* adequados torna-se, no mínimo, discutível.

Com o devido planejamento e conhecimento dos recursos disponíveis, a utilização de computadores na investigação da linguagem garante ao pesquisador que as operações sejam realizadas com taxa de erro desprezível, se comparadas às que ocorreriam se o processo fosse manual, pois, através de recursos computacionais, é possível contar palavras, buscar e recuperar as ocorrências de um termo específico, listar em ordem decrescente as palavras de um corpus de, por exemplo, um milhão de palavras, em tempo bem menor do que se fosse feito manualmente. Além disso,

através do computador e da tecnologia, há a possibilidade de descoberta de fatos novos da língua ou, também, a modificação no modo de se enxergar a linguagem.

2.5 Lingüística de Corpus: disciplina ou metodologia?

Há um debate na definição do status dessa área. Afinal, ela é disciplina ou metodologia? Não é uma disciplina como as outras subáreas da lingüística (Sociolingüística, por exemplo), pois o seu objeto de pesquisa não é delimitado como em outras áreas, mas sim, ela se ocupa de vários fenômenos comumente enfocados nas outras áreas (léxico, sintaxe, textura) para se fazer pesquisa lingüística.

McEnery e Wilson (1996, p. 2) esclarecem que “Corpus linguistics is not a branch of linguistics in the same sense as syntax, semantics, sociolinguistics and so on. All of these disciplines concentrate on describing aspect of language use. Corpus linguistics in contrast is a methodology rather than an aspect of language requiring explanation or description”.

Para Leech (1992, p. 106) ela é uma nova abordagem filosófica com uma base metodológica para se fazer pesquisa lingüística e argumenta: “...computer corpus linguistics defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject. The computer, as a uniquely powerful technological tool, has made this new kind of linguistics possible”.

Além disso,

Corpus linguistics refers not to a domain of study, but rather to a methodological basis for pursuing linguistic research. In principle (and often in practice) corpus linguistics combines easily with other branches of linguistics: we can study phonetics, syntax, sociolinguistics, and any other aspect of linguistics by means of corpora, and when we are doing this we can be said to be combining techniques of corpus linguistics with the subject-matter of phonetics, syntax, sociolinguistics, and so on. (ibid., p.105)

Hoey, citado por Berber Sardinha (2004, p. 37), considera que a “Linguística de Corpus não é um ramo da linguística, mas uma rota para a linguística.” Assim, ela é uma maneira de se chegar à linguagem, uma abordagem. Por isso, alguns influentes linguistas de corpus como Douglas Biber preferem o termo *abordagem baseada em corpus*, que foi o termo utilizado para esta pesquisa.

Através da metodologia da Linguística de Corpus, com a utilização de recursos computacionais, é possível verificar em um corpus eletrônico quais os usos reais da língua e direcionar o estudo ou a análise para qualquer ramo da linguística que se queira ou que seja mais adequado para os propósitos do pesquisador.

2.6 Tipos de pesquisas privilegiados

Cresce, a cada ano, o número de pesquisadores em trabalhos que se encaixam na Linguística de Corpus e, segundo Kennedy (1998, p.9), são quatro as suas principais áreas de concentração:

- 1- Compilação de corpus, com a coleta e armazenamento de textos;
- 2- Desenvolvimento de ferramentas para análise de corpora;
- 3- Descrição da linguagem, principalmente do léxico e da gramática das línguas;
- 4- Aplicação de corpora (há uma variedade de aplicações, como no ensino de línguas, reconhecimento de voz, tradução).

A área na qual há mais atividade é a terceira, a da descrição. Há um número considerável de trabalhos enfocando, principalmente, o léxico e a gramática a partir de exame de um corpus que fornece contextos para o estudo dos itens lexicais em uso e das estruturas gramaticais. Com isso, há mais facilidade para se trabalhar com investigações linguísticas empíricas.

O estudo da linguagem através da observação de dados (evidências empíricas) tem uma longa tradição em linguística, interrompida nos anos 50, no início da era *chomskyana*, quando foi suplantado pelo estudo da linguagem através da introspecção,

isto é, de dados que eram construídos na base do conhecimento de uma linguagem de um falante nativo.

Com o advento da invenção do computador, a criação de corpora eletrônicos foi possível e isso contribuiu decisivamente para o reaparecimento e fortalecimento da pesquisa lingüística baseada em corpus, já que havia críticas aos corpora manuais, por não serem considerados confiáveis, visto que o número de dados era enorme para que fossem manipulados por humanos.

Além da criação e aplicação de corpora eletrônicos, foram desenvolvidas ferramentas para analisá-los e cresceu o interesse pela descrição da linguagem e seus padrões.

2.7 Padrões de linguagem

Uma das áreas de pesquisa da LC é a da descrição da linguagem, que envolve estudos lexicais e sintáticos. Ao se utilizar corpora e computadores, é possível estudar os padrões de combinação de uma palavra com outra ou outras. Sinclair (1991) foi o precursor desse tipo de trabalho. Ao estudar o comportamento de palavras polissêmicas, ele observou que quando uma palavra tem vários sentidos e é utilizada em diversos padrões, cada um desses padrões ocorrerá com mais frequência com um dos sentidos e menos com os outros, ou seja, é a padronização que permite ao falante de uma língua descobrir com que sentido está sendo utilizada uma palavra. A descrição de uma palavra e de seus padrões não pode ser considerada uma descrição puramente lexical ou puramente gramatical, já que o sentido e a padronização estão associados.

A descrição de padrões já foi estudada por vários autores e segundo Hunston e Francis (2000, p.37, apud Berber Sardinha 2004, p. 39), “Os padrões de uma palavra podem ser definidos como todas as palavras e estruturas com as quais são regularmente associados e que contribuam para seu significado. Um padrão pode ser identificado se uma combinação de palavras ocorre com relativa frequência, se é dependente de uma palavra específica e se há um significado claro associado”.

As questões de que se ocupa a área da descrição de padrões são:

- Quais os padrões lexicais dos quais a palavra faz parte?
- A palavra se associa regularmente com outros sentidos específicos?
- Em quais estruturas ela aparece?
- Há uma correlação entre o uso/sentido da palavra e as estruturas das quais ela participa?
- A palavra está associada com uma certa posição na organização textual? (Hoey, 1997, p.3 – apud Berber Sardinha 2004, p.40).

A padronização se evidencia pela recorrência de uma colocação, coligação ou estrutura que se repete significativamente, demonstrando ser um padrão lexical ou léxico-gramatical (para ser recorrentes, os itens devem estar presentes pelo menos duas vezes). Para se saber se dois ou mais itens lexicais formam um padrão, é necessário saber se a coocorrência é estatisticamente significativa, ou seja, a frequência de determinada ocorrência não pode ser atribuída ao acaso. Para haver coocorrência, os itens devem estar na presença de outros. Eles obtêm significância na medida em que são interpretados como parte de um conjunto formado por outros itens, mas não é necessário que haja aparição sequencial. O horizonte de coocorrência pode ir de algumas palavras ao redor de um item às fronteiras do texto. (Berber Sardinha, 2004, p.90)

Os padrões podem ser formalizados em três conceitos principais ou três tipos de padrões léxico-gramaticais, isto é, que englobam o vocabulário e a gramática (sintaxe):

- 1- **Colocação:** refere-se à associação entre itens lexicais ou entre o léxico e campos semânticos. Firth (1957 apud Stubbs 1996, p. 173), propôs um método para estudar a distribuição das palavras, incluindo sua tendência de coocorrer com outras e criou uma famosa definição para colocação. Segundo ele, colocação é “the company a word keeps” e acrescenta “collocations are actual words in habitual company”. Em português, traz a idéia do “*diga-me com quem andas e eu te direi quem és*”.

Conforme Hunston (2002, p.68), “**Collocation** is the tendency of words to be biased in the way they co-occur. For example, the word *toys* co-occurs with *children* more frequently than *women* or *men*. This collocation might be said to be motivated, in that there is a logical explanation for it (toys belong to children, on the whole, rather than to adults).”

- 2- **Coligação:** é a associação entre itens lexicais e gramaticais. Por ex., *start* é mais comum com sintagmas nominais e orações *-ing*, enquanto *begin* é mais usado com um complemento *to*, de acordo com Biber et al. (1998). Ambos têm o mesmo sentido, mas um item lexical é seguido por verbo no infinitivo, como *begin to work* e o outro, por verbo no gerúndio, como *start working* (*começar a trabalhar*).

- 3- **Prosódia semântica:** “associação entre itens lexicais e conotação (negativa, positiva ou neutra) ou instância avaliativa. Certas palavras preparam o ouvinte, ou o leitor, para o conteúdo semântico que está por vir, da mesma maneira que a prosódia na fala indica para o interlocutor que tipos de sons estão por vir (Hoey, 1997). Por ex., *cause* tem uma prosódia semântica negativa, pois se associa a palavras desfavoráveis como *problem (s)*, *damage*, *death (s)*, *concern* e *câncer*”.

Dentre esses conceitos, o fenômeno da colocação é o mais tradicionalmente enfocado no estudo de corpus e ele se fundamenta na análise da palavra dentro de um pequeno contexto (espaço de texto), geralmente quatro palavras para cada lado do item lexical de interesse. A utilização de um pequeno espaço de texto se justifica, pois, na década de 1960, quando surgiu o estudo lingüístico através de corpus eletrônico, era inviável analisar a palavra num espaço maior de texto, já que a tecnologia da época não era tão avançada.

Neste trabalho, haverá enfoque nos padrões de colocação e de coligação, pois interessa saber a quais campos semânticos ou a qual item lexical o **que nem** se associa e, também, com qual item gramatical ele aparece comumente.

2.8 *Que nem* nas obras de referência: dicionários e gramáticas

Os materiais selecionados para a análise foram:

a) dicionários:

- Novo Aurélio (1986)
- Houaiss (2001)

b) gramáticas:

- *Gramática Metódica da Língua Portuguesa* – Napoleão Mendes de Almeida (1982)
- *Novíssima Gramática da Língua Portuguesa* - Domingos Paschoal Cegalla (1989)
- *Moderna Gramática Portuguesa* – Evanildo Bechara (2003)
- *Nova Gramática do Português Contemporâneo* – Celso Cunha e Lindley Cintra (1985)
- *Nossa Gramática – teoria e prática* - Luiz Antonio Sacconi (1999)

Entende-se por gramática, neste trabalho, a definição dada por Perini (1999), como um manual de regras do bom uso da língua que devem ser seguidas. Essa gramática é normalmente rotulada de Gramática Tradicional (GT).

c) outras obras de referência que tratam do assunto.

2.8.1 *Que nem* - de acordo com o dicionário

No dicionário **Novo Aurélio** (1986, p. 1427), **que nem** assume os seguintes significados, conforme acepção do que: **Que nem**: que só; tal qual; como se fosse; do mesmo modo. Ex.: Saiu que nem o pai.

Quanto ao **nem** (p.1187) [do lat.nec] conjunção:

1. **E não:** *Não conheço senão fatos sem causas, nem procurei conhecê-las.* (Thiers Martins Moreira, Os Seres, p.63)
2. **E sem:** *Ficou sem amigo nem dinheiro.*
3. **Ao menos, pelo menos; sequer:** *A minha vida se resume, desconhecida e transitória, / em contornar teu pensamento, / sem levar dessa trajetória, / nem esse prêmio de perfume/ que as flores concedem ao vento.* (Cecília Meireles, Obra Poética, p.103)
4. **Ou.** *É o crime mais negro que se pode cometer nem conceber.*

Segundo o dicionário **Houaiss** (2001, p. 2350 e 2006), os significados são os seguintes:

- a) **que:**relacionado ao **que nem:** conjunção comparativa – expressando informalmente - igualdade: (é teimoso *que nem* o pai).
- b) **nem:** relacionado ao **que nem** assume o significado de: do mesmo modo que; como (*estúpido que nem uma galinha*) (*falava que nem um francês*).

As demais acepções do **nem**, segundo o Houaiss (2001, p. 2006) são:

nem: conj. **1.** coord. serve para ligar palavras e orações negativas, indicando: **1.1** adt. conexão, ligação **1.1.1** e não, e sem, e ninguém, e nunca (*não chorou nem reclamou da dor*) (*ficou só, sem pai nem mãe*) (*ninguém sabe o dia de amanhã nem pode adivinhar*) (*nunca saiu do Brasil nem de sua cidade natal*) **1.2** alt. alternância: ou, ou...ou (*não foi à festa nem ao cinema; ficou em casa*) (*não o abalaram nem as lágrimas nem as ameaças*). adv. **2** exprime negação: **2.1** não (*nem pense em fazer isso*) **2.2** sequer, ao menos (*nem um bocado de pão havia na mesa*) * **nem mais nem menos:** exatamente ***nem que:** mesmo que, apesar de (*nunca iria deixá-lo, nem que fosse necessário*) ***que nem:** do mesmo modo que; como (*estúpido que nem uma galinha*) (*falava que nem um francês*) * Uso como adv. pode-se empregar sem o verbo: *Para cozinhar,*

nem sal, nem tempero. * ETIM lat. nec ‘nem, não, e não’, com apócope do –c e prov. influência da nasal inicial; f. hist. 1041 *nem*, sXIII *nen*, sXIII *n~e*, sXIV *nem*.

Pôde-se perceber que o dicionário Houaiss está mais completo e esclarecedor quanto à função da palavra e quanto ao sentido.

2.8.2 *Que nem* - de acordo com a Gramática Tradicional

A gramática tradicional apresenta o item lexical **que nem** como locução conjuntiva comparativa que equivale ao **como**. Não há outras classificações para ele, a não ser que seja observado separadamente o **que** e o **nem**.

O **que** não tem uma classificação única. Sacconi (1999, p. 539) reserva um capítulo à parte para mostrar que existem várias formas de analisá-lo. Ele pode ser substantivo, interjeição, advérbio, pronome, conjunção, expletivo ou preposição. No dicionário, ele também recebe essas classificações. Interessa para esta pesquisa o **que** pronome relativo ou conjunção, pois, conforme observado nos dados, como o **que** aparece no meio da oração e seguido pelo **nem**, ora ele será um ou outro.

Conforme Sacconi, o pronome relativo substitui substantivo ou pronome substantivo e pode exercer funções sintáticas as mais diversas. Ex. *Cão **que** ladra não morde.* (**que** = sujeito de **ladra**).

O **que** pode ser conjunção coordenativa ou subordinativa e vai ligar sintaticamente orações dependentes ou independentes. Ex. *Dizem **que** vai gear.* (**que** é conjunção subordinativa integrante; o verbo **dizer** é transitivo direto.)

Quanto ao **que nem**, locução conjuntiva comparativa, não há nenhuma menção nessa gramática.

Na *Gramática Metódica da Língua Portuguesa*, de Napoleão Mendes de Almeida (1982, p. 348), há referência ao **que nem** no capítulo que trata da classificação do **nem**, conforme exposto a seguir:

? conjunção aditiva: liga duas orações, aproximando-as meramente.

- a) **nem**, quando aditivo, supõe, antes, uma oração negativa e equivale analiticamente a “e não”: “*Não foi nem (= e não) deixou que outros fossem.*” Como a conjunção **nem** já equivale a “e não”, é hoje condenada a anteposição do **e** ao **nem**: “*Não foi e nem deixou...*” – Só é possível dizer “e nem” quando o **nem** não exerce função coordenativa, como nestes exemplos: “*Não foi e, nem que tivesse ido, não...*” – “*Ele não foi e nem por isso faltou à obrigação*” – “*Corriam alegres para a escola e nem sequer dos brinquedos de casa se lembravam*” - “*E nem da própria vida estou seguro*”.
- b) Quando repetido, o **nem** implica separação de idéias; diz-se então conjunção alternativa: “*Nem um, nem outro*” – “*Nem para trás, nem para diante*”.
- c) **Nem** desempenha às vezes função adverbial: “*Nem por sombra*” - “*Nem por isso*” - “*Nem tudo é bom*” - “*Nem sempre*”.
- d) A locução conjuntiva **que nem**, de significação igual a *como*, só é usada hoje por pessoas incultas: “*Ele caiu que nem uma pedra*”.

Esta última classificação é a que mais nos interessa, visto que a ênfase desta pesquisa centra-se nesta forma popular de se exprimir lingüisticamente a comparação. Sabe-se que essa é uma forma bastante utilizada pelas pessoas, sobretudo na língua falada. É possível que, na época da publicação dessa gramática, essa classificação fosse válida, apesar de discriminatória, mas hoje, com a evolução dos estudos lingüísticos e a tentativa de diminuição do preconceito acerca, inclusive, da linguagem, dizer que essa forma só é usada por *pessoas incultas* é, no mínimo, ofensivo.

Na *Novíssima Gramática da Língua Portuguesa*, de Domingos Paschoal Cegalla (1989, p.246), o **que nem** está presente no rol das conjunções subordinativas

comparativas: **que nem** (= como, do mesmo modo que). Por ex. *Os pedestres se cruzavam pelas ruas que nem formigas apressadas.*

Nessa mesma gramática (ibid., p.341), o autor considera que **em estruturas de cunho popular** (grifo meu) como: Ficou vermelho *que nem* brasa. *É forte como o diabo.* (Aurélio). *Luzia que só espelho!* (Luís Jardim), esses elementos devem ser analisados como adjuntos adverbiais de comparação e não como orações adverbiais comparativas. O autor apenas faz essa citação, mas não há maiores esclarecimentos.

Percebe-se que embora o **que nem** seja apresentado como um item lexical de cunho popular, ele não é visto de forma preconceituosa, mas aceito como qualquer outra conjunção, diferentemente da gramática de Almeida que destaca que essa expressão só é usada por pessoas incultas.

Bechara (2003, p.493) esclarece a noção de locuções conjuntivas comparativas, porém cita vagamente o **que nem**. Segundo ele, essas locuções podem ser assimilativas ou quantitativas, conforme exposto a seguir:

Comparativas: quando a subordinada exprime o ser com que se compara outro ser da oração principal.

A comparação é *assimilativa*, “quando consiste em assimilar uma coisa, pessoa, qualidade ou fato a outra mais impressionante, ou mais conhecida”.

É introduzida a oração subordinada comparativa desta espécie por *como* ou *qual*, podendo ainda estar em correlação com *assim* ou *tal* postos na oração principal:

“*Os importunos são como as moscas que, enxotadas, reverterem logo.*”

A comparação pode ainda ser *quantitativa*, quando “consiste em comparar, na sua quantidade ou intensidade, coisas, pessoas, qualidades ou fatos”.

Há três tipos de comparação quantitativa:

a) igualdade: introduzida por *como* ou *quanto* em correlação com *tanto* ou *tão* da oração principal, ou o *mesmo que*: “Nada conserva e resguarda *tanto* a vida *como* a virtude”; “Isto é o *mesmo que* nadar em ouro”.

b) superioridade: introduzida por *que* ou *do que* em correlação com *mais* da oração principal: “Um homem pode saber *mais do que* muitos, porém nunca tanto como todos”.

c) inferioridade: introduzida por *que* ou *do que* em correlação com *menos* da oração principal: “O governo dos loucos dura pouco, o dos tolos ainda *menos que* o dos velhacos.”

Quando se refere ao **como** (tido como elemento de comparação) é que esclarece: “Em vez de *como, do mesmo modo que, tanto como*, empregamos com frequência *que nem*: É forte *que nem* um touro.” (ibid., p. 496)

Essa é a única referência ao **que nem** na gramática de Bechara e, o fato de dizer que ‘empregamos com frequência o *que nem*’, generaliza o enunciado e não esclarece *quem* exatamente usa e em que situações.

Ainda em Bechara (2003, p.319), no capítulo referente às conjunções, mais especificamente às conjunções comparativas, não há nenhuma citação sobre esse item lexical.

Quanto ao **nem**, é classificado como conjunção aditiva, ou conector aditivo, usada para a adição de unidades negativas. Ex. *Ele não é rico nem honesto.*

Em Cunha e Cintra (1985, p. 574), o **que nem** faz parte do grupo de conjunções subordinativas comparativas, as quais iniciam uma oração que encerra o segundo membro de uma comparação, de um confronto: *que, do que* (depois de *mais, menos, maior, menor, melhor e pior*), *qual* (depois de *tal*), *quanto* (depois de *tanto*), *como, assim como, bem como, como se, que nem* (grifo meu).

Ex. Ele comeu-a **que nem** confeitos. (C. Castelo Branco, OS, I, 368)

Essa é a única referência ao **que nem** nessa gramática.

Neves e Hattner (2002, p. 125) criticam a Gramática Tradicional por não darem um tratamento cuidadoso às construções comparativas, pois são mencionadas as “orações subordinadas adverbiais comparativas” entre as adverbiais, mas é apenas na

parte destinada à morfologia, no capítulo que apresenta o grau de adjetivos, que alguma indicação do mecanismo de comparação é oferecida. Esta restrição teria justificativa se, pelo menos, a língua portuguesa apresentasse sufixo comparativo, mas, como esse não é o caso, fica absolutamente inconseqüente comporem-se os esquemas: *mais/menos... (do) que e tão/tanto... como/quanto*, como se eles só se aplicassem a adjetivos, pois é o que prescreve a gramática.

Partindo desse pressuposto, imagina-se que o **que nem** enquanto comparativo, esteja diretamente relacionado a adjetivos, ou seja, espera-se que um dos padrões mais encontrados na pesquisa seja justamente o do uso do **que nem** com adjetivos.

Observações: Nas gramáticas e nos dicionários, de modo geral, o **que nem** é classificado como locução conjuntiva comparativa, ou seja, possui uma classificação morfossintática – de relação entre as classes morfológicas e as funções sintáticas exercidas pelas palavras - relação do adjetivo com o substantivo, com o verbo. Essa classificação ora é superficial ora está relacionada com a modalidade oral e ora aparece como qualquer outra locução conjuntiva comparativa. Isso demonstra que nessas obras não há um consenso em relação à maneira como ela é utilizada. O que destoa ainda mais é a classificação dada por Cegalla: a de adjunto adverbial de comparação, que teria uma função sintática, como a do sujeito, objeto, adjunto adnominal.

2.8.3 *Nem* em outras obras de referência

Nem como operador argumentativo:

Segundo Koch (1992, p. 30), o termo *operadores argumentativos* foi cunhado por O. Ducrot, criador da Semântica Argumentativa (ou Semântica da Enunciação), para designar certos elementos da gramática de uma língua que têm por função indicar (“mostrar”) a força argumentativa dos enunciados, a direção (sentido) para o qual

apontam. Para explicar seu funcionamento, Ducrot utiliza duas noções básicas: as de *escala argumentativa* e *classe argumentativa*. Uma *classe argumentativa* é constituída de um conjunto de enunciados que podem igualmente servir de argumento para (apontam para) uma mesma conclusão (a que, por convenção, se denomina **R**).

Quando dois ou mais enunciados de uma classe se apresentam em gradação de força crescente no sentido de uma mesma conclusão, tem-se uma *escala argumentativa*.

Cada operador argumentativo é utilizado com um determinado propósito, para introduzir determinado tipo de argumento. Interessa, para essa pesquisa, como é o funcionamento do operador argumentativo **nem**, além do **mesmo**, **sequer**.

Para Koch (1993, p.104), o uso da linguagem é inerentemente argumentativo. Ela se fundamenta na tese defendida por Ducrot, Anscombe e Vogt de que a argumentatividade não constitui apenas algo acrescentado ao uso lingüístico, mas, pelo contrário, está inscrita na própria língua.

Ainda segundo Koch (1992, p.29),

Quando interagimos através da linguagem (quando nos propomos a jogar o “jogo”), temos sempre objetivos, fins a serem seguidos; há relações que desejamos estabelecer, efeitos que pretendemos causar, comportamentos que queremos ver desencadeados, isto é, pretendemos *atuar* sobre o(s) outro(s) de determinada maneira, obter dele(s) determinadas reações (verbais ou não verbais). É por isso que se pode afirmar que o uso da linguagem é essencialmente argumentativo: pretendemos orientar os enunciados que produzimos no sentido de determinadas conclusões (com exclusão de outras). Em outras palavras, procuramos dotar nossos enunciados de determinada força argumentativa.

São os operadores argumentativos (termo criado por O. Ducrot) que determinam o valor argumentativo dos enunciados, constituindo-se, assim, em marcas lingüísticas importantes da enunciação.

Koch (1993, p.105) mostra que a gramática tradicional considera alguns deles como elementos meramente relacionais – **conectivos**, como **mas**, **porém**, **já que**, **pois**, etc., e outros, como vocábulos que, segundo a N.G.B., não se enquadram em nenhuma das classes gramaticais. Rocha Lima chama-as de **palavras denotativas** e

Bechara de **denotadores de inclusão, de exclusão, de retificação...** ou **palavras** “essencialmente afetivas” para Celso Cunha.

A semântica argumentativa retoma esses elementos, por serem justamente eles que determinam o valor argumentativo dos enunciados e, para examiná-los, é conveniente retomar a noção de **escala argumentativa** formulada por Ducrot. Diz-se que **p** é um argumento para a conclusão **r**, se **p** é apresentado como devendo levar o interlocutor a concluir **r**. Quando vários argumentos – **p**, **p'**, **p''**... – se situam numa escala graduada, apontando, com maior ou menor força, para a mesma conclusão **r**, diz-se que eles pertencem à mesma escala argumentativa (Koch, 1993). Por exemplo:
r = Pedro é um político ambicioso.

| | | |
|--------------|-------------------|------------|
| Ele quer ser | até | presidente |
| | | governador |
| | pelo menos | prefeito |

Duas ou mais escalas orientadas no mesmo sentido, isto é, para uma mesma conclusão, constituem uma classe argumentativa.

Certos operadores estabelecem a hierarquia dos elementos numa escala, assinalando o argumento mais forte para uma conclusão **r** (**mesmo, até, até mesmo, inclusive**) ou, então, o mais fraco (**ao menos, pelo menos, no mínimo**), deixando, porém, subentendido que existem outros mais fortes, como no exemplo dado.

Havendo duas ou mais escalas orientadas no mesmo sentido, seus elementos podem ser encadeados por meio de operadores como **e, também, NEM** (grifo meu), **tanto... como, não só... mas também, além de, além disso, etc.**

Dependendo da intenção do falante e de que tipo de argumento ele queira introduzir, haverá um propósito diferente para utilizar este ou aquele operador.

Interessa-nos, especialmente, como é o funcionamento do operador argumentativo **nem**. Segundo Koch (1993, p.109, 173, 175, 231, 212):

- o **nem** é um operador argumentativo que seleciona o argumento mais forte da escala argumentativa orientada para uma conclusão **R** (sentido do vazio – negação universal: *nada, nenhum*).

- o **nem sempre** é um operador que marca a frequência da negação da qualidade expressa no enunciado.

- o **nem mesmo** indica o mínimo numa escala orientada no sentido do vazio (negação universal: *nada, nenhum*). Na escala de sentido negativo, o argumento mais forte vem introduzido por ele. Equivale ao **sequer** ou ao **nem sequer**.

- **nem assim** = equivale ao **mesmo assim não...** é um operador de valor concessivo.

Obs. Quando se tem escalas orientadas no sentido da afirmação plena (universal afirmativa: *tudo, todos*) ou da negação plena (universal negativa: *nada, nenhum*), os quantificadores selecionam determinados operadores capazes de dar seqüência ao discurso. (Koch,1993, p.109)

Conforme Ducrot (1977, p.152), há uma regra geral segundo a qual, *no nível do componente lingüístico*, os enunciados que comportem *mesmo* significam (além de outras coisas) tudo aquilo que significam sem *mesmo*.

O que ocorre é que se houver a utilização do *mesmo*, haverá um sentido de inclusão e de ênfase no enunciado que se tornará mais evidente se esse *mesmo* estiver precedido pelo *nem* (**nem mesmo**).

A negação **nem mesmo** afeta a escala argumentativa como um todo, no sentido que cria uma outra escala, de acordo com Ilari e Geraldi (2003, p.81). Afinal, o argumento mais forte vem introduzido por essa negação. Logo, subentende-se que haverá outra escala com argumentos mais fracos.

Kock (1992, p.39) conclui que os operadores argumentativos fazem parte da gramática da língua, mas têm merecido pouca atenção nos livros didáticos e nas aulas de língua portuguesa, (já que tradicionalmente são descritos como “meros elementos de relação, destituídos de qualquer conteúdo semântico”), porém são os responsáveis, em grande parte, pela força argumentativa dos textos.

Em relação ao **que nem** (locução conjuntiva comparativa), Koch (1993, p.196) o cita como uma expressão popular que pode substituir o ‘como’, ‘assim como’, ‘tanto quanto’. Esses elementos são operadores de comparação e, portanto, exercem

algum poder argumentativo. Segundo a autora, quando dois termos são comparados, mesmo em se tratando de comparativo de igualdade, evidencia-se a presença da oposição entre os dois termos, porque a compensação argumentativa opera a diferença na igualdade.

Vogt (1977, p.202) aprofunda esses conceitos ao fazer um estudo sobre a comparação. Ele afirma que:

o comparativo de igualdade se apresenta como um equilíbrio passageiro entre dois pesos de uma balança e que, por isso mesmo, se sustenta por uma oposição entre os dois termos comparados. Tal oposição, que é dinâmica, tanto pode realizar-se ao nível das conclusões que se podem tirar de cada um dos termos, ou ao nível dos próprios termos, enquanto argumentos equivalentes para uma mesma conclusão.

Ainda segundo Vogt, quando o termo comparado é tomado como um tema, o enunciado comparativo constitui um argumento favorável (argumentatividade forte) à existência da propriedade que lhe é atribuída: “**A** é tão **p** quanto **B**” enuncia uma certa igualdade de **A** e **B** relativamente a **p**, mas de modo a apreciar **A** como provavelmente **mais** e **B** como necessariamente **menos**.

A respeito do operador *como*, enquanto marca específica no comparativo de igualdade, Vogt (1977, p.198-199) considera que, mesmo no que concerne às expressões proverbiais, não deve haver exclusividade desse operador.

Assim, em português, para todas as expressões do tipo *doce como mel*, *feio como assombração*, *forte como um touro*, etc..., que são, de toda a evidência, expressões feitas e de tipo proverbial e onde o segundo termo deve ser considerado como exemplar quanto à qualidade comparada, podem-se encontrar expressões paralelas, e de uso muito comum, do tipo: *doce que nem mel*, *feio que nem assombração*, *forte que nem um touro*, *esperto que nem macaco*, etc.

Vogt esclarece que:

Essas expressões podem ser facilmente compreendidas como equivalendo a:

doce que nem mesmo mel é doce
feio que nem mesmo assombração é feia
forte que nem mesmo um touro é forte, etc.

Isto é, *nem* compreenderia, de alguma forma, um operador *mesmo*. O valor desse operador é hierarquizar argumentos, dando exclusividade àquele sobre o qual opera.

O comparativo *que nem*, evidencia, na igualdade, a presença da oposição (pelo valor negativo de *nem*) entre os dois termos comparados, aproximando-a das outras duas formas de comparativo. A combinação de *que nem mesmo* mostra, com toda força, que a igualdade já é sintoma de uma diferença, argumentativamente falando, porque agora tais enunciados significam necessariamente:

mais doce que mel
mais feio que assombração
mais forte do que um touro, etc.

Ainda segundo Vogt, essas expressões, mesmo tendo caráter proverbial e ainda que o segundo termo realize exemplarmente a qualidade comparada, nem por isso deixam de realizar o dinamismo argumentativo da compensação entre o termo comparado e o termo comparante.

É passível de observação que a igualdade argumentativa contenha já o elemento negativo próprio da adversidade ou da disjunção entre os dois termos relativos à propriedade comparada. Essa igualdade pode formalmente ser representada pela expressão *que nem*, de uso comum, na qual a negatividade aparece de forma evidente. (Vogt, 1977, p.204)

Esse tipo de comparação é também discutido por Moura (2002, p.79):

Consideremos por exemplo a forma com “que nem”. Comparar é avaliar dois ou mais indivíduos ou coisas sob a perspectiva de uma certa propriedade. Mas a comparação serve também para a escolha de opções. Ao comparar objetos sob uma certa ótica, visamos determinar a escolha de um ou mais objetos dentre uma série de objetos. A determinação de uma propriedade relativa a um certo objeto leva ao julgamento desse objeto no nosso sistema de crenças.

Qual é o processo da comparação popular “que nem”? Ela simultaneamente compara e formula um julgamento sobre o termo comparado.

Uma sentença como: “O aluno é esperto que nem o professor.”, normalmente não aparece nas gramáticas (foi inventada pelo autor), mas é uma sentença boa como raciocínio, pois o “que nem” não pode ser levado ao pé da letra, não pode simplesmente ser substituído pelo “como”, porque ele terá interpretações de acordo com as opções e intenções do falante. A intenção do falante, quando é o caso, projeta sobre o sentido inicial um outro sentido mais apropriado, segundo Moura (2002, p. 79).

Vejam os quais seriam esses sentidos. Numa primeira interpretação, há uma comparação de superioridade: “o aluno é esperto como nem o professor é”, ou seja, o aluno é mais esperto que o professor. Numa segunda interpretação, a partícula “nem” nega o atributo de esperteza ao professor: “o aluno é esperto como o professor não é”. Novamente interpreta-se que o aluno é mais esperto do que o professor ou que o professor não é esperto. É claro que não está se querendo dizer que o professor não é esperto, mas sim, o que se quer enfatizar é a esperteza do aluno. Além disso, o contraste entre a esperteza do aluno e do professor (nem o professor é tão esperto) visa ressaltar a esperteza do aluno, sem negar a esperteza do professor.

Enfim, uma partícula de comparação tão popular como o *que nem* suscita várias interpretações, ou seja, é rica de significados e cumpre sua função comunicativa, mas a gramática tradicional não valoriza: simplesmente exclui ou não comenta essas construções. Ela aceita construções como: “O aluno é esperto **como** o professor”, mas deixa de lado o tipo de comparação utilizada oralmente, como: “O aluno é esperto **que nem** o professor”. Portanto, valoriza-se a norma culta da língua, aquela que os gramáticos consideram mais representativa e correta, em detrimento das falas populares.

O que se observa em situações como essa é que, muitas vezes, as pessoas utilizam as palavras (até as mais banais) e, por desconhecimento ou outros motivos, não percebem a força argumentativa que elas têm ou o efeito que elas podem produzir. Outras vezes, as pessoas as utilizam intencionalmente, pois sabem qual será o efeito desejado e até poderão conseguir alguma vantagem com isso. O **que nem** é um exemplo desses tipos de usos.

2.9 Gêneros textuais

Nesta pesquisa pretende-se, também, descobrir em qual gênero textual é mais recorrente o uso do **que nem** – locução conjuntiva comparativa. Para isso, convém esclarecer quais são as noções de gênero.

Segundo Bakhtin (2000, p. 279), todas as esferas da atividade humana, por mais variadas que sejam, estão sempre relacionadas com a utilização da língua que se efetua em forma de enunciados (orais e escritos), concretos e únicos, que emanam dos integrantes duma ou doutra esfera da atividade humana. Qualquer enunciado considerado isoladamente é individual, mas cada esfera de utilização da língua elabora seus *tipos relativamente estáveis* de enunciados, que são denominados *gêneros do discurso*.

A relação entre os gêneros do discurso e texto, segundo Bakhtin (1985:295) apud Rodrigues (2001) é que o texto é concebido como enunciado, pois é o ponto de partida para o estudo do homem social e da sua linguagem. Nessa concepção, se se considera o texto como enunciado, os gêneros do discurso constituem-se como tipos históricos relativamente estáveis de texto. Nesse caso, pode-se falar de gêneros textuais como tipos históricos de textos.

Em suma, aquilo que Bakhtin chama de gêneros do discurso, Marcuschi (2003, p.22,23) chamará de gêneros textuais, que são tipos históricos de texto. Ele apresenta uma distinção entre gêneros e tipos textuais que é fundamental em todo trabalho com a produção e compreensão textual. Ainda de acordo com Marcuschi (2003), Entre os autores que defendem uma posição similar à que será exposta, estão Douglas Biber, John Swales, Jean-Michel Adam e Jean-Paul Bronckart. A definição das duas noções é a seguinte:

- a) Usamos a expressão *tipo textual* para designar uma espécie de seqüência teoricamente definida pela *natureza lingüística* de sua composição (aspectos lexicais, sintáticos, tempos verbais, relações lógicas). Em geral, os *tipos textuais* abrangem cerca de meia dúzia de categorias conhecidas como: *narração, argumentação, exposição, descrição, injunção*.

- b) Usamos a expressão *gênero textual* como uma noção propositalmente vaga para referir os *textos materializados* que encontramos em nossa vida diária e que apresentam *características sócio-comunicativas* definidas por conteúdos, propriedades funcionais, estilo e composição característica. Se os tipos textuais são apenas meia dúzia, os gêneros são inúmeros. Alguns exemplos de gêneros textuais seriam: *telefonema, sermão, carta comercial, carta pessoal, romance, bilhete, reportagem jornalística, aula expositiva, reunião de condomínio, notícia jornalística, horóscopo, receita culinária, bula de remédio, lista de compras, cardápio de restaurante, instruções de uso, outdoor, inquérito policial, resenha, edital de concurso, piada, conversa espontânea, conferência, carta eletrônica, bate-papo por computador, aulas virtuais* e assim por diante.

Além dos gêneros amplamente divulgados e conhecidos, pode-se perceber, nesses exemplos, que surgiram novos gêneros textuais nos últimos anos, graças às novas tecnologias, em especial às ligadas à área da comunicação. Instaurou-se, também, uma nova relação com os usos da linguagem. Em certo sentido, isso possibilitou a redefinição de alguns aspectos centrais na observação da linguagem em uso como, por exemplo, a relação entre a oralidade e a escrita, desfazendo ainda mais as suas fronteiras.

A abordagem sobre gêneros e a relação entre a oralidade e a escrita se faz necessária, visto que o que está em discussão é justamente como um elemento da oralidade (**que nem = como**) se faz presente em textos extraídos da língua escrita. Questiona-se, também, se esses textos escritos também fazem parte da oralidade.

Marcuschi (2003, p.33) discute essa questão, mostrando que há várias possibilidades de observação dos gêneros textuais. Em especial, destaca que é preciso ter cautela com a questão da relação oralidade e escrita no contexto dos gêneros textuais, porque eles se distribuem pelas duas modalidades (orais e escritas) num contínuo, desde os mais informais aos mais formais e em todos os contextos e situações da vida cotidiana. E esclarece:

Mas há alguns gêneros que só são recebidos na forma oral apesar de terem sido produzidos originalmente na forma escrita, como o caso das notícias de televisão ou rádio. Nós ouvimos aquelas notícias, mas elas foram escritas e são lidas (oralizadas) pelo apresentador ou

locutor... As jaculatórias, novenas e ladainhas, embora todas tenham sido escritas, seu uso nas atividades religiosas é sempre oral.

Essas características dos gêneros os tornam fenômenos bastante heterogêneos e por vezes híbridos em relação à forma e aos usos.

Segundo Biber (1988) apud Marcuschi (2003, p.34), os gêneros são geralmente determinados com base nos objetivos dos falantes e na natureza do tópico tratado, sendo assim uma questão de uso e não de forma. Por exemplo, o uso do gênero *piada* não condiz com determinados eventos como uma reunião.

Em suma, pode-se dizer que os gêneros textuais fundam-se em critérios externos (sócio-comunicativos e discursivos), enquanto os tipos textuais fundam-se em critérios internos (lingüísticos e formais).

Marcuschi (2001, p.44) propõe três conjuntos de gêneros textuais: os tipicamente orais; os tipicamente escritos; e os produzidos na interface dos orais e dos escritos. Ele cita como exemplo desse terceiro as entrevistas das páginas amarelas da Revista VEJA, que se constituem de um textos oralmente produzidos, mas transmitidos pela via escrita. O resultado é que as relações entre oralidade e escrita são profundamente imbricadas, ficando as diferenças por conta da natureza das práticas desenvolvidas que determinam os gêneros textuais como eventos sócio-comunicativos situados, culturais e históricos.

Nesta mesma linha da junção entre o oral e o escrito, Hoffnagel (2003, p.180-188) examina o gênero *entrevista* e a considera como uma constelação de eventos possíveis que se realizam como gêneros (ou sub-gêneros) diversos, pois existe a entrevista jornalística, entrevista médica, entrevista científica, entrevista de emprego. Embora tenham propósitos diferentes, elas têm em comum a estrutura marcada por “perguntas e respostas”. Seguem essa mesma estrutura o inquérito policial e o “exame oral” em que o professor pergunta e o aluno responde, mas os objetivos e a natureza dos atos praticados são diferentes.

Além disso, a entrevista é um gênero primordialmente oral. Quando publicada em jornais e revistas, na maioria das vezes, a entrevista foi feita oralmente e depois transcrita para a publicação. Quanto aos objetivos, toda entrevista pretende obter informações. O tipo e uso destas informações podem servir a vários fins, dependendo

do público-alvo, que é diferente. A linguagem usada nas entrevistas varia de acordo com o propósito, o tópico da entrevista e o público-alvo. Nas revistas dirigidas aos jovens que entrevistam artistas, a linguagem é coloquial, tanto na parte do entrevistador como do entrevistado, pois o efeito no leitor é de que ele está compartilhando uma conversa íntima entre pessoas do mesmo mundo. Nas entrevistas com autoridades ou especialistas, a linguagem é mais formal, ‘objetiva’ e isso não depende só da revista, por exemplo, mas de quem está sendo entrevistado.

Pode-se perceber que o gênero *entrevista*, apesar de ser sempre constituído pelo entrevistador e pelo entrevistado, é um gênero marcado por certa liberdade de expressão tanto por parte de um, como do outro. Essa liberdade se reflete tanto nas respostas ou perguntas como na linguagem utilizada.

Outro fator importante para se observar no que diz respeito aos gêneros é o fator da individualidade. Para Bakhtin (2000, p.283), o enunciado (texto) – oral e escrito, primário (simples) e secundário (complexo), em qualquer esfera da comunicação verbal – é individual, e por isso pode refletir a individualidade de quem fala (ou escreve). Em outras palavras, possui um estilo individual. Mas nem todos os gêneros são igualmente aptos para refletir a individualidade na língua do enunciado, ou seja, nem todos são propícios ao estilo individual. Os gêneros mais propícios são os literários – neles o estilo individual faz parte do empreendimento enunciativo enquanto tal e constitui uma das suas linhas diretrizes. As condições menos favoráveis para refletir a individualidade na língua são as oferecidas pelos gêneros do discurso que requerem uma forma padronizada, tais como a formulação do documento oficial, da ordem militar, da nota de serviço.

3 METODOLOGIA

3.1 Corpus da pesquisa

Neste trabalho, serão descritos padrões de uso do item lexical **que nem**, em especial da locução conjuntiva comparativa, através da *abordagem com base em corpora lingüísticos*, a fim de delimitar os contextos lingüísticos em que tal forma aparece.

A concordância (lista de ocorrências desse item) foi extraída do corpus do NILC - Núcleo Interinstitucional de Lingüística Computacional – órgão criado em 1993, a partir da formação de uma equipe multidisciplinar para o desenvolvimento de um revisor gramatical automático, num projeto de parceria entre a Itaotec-Philco e a Fundação de Apoio à Física e à Química (FAFQ), do campus da USP de São Carlos.

O corpus de português do Brasil, compilado pelo NILC, contém cerca de 35 milhões de palavras, constituído de textos em prosa (textos em versos não constam do corpus), predominantemente jornalísticos, utilizados especificamente para pesquisa. Esse corpus está dividido em textos corrigidos, não-corrigidos e semicorrigidos. Esta classificação, que teve como critério a maior ou menor aproximação ao português escrito padrão, deveu-se ao primeiro objetivo da reunião do *corpus*: o desenvolvimento de um revisor gramatical.

Os textos classificados como corrigidos, num total de aproximadamente 32.590.000 palavras, são aqueles publicados para grande número de leitores (livros, jornais, revistas, etc.), que são, portanto, supostamente corrigidos por especialistas em revisão de textos. É composto por cerca de 4.300 textos de: livros (de literatura brasileira; didáticos – biologia, química, física, história, geografia; enciclopédias; temas - arte, ciências, etc.); revistas; constituição brasileira e textos jurídicos; jornais.

Os textos não corrigidos, totalizando cerca de 738.000 palavras, são textos autênticos, escritos por universitários ou por pessoas de nível médio de escolaridade (2º grau). É composto por 2.400 textos que incluem redações, monografias e textos de publicidade, por exemplo.

Finalmente, os textos semicorrigidos, em número de 238, com aproximadamente 1.150.000 palavras, são textos publicados para um pequeno número de leitores, ou não publicados, que são corrigidos, mas, geralmente, não por especialistas em revisão de textos. Estão entre estes, contratos, relatórios, dissertações acadêmicas e outros.

Todos os textos do corpus se apresentam sem formatação, gráficos, tabelas, ilustrações, fórmulas ou quaisquer outros sinais. Os textos, integrais ou partes de textos, constituem-se em textos por fazerem sentido ao leitor humano; não são, portanto, conjuntos de frases.

A linguagem predominante no corpus é a linguagem jornalística, que, na visão de Partington (1998, p.13), “... é uma soma de grande número de tipos de textos”. Muitos desses textos sugerem características em comum com a linguagem falada, por serem produzidos em curto espaço de tempo, como é o caso dos comentários esportivos e artigos, porém isso não os faz nem mais nem menos representativos que qualquer outra modalidade de textos. Sendo assim, todas as modalidades de textos presentes no corpus servem como base para investigação lingüística e nenhuma modalidade de linguagem é superior à outra.

Para Perini (1985) apud Bezerra (2003, p. 46), “textos jornalísticos e técnicos (jornais, revistas, livros) apresentam uma regularidade gramatical e mesma estilística em todo o Brasil, de modo que não se distingue lingüisticamente um texto jornalístico ou técnico publicado em uma região ou outra do país (o que não ocorre com textos literários). Por essa razão, esse autor defende que a variedade padrão do Português brasileiro é encontrada nesses textos”.

Nesta pesquisa, através do programa concordanciador *WordSmith Tools Concord*, gerou-se uma concordância de 630 ocorrências do **que nem**, encontradas em 7.253 arquivos de 101 diretórios do Corpus NILC e, conforme exposto, os textos são predominantemente de jornais.

As ocorrências foram provenientes do corpus inteiro, de arquivos de todas as modalidades de textos presentes, tanto de textos corrigidos, semicorrigidos, quanto não-corrigidos e, para se obter uma amostra equilibrada, optou-se por uma busca aleatória, feita numa regulagem de 1 por 2, ou seja, de cada duas ocorrências do **que**

nem no corpus, uma foi extraída para fazer parte da amostra, perfazendo uma análise de metade da amostra pesquisada, a fim de delimitar o assunto.

Após a verificação, observou-se que havia ocorrências repetidas, as quais foram excluídas, perfazendo, assim, um total de 606 ocorrências que serão descritas neste trabalho.

3.2 O Programa *WordSmith Tools*

O programa concordanciador selecionado para uso nesta pesquisa é o *WordSmith Tools*, desenvolvido por Mike Scott (1995) e publicado pela *Oxford University Press*. Esse instrumento computacional é utilizado para análise de corpora e através dele foram extraídas as concordâncias que serão analisadas.

Conforme descrito por Berber Sardinha (2004, p.86), o *WordSmith Tools* é um conjunto integrado de programas que permite analisar como as palavras se comportam em um texto. Ele é composto de três ferramentas para análise lingüística: o *WordList*, o *KeyWords* e o *Concord*. Essas ferramentas são capazes de selecionar os arquivos-texto, indicar a frequência com que os dados ocorrem na amostra, compor a concordância desejada, além de outras possibilidades de manipulação do programa.

Numa breve descrição, são assim caracterizados:

Wordlist: Através dele são criadas as listas de palavras. Uma é ordenada alfabeticamente e a outra, por ordem de frequência das palavras. Numa terceira lista, aparecem estatísticas simples a respeito dos dados.

KeyWords: Permite a seleção de itens de uma lista de palavras (ou mais) por meio da comparação de suas frequências com uma lista de referências. O resultado do contraste é uma lista de palavras-chave, ou palavras cujas frequências são estatisticamente diferentes no corpus de estudo e no corpus de referência. Essa ferramenta não foi utilizada neste trabalho, pois esse tipo de análise não era o nosso objetivo.

Concord: Produz concordâncias ou listagens das ocorrências de um item específico (chamado palavra de busca ou nóculo, que pode ser formado por uma ou mais palavras) acompanhado do texto ao seu redor (o cotexto). Para descrever a padronização de um item lexical, é necessária a observação dos contextos de ocorrências desse item (correlação entre características lingüísticas e situacionais), com a utilização de instrumentos como a concordância, usada neste trabalho, através da ferramenta *Concord*.

Pelo exposto, observa-se que no contexto da metodologia da LC, o termo “concordância” não tem relação com o sentido gramatical, que consiste em se adaptar a palavra determinante ao gênero, número e pessoa da palavra determinada, podendo ser verbal ou nominal, conforme Bechara (2003, p. 543).

Dentre as possibilidades de se fazer concordância na metodologia da Lingüística de Corpus, a mais comum é a busca da palavra-chave no contexto, em que a palavra aparece centralizada e rodeada de seu contexto. Isso é possível, graças a alguns comandos que realizam essa tarefa.

Além de comandos específicos, o *Concord* disponibiliza alguns instrumentos de análise, como:

- Concordância (*concordance*): lista que contém uma palavra específica (de busca);
- Lista de colocados (*collocates*): lista de palavras que ocorrem ao redor da palavra de busca, em posições determinadas;
- Lista de agrupamentos lexicais (*clusters*): lista de seqüências de, pelo menos, três palavras recorrentes na concordância;
- Lista de padrões colocados (*patterns*): lista do resumo dos colocados nas posições em que são mais freqüentes;
- Gráfico de distribuição da palavra de busca (*plot*): as ocorrências da palavra de busca são representadas por uma barra.

Desses instrumentos, utilizou-se neste trabalho a concordância e a lista de agrupamentos lexicais e, para facilitar a leitura das listas de concordância, foi utilizado

o que se chama *anotação de corpus*, para que fosse possível definir a classe gramatical das formas selecionadas.

Leech (1997, p.2) define anotação como a prática de somar informação e interpretação lingüística de dados da língua falada e escrita a um corpus. É uma importante contribuição para a pesquisa, pois durante o processo de anotação, são inseridos rótulos ou etiquetas cuja função é indicar as categorias de determinada palavra que servirão de informação para posteriores interpretações.

Observe-se a figura a seguir que apresenta uma lista de concordância.

Figura 1: Lista de concordância

| N | Concordance | Set | Tag | Word No | File | % |
|----|---|-----|-----|---------|------|---|
| 47 | de reciclagem de lixo que nem as cidades jap | | | | | |
| 48 | A questão para ele é que nem as queria à fo | | | | | |
| 49 | destreza se defendia que nem assim o podia | | | | | |
| 50 | peessoas se lembrem de que nem "Bambi" nem | | | | | |
| 51 | ma geração de artistas que nem bem chegou a | | | | | |
| 52 | Eduardo Regianno, 15, que nem bem mudou d | | | | | |
| 53 | m! É boa! Trabalhara que nem besta de carga | c | | | | |
| 54 | via de ser aperreada que nem boi ladrão?... | c | | | | |
| 55 | nda estar aí, vivinho, que nem Braguinha, | | | | | |
| 56 | -Mas o senhor fala que nem cachoeira. | c | | | | |
| 57 | Homens são... ``bobos, que nem cachorro'' | c | | | | |

Nessa figura está representada apenas a concordância propriamente dita com as palavras que se encontram antes e depois da palavra de busca (que nem). As palavras que sucedem o **que nem** estão ordenadas alfabeticamente. O mesmo pode ser feito com as que antecedem. Na coluna *Set*, podem ser utilizadas letras do alfabeto como códigos para classificar as linhas da concordância. No caso, foi utilizada a letra *c* para indicar que, naquela linha, há uma locução conjuntiva comparativa. Há, nessa lista, um pequeno exemplo de como pode aparecer a concordância, ou seja, com as frases incompletas, porém, ao se acionar determinados recursos, é possível visualizar a frase ou o texto inteiros.

O uso dessa ferramenta possibilita também:

- verificar o número sequencial da linha da concordância, visível na extremidade esquerda da tela, na coluna *N*;
- inserir também na coluna *Set* e *Tag* códigos para classificar as linhas de concordância;
- verificar, na coluna *Word No*, o número correspondente à posição de cada palavra no corpus de estudo;
- encontrar, na coluna *File*, o arquivo em que cada ocorrência se localiza;
- localizar, na coluna %, em que parte do arquivo as ocorrências se encontram, por meio de percentuais em relação ao total de palavras e, a partir disso, delimitar o total de ocorrências no arquivo (%).

Após se obter a concordância, foi possível agrupar as palavras recorrentes através dos *clusters* – que é uma lista dos agrupamentos lexicais. Essa lista possui sequências fixas de palavras recorrentes na concordância, ou seja, são multipalavras que incluem a palavra de busca, no caso, o **que nem**, mas podem não ser, pois o programa busca os itens recorrentes na concordância, sem se limitar a trechos nos quais aparece a palavra de busca.

Vejamos o exemplo do *cluster* (agrupamento lexical) de algumas ocorrências do **que nem** na amostra pesquisada.

Figura 2: Agrupamento lexical (*cluster*)

| N | cluster | Freq |
|----------|----------------|-------------|
| 1 | que nem sempre | 71 |
| 2 | que nem o | 43 |
| 3 | que nem todos | 38 |
| 4 | que nem mesmo | 36 |
| 5 | é que nem | 26 |
| 6 | nem todos os | 23 |
| 7 | de que nem | 22 |

Através do *cluster*, foi possível perceber quão recorrente é a expressão “que nem sempre”, com 71 ocorrências na amostra pesquisada, porém foram analisadas 69, pois 02 estavam repetidas.

3.3 A pesquisa

Conforme exposto anteriormente, nesta pesquisa, foi feita a descrição do uso do item lexical **que nem**, com ênfase na locução conjuntiva comparativa. Para isso, extraiu-se uma listagem de ocorrências desse item (concordância) e, a seguir, utilizando-se o *cluster* foi possível observar quais agrupamentos de palavras são recorrentes com o **que nem** e em quais contextos ele aparece. Assim, partiu-se para uma classificação, utilizando-se a anotação de corpus, conforme os sentidos encontrados em 606 ocorrências. Procurou-se verificar quais os padrões mais recorrentes e procedeu-se à classificação.

Pode-se considerar que esse item possui “padrões”, porque, conforme explanado no cap. 2, os padrões de uma palavra podem ser definidos como todas as palavras e estruturas com as quais são regularmente associados e que contribuam para seu significado. Além disso, devem ocorrer com relativa frequência, segundo Hunston e Francis (2000, p.37).

O padrão encontrado diz respeito ao fenômeno da “colocação”, pois se refere à associação entre itens lexicais ou entre o léxico e campos semânticos. Ela evidencia que as palavras não se combinam ao acaso, mas seguem regras ou motivações do mundo real.

4 RESULTADOS E ANÁLISE

A seguir, serão analisados os padrões do **que nem** encontrados na amostra pesquisada. É bastante recorrente o aparecimento do **que nem** com determinado sentido, por exemplo, **nem** com o sentido de **sequer** ou associado a outros itens lexicais, como o **que nem** associado ao **todos**. Para esta análise, serão levados em conta, principalmente, os sentidos do **nem** que, associado ao **que** ou a outro item lexical, seguirá um padrão ou outro. Quanto ao **que**, sabe-se que ele pode pertencer a várias categorias gramaticais, exercendo as mais diversas funções sintáticas. Nas ocorrências analisadas, ora é conjunção, ora é pronome relativo. O fato de o **que** ter uma ou outra classificação não interfere na padronização, que pode ser feita com base apenas na coocorrência dos itens.

O **que nem** mais recorrente é o que apresenta o **nem** com o sentido de *sequer*, *ao menos*, com 132 exemplos, o que corresponde a 21,78% da amostra pesquisada. O mais raro é aquele sucedido pelo *assim* e pelo *mais*, com um exemplo de cada, correspondendo a 0,16%. A locução conjuntiva comparativa **que nem** = *como*, que é o principal alvo desta pesquisa e, portanto, foi a mais explorada, apareceu em 3º lugar em número de ocorrências.

As 606 ocorrências do item lexical **que nem** foram contempladas na tabela a seguir.

4.0 Classificação geral do que nem

Tabela 1: Padrões do que nem: ocorrências e frequência

| Padrões | Total de ocorrências | Frequência em % |
|---|----------------------|-----------------|
| 1º que nem = sequer, ao menos | 132 | 21,78 |
| 2º que nem = não | 129 | 21,29 |
| 3º que nem = como | 97 | 16,00 |
| 4º que nem todo(s), toda(s), tudo | 76 | 12,54 |
| 5º que nem sempre | 69 | 11,39 |
| 6º que nem ... nem (alternância) | 36 | 5,94 |
| 7º que nem mesmo | 34 | 5,61 |
| 8º que nem sequer | 19 | 3,13 |
| 9º que nem ao menos | 5 | 0,83 |
| 10º que nem + advérbio (só, bem, mais, assim) | 9 | 1,49 |
| TOTAL | 606 | 100.00 |

4.1 1º padrão: que nem = sequer, ao menos

O padrão mais recorrente na amostra representa 21,78% do total analisado. Com 132 ocorrências, o **nem** com o sentido de *sequer, ao menos* é o mais utilizado pelos falantes do português brasileiro quando precedido pelo **que**.

(01) Marques disse **que nem** o prefeito entra na festa.

Nesse caso, poder-se-ia dizer: “Marques disse que **sequer** o prefeito entra na festa.” Ou: “Marques disse que **nem sequer** o prefeito entra na festa”. A conjunção integrante **que** o acompanha nesse exemplo e na maioria dos casos analisados.

Deve-se observar que quando o **nem** assume o sentido acima exposto, ele não pode ser substituído pelo não. Seria agramatical uma sentença assim: “Marques disse

que não o prefeito entra na festa.” Se fosse: “Marques disse que o prefeito não entra na festa”, seria uma sentença possível, mas sem a força argumentativa daquela em que aparece o **nem**, pois quando se diz: “Marques disse **que nem** o prefeito entra na festa”, parece que há uma expectativa rompida, pois se nem o prefeito entra na festa (e ele, provavelmente, está num nível mais alto numa determinada escala), tampouco os demais entrarão. É claro que uma interpretação vai depender do contexto em que a sentença é proferida, mas se o falante utilizou o **nem**, é porque, provavelmente, quis dizer algo mais, que ficou subentendido.

Outro aspecto interessante desse padrão é que, depois desse **que nem**, é rara a utilização de verbo. Observe-se que, no exemplo analisado, o elemento que sucede o **que nem** é um substantivo (o prefeito). Nas ocorrências em que há verbo (somente 03 ocorrências com o verbo *querer* = *quer*), ele está formando uma locução verbal, mas como o **nem** pode ser substituído pelo **não**, ele está sendo considerado como pertencente ao 2º padrão, como é o caso do exemplo a seguir:

(2) Não quis analisar a manutenção de Mazinho em seu lugar ("Vocês podem conversar com o técnico...") e disse que, no momento, o importante é que ele está torcendo pela seleção. E está tão concentrado nela **que nem quer** falar sobre sua possível transferência para o Mônaco.

Nesse exemplo é possível dizer: *...está tão concentrado que não quer falar...*, mas, em virtude da sonoridade, torna-se estranho dizer: *...está tão concentrado que sequer quer falar...*

Já uma sentença do tipo “A criança disse **que nem brincar** ela podia” não apareceu na amostra do corpus. Nesse caso, o *nem* pode ser substituído pelo *sequer*, mas não pode ser substituído pelo *não*. Portanto, conforme verificado na amostra do corpus, é rara a utilização de verbo depois do **que nem** pertencente a esse padrão.

4.2 2º padrão: **que nem = não**

Em segundo lugar em número de ocorrências (129 do total) correspondendo a 21,29%, aparece o **nem** com o sentido de **não** em contexto de negação, acompanhado pelo **que** (pronome relativo na maioria dos casos ou conjunção em menor escala).

(03) Santos afirmou que **nem sabia** da existência de uma crise.

Nesse caso, **que** é conjunção integrante do verbo *dicendi*.

(04) “...paguei o refrigerante que **nem cheguei** a beber...”

Aqui o **que** é pronome relativo. Substitui o substantivo “refrigerante”.

O curioso, nos dois casos e em todas as demais ocorrências desse padrão, é que o elemento que sucede o **que nem** é sempre um verbo em 110 casos. Nas outras 19 ocorrências, também há verbo, mas precedido por pronome oblíquo, como no exemplo:

(05) “...pode ser até **que nem me ouviu**.”

Conclusão: na estrutura do **que nem**, quando o **nem** tem sentido de **não** ou pode ser substituído pelo **não**, certamente na seqüência aparecerá um verbo ou um pronome oblíquo + verbo. No exemplo a seguir, é perfeitamente possível construir a sentença com a utilização do **não** no lugar do **nem**, porém novamente se deve levar em conta o contexto em que ela foi proferida e a intenção do falante, pois, com a utilização do **nem**, haverá uma força argumentativa maior, mais incisiva ou a demonstração de indignação. A construção com o **não** fica mais branda.

(06) Como alguém pode governar um povo **que nem conhece**?

(07) Como alguém pode governar um povo **que não conhece**?

Observe-se que o **que** funciona como pronome relativo e, sucedendo o **que nem**, há um verbo.

4.3 3º padrão: **que nem = como**

O grande foco desta pesquisa centra-se neste 3º padrão do **que nem** quando equivale ao **como** - locução conjuntiva comparativa. As 97 ocorrências desse item lexical (16% do total) foram classificadas em duas categorias:

- 1- categoria de natureza léxico-gramatical
- 2- categoria de gêneros textuais

4.3.1 - Categoria de natureza léxico-gramatical

A associação entre itens lexicais e gramaticais (o item lexical **que nem** se associa aos verbos, substantivos, que são itens gramaticais) forma uma coligação.

Na tabela a seguir, são apresentadas as associações mais frequentes e menos frequentes do **que nem** (locução conjuntiva comparativa) e dos elementos gramaticais que o acompanham, que coocorrem com ele nas 97 ocorrências do corpus pesquisado.

Tabela 2: Coligações 1: ocorrências e frequência

| Coligações | Total de ocorrências | Frequência em % |
|-------------------------------------|-----------------------------|------------------------|
| verbo + que nem + substantivo | 30 | 30,93 |
| adjetivo + que nem + substantivo | 22 | 22,68 |
| substantivo + que nem + substantivo | 20 | 20,62 |
| substantivo + que nem + pronome | 11 | 11,34 |
| pronome + que nem + substantivo | 04 | 4,13 |
| substantivo + que nem + verbo | 02 | 2,06 |
| verbo + que nem + pronome | 02 | 2,06 |
| pronome + que nem + pronome | 02 | 2,06 |
| adjetivo + que nem + pronome | 01 | 1,03 |
| adjetivo + que nem + verbo | 01 | 1,03 |
| outros | 02 | 2,06 |

Percebe-se que, na maior parte das ocorrências, os verbos, substantivos e adjetivos é que se associam ao **que nem** para se fazer uma comparação. Em menor escala, estão os pronomes, por exemplo. Desse modo, pode-se estabelecer como categoria de natureza léxico-gramatical:

4.3.1.1 Categoria 1 - verbo + **que nem** + substantivo (com ou sem artigo) = 30 ocorrências

Essa coligação corresponde a 30,93% do total das ocorrências. Esse fenômeno causa certa surpresa, pois, como visto no capítulo 2, quando aparecem exemplos de comparações nas gramáticas ou dicionários, o adjetivo é que normalmente funciona como elemento de comparação, como se fosse possível só haver comparação com o uso de adjetivos, conforme Houaiss (2001) “é teimoso que nem o pai”. Pode-se perceber claramente que isso não é verdade, pois o verbo foi o item gramatical mais recorrente na comparação. Das 30 ocorrências com verbo, 14 delas contêm verbos de ligação.

(08) Mané Candeeiro falava pouco, a não ser que se tratasse de coisas de caça; mas cantava **que nem** passarinho.

(09) "Eu sou **que nem** o Tyson. Não jogo a toalha nunca. Taco o porrete no adversário e venço a luta." Orestes Quércia, candidato do PMDB.

Nesses exemplos, há dois termos que se comparam: o 1º termo tem como elemento mais forte um verbo e o 2º termo, um substantivo. Temos um padrão de natureza gramatical que assume força argumentativa na comparação. A ação de “cantar” em (08) assume mais força argumentativa do que o substantivo “passarinho”. Por si só, o verbo já esclarece a comparação e o restante da oração funciona como um complemento que, de certa forma, retoma os adjetivos subentendidos referentes a passarinho (que é bom cantor, é afinado). Da mesma forma em (09), o verbo “ser”

demonstra uma certeza diante da comparação, supervalorizando o termo comparado, ou seja, a pessoa se considera melhor do que o Tyson.

Isso confirma a análise de Vogt (1977, p. 202) sobre a comparação, pois parece haver uma relação de equilíbrio entre os termos comparados. Porém, há a presença de oposição entre eles, visto que o **nem** tem valor negativo. Logo, “cantava **que nem** passarinho” assume o sentido de “cantava que nem mesmo passarinho canta”, ou seja, cantava melhor do que passarinho. Dessa maneira, quando houver a locução conjuntiva **que nem** introduzindo uma comparação, o 1º termo terá mais força argumentativa e se sobressairá ao 2º termo. Além disso, ocorre nesses casos uma comparação *assimilativa*, que consiste em assimilar uma coisa, pessoa, qualidade ou fato a outra mais impressionante ou mais conhecida. Assimila-se o fato de cantar com aquele que é mais conhecido no ato de cantar: o passarinho: “...cantava *que nem passarinho*.”

4.3.1.2 Categoria 2 - adjetivo + **que nem** + substantivo = 22 ocorrências

Já vimos que a utilização de adjetivo numa construção comparativa era esperada, embora de maneira mais recorrente do que realmente foi, visto que é o tipo de construção que aparece nos exemplos de obras de referência.

(10) A noite estava clara **que nem** dia.

Nessa comparação de igualdade, pode-se interpretar que há uma comparação de superioridade, pois se percebe que o 2º termo “dia” é intenso, mais forte e exemplar quanto à qualidade comparada “clara” (1º termo). Assim, se a noite estava clara comparando-se ao dia, é porque ela estava muito mais clara do que normalmente é. Uma outra interpretação parece sugerir que o dia não é tão claro. Porém, não se quer enfatizar que o dia não é claro, mas sim, que a noite é clara, ou seja, a noite estava tão clara que nem mesmo o dia é tão claro.

4.3.1.3 Categoria 3 - substantivo + **que nem** + substantivo = 20 ocorrências

(11) “A gente fala pro pessoal que o baile é **que nem** uma igreja.”

Normalmente, “igreja” é um substantivo que parece ter conotação positiva. Ao se fazer essa comparação, parece que, nas entrelinhas, há uma retomada das qualidades, dos adjetivos referentes à igreja. Se o “baile” é comparado à igreja é porque ele também assume essa conotação, tornando-se tão positivo que nem mesmo a igreja o é. Novamente se sobressai o 1º termo, que adquire argumento mais favorável para si em relação à propriedade com o qual ele é comparado.

4.3.1.4 Categoria 4 - substantivo + **que nem** + pronome = 11 ocorrências

(12) Depois de assistir ao empate entre Camarões e Suécia (2 a 2) e outros vídeos com partidas da seleção africana, Parreira concluiu: "Camarões tem um balanço **que nem** o nosso".

(13) Esse homem é brasileiro **que nemeu**.

4.3.1.5 Categoria 5 - pronome + **que nem** + substantivo = 04 ocorrências

(14) Estava tudo **que nem** um brinco.

4.3.1.6 Categoria 6 - substantivo + **que nem** + verbo = 02 ocorrências

(15) O inferno é um sem-fim **que nem** não se pode falar.

4.3.1.7 Categoria 7 - verbo + **que nem** + pronome = 02 ocorrências

(16) Dos gregos a Gerald Thomas, todos me devem notável contribuição ao teatro universal. Sou **que nem** aquele cara que diariamente ganhava 500 cruzeiros porque não cuspi no elevador de seu prédio. Havia a placa: "É proibido cuspir. Multa: 500 cruzeiros". Ele não cuspi. Ano passado, neste mesmo espaço, lembrei minha única experiência teatral.

4.3.1.8 Categoria 8 - pronome + **que nem** + pronome = 02 ocorrências

(17) A escuta de um "si" profundo, que talvez favoreça por extensão, a melhor escuta e aproximação daquele outro que "não é **que nem** eu" e produz sua cultura que nem eu. E eu quem sou?

Nesse caso, a comparação se dá não com um elemento imediatamente anterior ao **que nem**, mas com o "daquele outro" (pronome) que é comparado ao "eu" (pronome).

4.3.1.9 Categoria 9 - adjetivo + **que nem** + pronome = 01 ocorrência

(18) Isso, um cachorro. Mansinho **que nem** ele só. Quer ver?

4.3.1.10 Categoria 10 - adjetivo + **que nem** + verbo = 01 ocorrência

(19) Na quinta pergunta, temos: "Padeci com jeito. E o governo da vida? Anos, que me foram de gentil sujeição, custoso **que nem** guardar chuva em cabaça, picar fininho a couve." (p. 47) A própria pergunta gera uma expectativa no sentido de haver uma mudança de decisões da protagonista. No entanto, há quebra desta expectativa, pois Flausina continua "submissa", ou pelo menos fingindo que é submissa.

As coligações presentes do 4º ao 10º exemplos podem ser analisadas da mesma forma que as anteriores. Destaca-se o 1º termo da comparação, que terá supremacia sobre o 2º, em relação à sua força argumentativa, não importando qual é a

classe gramatical a que ele pertence. Importa perceber o sentido que eles assumem em relação à argumentatividade do que saber qual é a classe gramatical do elemento.

4.3.1.11 Categoria 11 - outros = 02 ocorrências

(20) Você deve ajudar e sempre. Lá a gente vê crianças muito doentes, que tem o câncer muito evoluído e sofrem demais. **Que nem** eu falava pra D. Luiza: eu vou continuar como colaboradora, não vou ser voluntária.

(21) Isso tá indo bem, este trabalho psicológico? FV - Tá indo bem. O que ele coloca é que a gente tem que fazer sempre... **que nem**: já que a gente tá aqui fazendo, buscando, a gente vai buscar uma medalha fazer o melhor possível no momento, entendeu? Dá os 110 por cento. É aquele 10 por cento a mais do que as pessoas dão normalmente. Então se a gente chega em julho e conseguir a medalha de prata, ótimo...

Nesses casos, o **que nem** (comparativo), com o sentido de “como” parece assumir um valor de conectivo. Há uma comparação da situação exposta, mas, ao mesmo tempo, ele está ligando as orações. É um caso bem típico da oralidade.

4.3.1.12 Conclusões sobre as categorias de natureza léxico-gramatical :

Como se pôde perceber, a coligação “verbo + **que nem** + substantivo (com ou sem artigo)” é a mais recorrente na amostra pesquisada a respeito da locução conjuntiva comparativa. Portanto, do ponto de vista probabilístico, os dois elementos que vão acompanhar o **que nem** (locução conjuntiva comparativa) com mais frequência são os verbos e os substantivos, seguidos pelos adjetivos e, por último, pelos pronomes.

Em situações reais de uso da língua, parece mais comum comparar determinada ação dos indivíduos (verbo) com determinado elemento (substantivo), do

que comparar suas qualidades ou defeitos (adjetivos) com algum elemento. No exemplo: *Trabalhara **que nem** besta de carga...*, comprova-se essa idéia.

Em segundo lugar, aparecem as comparações com adjetivos como em...*vermelho **que nem** tomate*. O que parece comum é a utilização do substantivo como o 2º termo da comparação, ou seja, os verbos (da 1ª categoria), os adjetivos (da 2ª categoria) e os substantivos (da 3ª categoria) formam o 1º termo da comparação e são comparados a substantivos, mediados pelo **que nem**. Já a comparação de uma pessoa com outra (com a utilização de pronomes) é bem menos evidente: *Ele é **que nem** aquele cara*.

Se levarmos em conta que, do ponto de vista sintático, podemos agrupar os pronomes aos substantivos quando estes são núcleos de sintagmas nominais (ex. meu pai) e que, normalmente, junto a esses pronomes estão verbos de ligação, o sintagma nominal é que vai predominar (substantivo + que nem + substantivo) e não o sintagma verbal (verbo + que nem + substantivo), já que há 14 ocorrências com verbos de ligação, como em “*Eu sou **que nem** o Tyson*. Portanto, fazendo-se esses agrupamentos e unindo-se as ocorrências de verbos de ligação à dos substantivos, a totalidade dessas ocorrências seria:

Tabela 3: Coligações 2: ocorrências e frequência

| Coligações | Total de ocorrências | Frequência em % |
|-------------------------------------|----------------------|-----------------|
| substantivo + que nem + substantivo | 34 | 35,05 |
| adjetivo + que nem + substantivo | 22 | 22,68 |
| verbo + que nem + substantivo | 16 | 16,50 |
| substantivo + que nem + pronome | 11 | 11,34 |

Com base nesses resultados, pode-se considerar que a locução conjuntiva comparativa **que nem** segue um padrão de linguagem em que predomina o sintagma nominal: substantivo + que nem + substantivo. Bechara (2003, p.493) esclarece que a comparação é *assimilativa*, “quando consiste em assimilar uma coisa, pessoa, qualidade ou fato a outra mais impressionante, ou mais conhecida”. Há um ser que se

compara a outro ser. Portanto, há um nome comparado a outro nome, ou seja, substantivos comparados a substantivos.

As demais coligações são mais raras (com 1, 2 ou 3 ocorrências cada). Isso significa que, nas comparações, prevalecem as estruturas gramaticais formadas com substantivos, adjetivos e verbos comparados a substantivos ou, em menor escala, a pronomes.

4.3.2 Categoria de gêneros textuais

A locução conjuntiva comparativa **que nem = como** é recorrente em determinados gêneros textuais e, provavelmente, inexistente em outros, como nos textos jurídicos ou religiosos, pois são textos mais formais e a locução é mais informal.

Na tabela a seguir, são apresentados o total de ocorrências e a frequência do **que nem = como** nos gêneros textuais em que esse item apareceu na amostra pesquisada.

Tabela 4: Gêneros textuais: ocorrências e frequência

| Fontes: gêneros textuais | Total de ocorrências | Frequência em % |
|---------------------------------|-----------------------------|------------------------|
| literários | 43 | 44,33 |
| jornalísticos | 19 | 19,59 |
| entrevistas | 17 | 17,53 |
| livros técnicos | 16 | 16,49 |
| monografia | 01 | 1,03 |
| dissertação | 01 | 1,03 |
| Total | 97 | 100,00 |

É mais recorrente o uso do **que nem** em textos literários e jornalísticos. Conforme exposto no capítulo 3, o corpus pesquisado é composto por 35 milhões de

palavras (é considerado um corpus grande) e há predominância de textos jornalísticos. Através da ferramenta *WordList*, apresentada no capítulo 3, foi possível verificar qual é a quantidade de palavras presentes em cada gênero. Constam, nos textos literários, 2.232.840 palavras e nos textos jornalísticos, 30.799.136 palavras, portanto, quase a totalidade do corpus.

Vejamos como as ocorrências se apresentam na amostra, para, a seguir, tecermos algumas considerações.

4.3.2.1 Registros literários

Como se pode observar, é recorrente a utilização **que nem = como** nos textos literários, totalizando quase a metade da amostra pesquisada (44,33%).

O programa permite que, ao se acionar determinado recurso, seja possível visualizar as fontes de onde foram extraídas as ocorrências. Essas fontes são obras literárias, que, pelo contexto, é possível deduzir que são obras conhecidas, como, por exemplo, *Inocência* cujo autor é Visconde de Taunay; *O cortiço* de Aluísio Azevedo e *Triste fim de Policarpo Quaresma* de Lima Barreto. Esses autores se destacaram na Literatura Brasileira, entre outras coisas, por retratar o universo popular, a linguagem popular e espontânea do povo brasileiro. Isso significa que predomina nessas obras o aspecto informal e coloquial da linguagem, comprovando que a utilização do **que nem** (locução conjuntiva comparativa) está relacionada mais à oralidade do que à linguagem formal, mesmo que apareça no discurso do narrador, conforme o exemplo.

Discurso do narrador:

(22) Como que se sentiam ainda na indolência de ~~re~~blina as derradeiras notas da última guitarra da noite antecedente, dissolvendo-se à luz loura e tenra da aurora, **que nem** um suspiro de saudade perdido em terra alheia. A roupa lavada, que ficara de véspera nos coradouros, umedecia o ar e punhalhe um farto acre de sabão ordinário.

Discurso do personagem:

(23) - Bem, respondeu o amanuense num tom seco, pondo o chapéu sobre uma cadeira. E logo: - Homem, isto está **que nem** um paraíso! - Qual paraíso! Está nos debicando?... - Não senhora, longe de mim tal pensamento. O que digo é a verdade.

Além disso, o gênero que mais pode refletir a individualidade de quem fala (ou escreve) é o texto literário, conforme Bakhtin (2000, p.279). Nesses exemplos, há marcas da individualidade do narrador, do personagem e do próprio autor. Isso também confere maior liberdade em relação ao uso da linguagem que, nesses casos, tende a ser mais coloquial.

4.3.2.2 Registros jornalísticos

Em segundo lugar, em 19,59% das ocorrências, estão os textos jornalísticos pertencentes, na maior parte, à Folha de São Paulo e alguns, ao Jornal do Brasil. Eles se dividem em:

- textos de opinião (de colunistas) e crônicas = 13 ocorrências
- comentários ou depoimentos esportivos/ políticos = 5 ocorrências
- reportagem = 1 ocorrência

Nos textos de opinião, muitas vezes, como forma de convencer ou emocionar o leitor, utiliza-se uma linguagem que se aproxima da linguagem dele, ou mais coloquial, ou despojada como a do colunista conhecido como Macaco Simão (colunista que aparece na amostra). No exemplo a seguir, parece que há um apelo para a emoção do leitor.

(24) Para se ter uma idéia: os personagens viviam entocados em seus apartamentos, com medo de serem alvejados por um tiro. Não podiam dar sopa em janelas e arrastavam-se pelo chão da sala **que nem** soldados em campo de batalha. Uma Bósnia "avant la lettre"; ou melhor, "avant le deluge". O que parecia uma caricatural distopia no final dos anos 60 está se tornando realidade nos anos 90. Não em Nova York, mas em certas regiões da zona sul do Rio. (FOLHA OPINIÃO)

As crônicas são, por natureza, textos escritos para serem publicados em jornais ou revistas. O cronista normalmente extrai os seus temas do cotidiano e emite comentários pessoais ao escrever sobre algo que é circunstancial. Assim, esse gênero textual também apresentará marcas da oralidade e o **que nem** é uma delas.

Já a reportagem parece que constitui um caso mais raro na utilização do **que nem**, visto que, normalmente, é apresentada na modalidade culta da língua.

Quanto aos comentários esportivos ou políticos, é provável que, inicialmente, eles tenham sido feitos oralmente e, posteriormente, foram publicados (como se fossem uma transcrição do que se ouviu). É um gênero textual que permite haver marcas da oralidade.

(25) Depois de assistir ao empate entre Camarões e Suécia (2 a 2) e outros vídeos com partidas da seleção africana, Parreira concluiu: "Camarões tem um balanço **que nem** o nosso". O técnico espera que a marcação individual aplicada sobre Romário e Bebeto pela defesa da Rússia se repita na partida de hoje.

4.3.2.3 Entrevistas

As entrevistas aparecem em 3º lugar no total das ocorrências (17), correspondendo a 17,53%. Talvez mais do que qualquer outro texto, a entrevista é um gênero tipicamente oral que, posteriormente, é publicada. E, dependendo do tipo de entrevista ou do público-alvo, as marcas da oralidade estarão presentes em maior ou menor grau nesse gênero, conforme exposto no capítulo 2.

(26) Eu tenho medo que a música me abandone já que estou me afastando dela; os compositores que eu conheço depois de certo tempo param de fazer suas músicas e vão pescar **que nem** Dorival Caymmi. Às vezes eu me pergunto: porque os todos os compositores fazem suas músicas até seus trintas e tantos anos e depois começam a parar. Estou só exibindo a minha ignorância a respeito do assunto e a minha indagação.

4.3.2.4 Livros técnicos

No arquivo referente aos livros técnicos, que é o termo utilizado no corpus, 16 ocorrências de uma locução conjuntiva comparativa (**que nem**), representam um número significativo para um total de 97 da amostra, visto que essa locução é tida como de cunho popular. Então por que apareceria tanto em livros técnicos tidos como mais formais? No exemplo a seguir, parece que há uma análise de uma obra de Guimarães Rosa, portanto, de um texto literário. O **que nem** aparece justamente na fala de um personagem que, normalmente, na obra de Rosa, é marcada pela variante caboclo-sertaneja da língua portuguesa.

(27) Deus é definitivamente; o demo é o contrário Dele. (...) Mas a gente quer Céu é porque quer um fim: mas um fim com depois dele a gente tudo vendo (...). O inferno é um sem-fim **que nem** não se pode ver. (GSV p. 49, grifos nossos). Contra estas afirmações pertencentes ao imaginário estável de compadre Quelemém, e que Riobaldo nunca interiorizará completamente, as suas interrogações visam precisamente a elucidar o que é o inferno - trazendo à visão e ao conhecimento especulativo a consistência interna, a maneira de ser da negatividade, a natureza do não-ser.

Em outros exemplos, a utilização do **que nem** está presente em outras falas do povo brasileiro, que estão sendo analisadas pelo autor do livro técnico ou simplesmente estão sendo citadas para fortalecer a argumentação.

(28) Seu saber é outro e foi aprendido às duras penas da miséria e da exploração. Tão afastado no espaço e na classe, que o reconhecimento de uma identidade comum "brasileiro **que nem** eu", determinada pela existência de uma nação comum, soa como estranheza. Brasileiro que nem eu? O seringueiro? Mas, se não se anula classe, Mário, todavia anula a distância, deixa a Lopes Chaves e viaja. Turista aprendiz, vai ao encontro do concreto.

A conclusão a que se chega é a de que, o **que nem** aparece nessas obras (livros técnicos, portanto, mais formais quanto ao uso da linguagem), não porque o autor empregou “equivocadamente” um item típico da oralidade num texto escrito, mas porque esse item lexical faz parte de uma obra literária analisada por esse autor. Se o autor analisa as falas populares, esse item fará parte dessa análise, já que é tido como

de cunho popular e oral. Essa mesma justificativa serve para o 5º e 6º gêneros, a seguir.

4.3.2.5 e 4.3.2.6 Monografia e dissertação

Numa monografia (exemplo 29), em que se analisa uma obra literária, houve uma ocorrência do **que nem**, justamente na fala do personagem. E numa dissertação (exemplo 30) relacionada a clichês ou à fala de personagens, também apareceu uma ocorrência.

(29) Logo após, o narrador fala da "luta" que teve de travar com os Lopes para adquirir independência quanto aos rumos da sua vida; e mesmo assim, não conseguiu, continuando "submissa" a eles. Na quinta pergunta, temos: "Padecei com jeito. E o governo da vida? Anos, que me foram de gentil sujeição, custoso **que nem** guardar chuva em cabaça, picar fininho a couve." (p. 47) A própria pergunta gera uma expectativa no sentido de haver uma mudança de decisões da protagonista. No entanto, há quebra desta expectativa, pois Flausina continua "submissa", ou pelo menos fingindo que é submissa. Na sexta pergunta, o narrador põe em dúvida a paternidade das duas últimas crianças...

(30) No corpus da presente pesquisa, apareceram alguns clichês, como: "moço lindo como um cravo"; "em busca da felicidade"; "um príncipe corajoso e valente"; "lábios **que nem** docê de mel" [sic]; entre outros. Foi com a atividade 4, Procurando firme, aplicada à 5ª série que apareceu um maior número de clichês. A presença de discurso direto para indicar a fala dos personagens apareceu nos textos dos alunos de todas as séries, sobretudo nos do CB-2. Talvez isso se deva à influência do texto-base que apresenta um diálogo entre os dois personagens.

4.3.3 Conclusões sobre a categoria de gêneros textuais:

Com base na observação e na análise dos dados da amostra do corpus pesquisado, no que diz respeito aos gêneros textuais, foi possível chegar a algumas conclusões. O questionamento que se fazia no início desta pesquisa estava relacionado à recorrência ou não da utilização de um item lexical (considerado de cunho popular pela gramática normativa e mais utilizado na modalidade oral da língua), em textos extraídos da língua escrita, reconhecidos como pertencentes à norma padrão da língua. Também interessava saber em que gêneros textuais o item lexical **que nem** equivalente ao **como** - locução conjuntiva comparativa - se fazia presente. Antes de passar às conclusões, faz-se necessário explicitar alguns conceitos sobre língua.

De acordo com Possenti, (1996, p. 74), para a gramática normativa, a língua corresponde às formas de expressão produzidas por pessoas cultas, de prestígio. Nas sociedades que têm língua escrita, é principalmente esta modalidade que funciona como modelo, acabando por representar a própria língua. Eventualmente, a restrição é ainda maior, tomando-se por representação da língua a expressão escrita elaborada literariamente. É a essa variante que se costuma chamar “norma culta” ou “variante padrão” ou “dialeto padrão”. Em casos mais extremos, mas não raros, chega-se a considerar que esta variante é a própria língua. A gramática normativa exclui de sua consideração todos os fatos que divergem da variante padrão, considerando-os “erros”, “vícios de linguagem” ou “vulgarismos”.

O grande problema da gramática normativa é o de não considerar as outras variantes da língua, o que gera preconceito.

Para Signorini (2001, p.107), não se pode considerar a escrita como a mais perfeita concretização do sistema abstrato da língua. A sociolinguística variacionista considera que há contraposição entre língua e fala, mas a questão da diferença e semelhança entre fala e escrita é deslocada para a das relações cruzadas entre as variedades linguísticas (de origem histórica, geográfica, social e estilística) e os estilos formal e informal: quanto mais formal o estilo, mais próximo da modalidade escrita e igualmente mais próximo da variedade tida como padrão; quanto mais informal o

estilo, mais próximo da modalidade oral da “fala espontânea” e, em consequência, mais sujeito às variações em relação ao padrão.

Levando-se em conta esses conceitos, o item lexical **que nem = como** (locução conjuntiva comparativa) realmente aproxima-se mais da modalidade oral da língua, sendo usado mais informalmente. Apesar disso, foi significativa sua ocorrência nos diversos gêneros textuais e as possíveis conclusões, embora não definitivas, já que a amostra não era tão grande, são as seguintes: o item lexical **que nem = como** (locução conjuntiva comparativa) teve 97 ocorrências (o que equivale a 16% de um total de 606 ocorrências) na amostra do corpus pesquisado. Esse item ocorreu em textos literários, jornalísticos, entrevistas, livros técnicos, monografia e dissertação. Em todos esses gêneros textuais houve marcas da oralidade, ou melhor, os textos, embora fossem extraídos da língua escrita, representavam a linguagem oral. Portanto, prova-se que o **que nem** é um elemento mais voltado à oralidade e, se apareceu em textos escritos, não foi porque houve “erro” na transcrição ou porque as pessoas estão utilizando-o como se ele fosse um elemento da norma padrão, mas porque ele estava unicamente representando a modalidade oral da língua. Por exemplo, se ele ocorreu no gênero *entrevista*, ele apareceu primeiro oralmente, para depois ser publicado. Nos textos literários, pertencentes à modalidade escrita, o **que nem** está justamente representando a fala popular, principalmente de personagens literários do interior do Brasil, além de marcar a individualidade do autor, que escreve segundo seus propósitos, como a de retratar a vida das pessoas das camadas mais populares e nisso se inclui sua linguagem. Nos textos jornalísticos, quando expressam a opinião do colunista, também há liberdade de escolha do vocabulário que ele vai utilizar. Se estiver mais adequado aos seus propósitos, o colunista utilizará o **que nem** ao invés do **como** para estabelecer comparações. Os comentários ou depoimentos esportivos quase sempre são produzidos com linguagem coloquial, pois também dizem respeito à opinião do povo. E os livros técnicos, monografia e dissertação que, por natureza, devem ser escritos com linguagem mais formal, na amostra do corpus revelaram esse aspecto, mas também revelaram que o assunto analisado nessas obras era relacionado às falas populares, portanto o **que nem** e outras expressões populares poderiam

aparecer como parte do que estava sendo analisado, que é também o que acontece neste trabalho.

Ainda em relação aos gêneros textuais, se levarmos em consideração que a quantidade de textos jornalísticos é muito maior do que a quantidade de textos literários no corpus pesquisado, chega-se à conclusão de que a porcentagem de ocorrência do **que nem** nos textos literários é bastante grande. Ocorre que, em corpora maiores, há mais chance de itens raros aparecerem. Assim, dada a quantidade de textos e dada a quantidade de ocorrências, pode-se pensar que, nos textos jornalísticos, a ocorrência do **que nem** é muito baixa em comparação com a ocorrência nos textos literários.

Frente a essas constatações, devemos questionar até que ponto é válida a utilização de obras literárias como referência para a língua padrão, já que em dicionários ou gramáticas, muitas vezes, são utilizados exemplos de textos literários como bons exemplos da norma culta. De acordo com as ocorrências do **que nem** no corpus, podemos perceber que o que predomina nos textos literários são as marcas da oralidade, portanto, diferentes das marcas da língua padrão. Isso não quer dizer que todo e qualquer texto literário dará ênfase ou se restringirá a utilizar marcas da oralidade.

Em suma, o **que nem = como** continua sendo uma expressão oral e de cunho popular, não havendo interferência do oral no escrito. As ocorrências demonstraram que seu uso predomina na oralidade. As pessoas não o utilizam em textos de língua escrita, tidos como pertencentes à norma padrão da língua, mais formais, a não ser que haja um bom motivo, como a análise de expressões populares em um livro técnico. Se fosse utilizado um corpus de linguagem oral, provavelmente as ocorrências seriam muito maiores.

Essas considerações estão relacionadas ao 3º padrão do **que nem** (locução conjuntiva comparativa). Na seqüência, temos o 4º padrão que é do **que nem** e quantificadores.

4.4 4º padrão: **que nem todo (s), toda (s), tudo**

Na observação da amostra do corpus pesquisado, constatou-se que, com os quantificadores **todo(s), toda(s), tudo**, o **que nem** também é bastante recorrente, principalmente com o **todo (s)**, apresentando 43 ocorrências de um total de 76. Com o **toda (s)**, houve 16 ocorrências e com o **tudo**, 17 ocorrências, totalizando 12,54% da amostra.

(31) A partir da Semana de Arte Moderna, a criação literária no Brasil foi intensa e vigorosa. É bem verdade **que nem todos os** autores seguiram uma mesma linha de idéias, pois se formaram vários grupos com tendências literárias diferentes.

(32) Quem é inquilino, entretanto, deve saber **que nem todas as** despesas são de sua responsabilidade. Muitas devem ser pagas pelo proprietário do apartamento

(33) Segundo a assessoria de imprensa do órgão, estocar água aumenta o desperdício já que **nem todo o** volume armazenado é consumido.

(34) A quem interessa essa carnificina brutal? Haverá alguém lucrando ou "levando vantagem" com tudo isso? O trânsito brasileiro mostra **que nem tudo o** que acontece de trágico na convivência humana é produto da vontade ou da intenção de algum indivíduo, organização ou grupo social.

Nesses exemplos, há uma idéia de totalidade expressa pelo quantificador **todos** e suas variantes, porém, usado com o **nem**, não há generalização e nem totalidade, visto que, nesses casos, o **nem** funciona como advérbio de negação, negando a totalidade. Nesses exemplos, o **que** funciona com conjunção.

Assim, entende-se que, no exemplo 31, os autores, a partir da Semana de Arte Moderna, não seguiram uma mesma linha de idéias. O uso do **nem** nega que **todos** os autores seguiram a mesma linha de idéias.

4.5 5º padrão: **que nem sempre**

Com 69 ocorrências, o padrão **que nem sempre** corresponde a 11,39% do total da amostra.

(35) No entanto nota-se entre os meios de comunicação brasileiros, principalmente a televisão, a propagação de ideologias **que nem sempre** contribuem para o exercício democrático.

Conforme exposto no capítulo 2, o **nem sempre** é um operador argumentativo que marca a freqüência da negação da qualidade expressa no enunciado. Exerce uma função adverbial. Com sua utilização, demonstra-se, por exemplo, que a propagação de ideologias não contribui o tempo todo para o exercício democrático. Não é freqüente. O **que** é um pronome relativo.

4.6 6º padrão: **que nem ... nem (alternância)**

(36) O piloto, Doug Roesch, de 34 anos, também morreu no acidente, ocorrido em meio a mau tempo. Alan Pollock, porta-voz da Junta de Segurança do Transporte Nacional, disse **que nem** o piloto **nem** o helicóptero estavam autorizados a voar sob as Normas de Vôo com Instrumentos.

Conforme Almeida (1982, p. 348), quando repetido, o **nem** implica separação de idéias; diz-se então conjunção alternativa: “*Nem um, nem outro*”. O **que** é conjunção integrante do verbo *dicendi*. Na maior parte das 36 ocorrências (5,94%), o **que** exerce função de conjunção.

4.7 7º padrão: **que nem mesmo**

(37) Enquanto aumentar a distância entre o Primeiro e Terceiro Mundo, a Aids estará atingindo gente **que nem mesmo** sabe da existência do HIV, porque é desprovida de seus direitos mais básicos como alimentação, emprego, saúde, saneamento e habitação.

4.8 8º padrão: **que nem sequer**

(38) A gritaria não é maior porque, em essência, quem paga mais são os setores mais vulneráveis – os mais organizados, inclusive os sindicais, defendem-se em maior ou menor grau. Nada parecido com o pobre coitado **que nem sequer** tem conta em banco e não consegue abrir uma caderneta de poupança. E gente só ganha destaque quando está morrendo de fome no Nordeste, provocando arrastão nas praias, roubos ou seqüestros.

Os padrões 7 e 8 são equivalentes e, se somados, totalizarão 53 ocorrências (**que nem mesmo** = 34 = 5,61% e **que nem sequer** = 19 = 3,13%).

Conforme exposto no cap. 2, a negação **nem mesmo** afeta a escala argumentativa como um todo, no sentido que cria uma outra escala. Afinal, o argumento mais forte vem introduzido por essa negação. Logo, subentende-se que haverá outra escala com argumentos mais fracos. O que ocorre é que se houver a utilização do *mesmo*, haverá um sentido de inclusão e de ênfase no enunciado que se tornará mais evidente se esse *mesmo* estiver precedido pelo *nem* (**nem mesmo**). Equivale ao **sequer** ou ao **nem sequer**.

No exemplo (37), com a utilização do **nem mesmo**, percebe-se que há ênfase no enunciado e que realmente o argumento mais forte vem introduzido por essa negação, pois, no caso, se se diz que a Aids estará atingindo gente que nem mesmo sabe da existência do HIV, é porque já atingiu aqueles que sabem dessa existência.

Com o **nem sequer**, ocorre o mesmo tipo de justificativa. *O pobre coitado nem sequer tem conta em banco...* A negação introduz o argumento mais forte, revelando a desigualdade.

O **que**, para ambos os casos, funciona como pronome relativo, mas houve ocorrências em que é conjunção.

4.9 9º padrão: **que nem ao menos**

(39) João não pregara olhos, pensativo, com a calva entre as mãos, ao lado da afilhada. - Era o diabo, era o diabo! Até lhe doía a cabeça! Grandíssima besta, a parteira, **que nem ao menos** soubera apanhar a criança! Estúpida! Deixar morrer uma criança tão bem-feita e nutrida! Isso só acontecia a ele, João da Mata.

O padrão **que nem ao menos** (com 05 ocorrências, totalizando 0,83% da amostra) orienta para o mínimo numa escala argumentativa. A parteira não conseguiu fazer nem o mínimo (nem ao menos) o que se podia esperar que fizesse, quanto mais outras ações.

4.10 10º padrão: **que nem + advérbio (só, bem, mais, assim)**

Há uma recorrência bem menor do **que nem** com esses advérbios, totalizando 09, que se dividem em: **que nem só** = 05; **que nem bem** = 02; **que nem mais** = 01; **que nem assim** = 01. Isso significa que esse padrão, assim como o 9º padrão, são mais raros no português brasileiro ou, talvez, nem devam ser considerados padrões, por não haver recorrência de seus termos e por serem diferentes dos demais.

A seguir, um exemplo com cada um dos advérbios que representam exclusão (só), tempo (bem), concessão (assim = mesmo assim não) ou intensidade (mais).

(40) Os plantadores de trigo, os moinhos e as padarias poderão alegar ao ministro –que é homem de religião e cultura– **que nem só** de pão vive o homem. É um tipo de ameaça equivalente àquela que o Criador fez para Adão: novos aumentos deverão estar a caminho tornando mais problemática nossa capacidade de ~~sciar~~ outras fomes que herdamos do pai comum a partir de sua expulsão do paraíso.

(41) A 22.^a Bienal Internacional de São Paulo vai consagrar uma geração de artistas **que nem bem** chegou ao mercado –a dos anos 90. Dos 23 escolhidos, 14 nunca participaram da Bienal. Dos 14 estreadores, dez começaram a fazer exposições individuais nesta década.

(42) Era também agilíssimo no jogo da faca; com os pés atados bem juntos e com uma faca em punho desafiava a qualquer, e com tal destreza se defendia **que nem assim** o podiam tocar. Era também hábil em manejar as armas

dos índios, e sabia atirar a flecha como o mais destro dos Caiapós ou dos Chavantes.

(43) O Carnaval é um momento de grande alegria e onde muitos aproveitam para se soltar e descarregar todas as neuras e tensões. Para o povo brasileiro, o Carnaval é como um jato de esperança para muitos **que nem mais** acreditam em seus governantes e nas mudanças que estes poderiam eventualmente fazer.

5 CONSIDERAÇÕES FINAIS

Através desse breve estudo sobre o item lexical **que nem**, foi possível demonstrar que a análise lingüística com base em corpus é um grande recurso para investigar os usos da linguagem, já que essa abordagem realmente permite conhecer uma língua e saber como as pessoas a estão utilizando. Isso é possível graças à observação de uma grande quantidade de exemplos de usos reais da língua, extraídos de um corpus volumoso e diversificado. A utilização de recursos computacionais facilita o trabalho do pesquisador quanto à coleta e seleção dos dados, devido à agilidade do computador para processar informações, além de trazer resultados mais confiáveis que podem contribuir para um aprofundamento do estudo da língua portuguesa. Além disso, os resultados podem auxiliar na compreensão que se tem da língua, já que a discussão partiu dos dados reais que correspondem a um fato da língua em evidência em textos escritos.

Quando se consulta uma obra de referência como a gramática ou o dicionário, normalmente os exemplos ali contidos são inventados ou são de autores consagrados e as respostas e/ou análises são breves e prontas, sem contar que, muitas vezes, há ocorrências que nem são citadas nessas obras ou não são tão esclarecidas quanto deveriam. Isso ficou comprovado com essa pesquisa. Fica-se, então, com uma visão muito restrita sobre os fatos da língua, que é dinâmica. Além disso, os dados da realidade são muito mais ricos do que um lexicógrafo ou gramático poderia imaginar. Assim, uma abordagem baseada em corpus permite também que a compreensão sobre a língua seja mais abrangente em comparação com a compreensão adquirida através das referidas obras.

Há outras obras de referência (como as citadas nesse trabalho) que são de uso mais limitado. Certos usos da linguagem ficam conhecidos apenas por pesquisadores ou interessados no assunto. Conseqüentemente, talvez por desconhecimento ou por falta de interesse dos professores, nas aulas de Língua Portuguesa, muitas vezes os conteúdos são tratados com superficialidade, da mesma forma como são tratados pela Gramática Tradicional, que ainda é a principal fonte de consulta sobre a língua na comunidade escolar, quando o é. Isso significa que, utilizar um corpus como fonte de

informações sobre as línguas humanas, já que nele ocorrem exemplos que correspondem aos reais usos da língua, enriquece as pesquisas lingüísticas, pois contempla dados diferentes dos conhecidos e torna os resultados dessas pesquisas mais confiáveis.

Nesta pesquisa, pôde-se verificar que as gramáticas ou dicionários contemplaram o uso do **nem** com o sentido de **sequer, ao menos, como**; do **nem** com função de advérbio, exprimindo **negação** ou como conjunção **alternativa**. Esses sentidos ou usos foram confirmados através dos dados. Outros usos nem apareceram na amostra do corpus, como do **nem** – conjunção aditiva. Isso se justifica, pois o objeto de estudo era o **que nem**. Não caberia, portanto, o uso de **nem** como conjunção aditiva se ele estava acompanhado pelo **que**. Pela mesma razão, não há ocorrências no corpus do **nem** com o sentido de **e não, e sem e ou**.

E quanto aos outros casos de padrões encontrados na amostra que não são contemplados pelas obras de referência? Deve-se agir como se eles não existissem? Deve-se negar sua importância para a construção do texto? Ou será preciso aprofundar seu estudo e criar condições para que sejam conhecidos pelos falantes? Conhecidos eles já são, ou melhor, são utilizados pelos falantes, porém o que muitos não sabem é que esses elementos possuem grande força argumentativa. Trata-se dos operadores argumentativos como o **que nem sempre, que nem mesmo, que nem sequer, que nem ao menos**. Esses elementos operam no discurso de forma a conduzir o interlocutor a determinada conclusão e são os responsáveis, em grande parte, pela força argumentativa dos textos, conforme Koch (1992, p.39), porém têm merecido pouca atenção nos livros didáticos e nas aulas de língua portuguesa.

É bem provável que pesquisas baseadas em corpora tragam à tona muitos elementos que não estão sendo contemplados pelas obras de referência, mas que fazem parte do real uso da língua. Essas descobertas suscitam reflexões. Talvez fosse necessário integrar, de alguma forma, esses elementos a tais obras ou, talvez, deveria haver uma espécie de reformulação no estudo/ensino da língua de forma a abranger as novas descobertas.

Outra questão levantada pela pesquisa estava relacionada à utilização do **que nem = como** (locução conjuntiva comparativa) em textos extraídos da língua escrita.

Na amostra pesquisada, foi interessante descobrir que esse item lexical é recorrente em textos de língua escrita, mas como parte da oralidade, ou seja, ele apareceu, com maior destaque, na fala de personagens de textos literários, em entrevistas e em textos de opinião dos jornais. Portanto, textos com marcas de oralidade. Sendo assim, é um item que pertence à modalidade oral da língua, realmente. Portanto, é muito provável que, em gêneros textuais mais formais como os textos jurídicos, ela não se fará presente.

Outro aspecto observado é que, a locução conjuntiva comparativa é mais utilizada em sintagmas nominais. Os termos de uma comparação normalmente são substantivos comparados a substantivos (e não tanto adjetivos, como se esperava), por exemplo: “...*cabelos compridos que nem esfregão de cabeça para cima..*”. “Cabelos” e “esfregão” são substantivos e estão sendo comparados. Porém esperava-se uma maior quantidade de adjetivos sendo comparados com substantivos, como “*vermelho que nem tomate*”.

Por ser este um breve estudo sobre um item lexical presente na nossa língua, com certeza há possibilidade de ampliar as discussões e de aprofundá-lo. Por não haver discussão suficiente para servir de apoio ou por ser um primeiro contato com o assunto, classificar e analisar os dados, por exemplo, em algumas etapas, revelou-se uma tarefa árdua, porém instigante, porque foi possível observar como as pessoas lidam com a linguagem. Nesse sentido, pode-se considerar que o estudo foi proveitoso, pois suscitou reflexões a respeito da língua e de seus usos, principalmente no que diz respeito à compreensão das relações que as expressões são capazes de fazer no contexto da sentença e isso é mais importante do que simplesmente classificar determinado item conforme a Gramática Tradicional apregoa, até porque, conforme visto, o estudo demonstrou que, no uso real, as palavras contêm algo a mais do que aquilo que está estabelecido nas obras de referência. Isso comprova que a abordagem com base em corpus é produtiva e o conhecimento sobre milhares de palavras da língua seria enriquecido a partir de análises que reunissem os usos que as circundam. Assim, poderia haver um melhor entendimento do léxico-gramática do português e uma maior respeito aos seus usuários, até um aperfeiçoamento de um sistema de informática para processar línguas humanas.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, Napoleão M. de. **Gramática metódica da língua portuguesa**. 31ª ed. – São Paulo: Saraiva, 1982.

AURÉLIO. **Novo dicionário da língua portuguesa**. – 2. ed. revista e aumentada (5ª impressão) – Rio de Janeiro: Nova Fronteira, 1986.

BASÍLIO, Margarida. **Teoria lexical**. São Paulo: Ática, 1998.

BAKHTIN, Mikhail Mikháilovitch. **Estética da criação verbal**. 3ª ed. São Paulo: Martins FONTES, 2000.

BECHARA, Evanildo. **Moderna gramática portuguesa**. - 37ª ed. revista e ampliada – Rio de Janeiro: Lucerna, 2003.

BERBER SARDINHA, T. **Linguística de corpus**. Barueri, SP: Manole, 2004.

BEZERRA, M. A. Ensino de língua portuguesa e contextos teórico-metodológicos. In: DIONÍSIO, Ângela Paiva; MACHADO, Ana Raquel; BEZERRA, Maria Auxiliadora (Org.). **Gêneros textuais & ensino**. 2ª ed. – Rio de Janeiro: Lucerna, 2003.

CEGALLA, Domingos Paschoal. **Novíssima gramática da língua portuguesa**. – 31ª ed. – São Paulo: Companhia Editora Nacional, 1989.

CUNHA, C. e CINTRA, L. **Nova gramática do português contemporâneo**. 2ª ed. (17ª impressão). Rio de Janeiro: Nova Fronteira, 1985.

DUCROT, Oswald. **Princípios de semântica lingüística**. Trad. Brasileira: São Paulo, Cultrix, 1977. (Original francês: 1972)

HOFFNAGEL, Judith C. Entrevista: uma conversa controlada. In: DIONÍSIO, Ângela Paiva; MACHADO, Ana Raquel; BEZERRA, Maria Auxiliadora (Org.). **Gêneros textuais & ensino**. 2ª ed. – Rio de Janeiro: Lucerna, 2003.

HOUAISS, A. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2001.

HUNSTON, S. **Corpora in applied linguistics**. United Kingdom: Cambridge University Press, 2002.

ILARI, R. e GERALDI, J.W. **Semântica**. 10ªed. (6ª impressão) – São Paulo: Ática, 2003.

KENNEDY, G. **An introduction to corpus linguistics**. London and New York: Longman, 1998.

KOCH, Ingedore Grunfeld Villaça. **A inter-ação pela linguagem**. (Coleção repensando a língua portuguesa). São Paulo: Contexto, 1992.

_____. **Argumentação e linguagem**. 3ª ed. – São Paulo: Cortez, 1993.

_____. **Desvendando os segredos do texto**. – 2ª ed. – São Paulo: Cortez, 2003.

LEECH, G. Corpora and theories of linguistic performance. In: **Corpus linguistics proceedings of Nobel Symposium 82**. Stockholm 4-8 August. Berlim, 1992.

_____. Introducing corpus annotation. In: **Corpus annotation: linguistic information from computer text corpora**. London and New York: Longman: 1997.

MARCUSCHI, L.A. Gêneros textuais: definição e funcionalidade. In: DIONÍSIO, Ângela Paiva; MACHADO, Ana Raquel; BEZERRA, Maria Auxiliadora (Org.). **Gêneros textuais & ensino**. 2ª ed. – Rio de Janeiro: Lucerna, 2003.

_____. Letramento e oralidade no contexto das práticas sociais e eventos comunicativos. In: MARCUSCHI...[et al.]; SIGNORINI, Inês (org.). **Investigando a relação oral/escrito**. Campinas, SP: Mercado das Letras, 2001. (Coleção Idéias sobre Linguagem)

McENERY, T. e WILSON. **Corpus linguistics**. Edinburgh: Edinburgh University Press, 1996.

MOURA, H. M. de M. A língua popular tem razões que os gramáticos desconhecem. In: SILVA, F. L. da e MOURA, H. M. de M. (Org.). **O direito à fala: a questão do preconceito lingüístico**. 2ª ed. rev. - Florianópolis: Insular, 2002.

NEVES, Maria Helena de Moura; HATTNER, Marize Dall'Aglio. **Construções comparativas**. In: ABAURRE, Maria Bernardete; RODRIGUES, Ângela (orgs.). **Gramática do português falado**. Campinas, SP: Editora da UNICAMP, 2002.

PARTINGTON, A. **Patterns and meanings: using corpora for English Language Research and Teaching**. Amsterdam, John Benjamins, 1998.

PERINI, M.A. **Para uma nova gramática do Português**. 9ª ed. Série Princípios. São Paulo: Ática, 1999.

POSSENTI, Sírio. **Por que (não) ensinar gramática na escola**. Campinas, SP: Mercado das Letras, 1996.

ROCHA, Marco. O uso de corpora computadorizados no ensino de língua portuguesa: metodologia e avaliação. In: CABRAL, L.G.; SOUZA, P.; LOPES, R.V.; PAGOTTO, E.G. (org.). **Lingüística e ensino: novas tecnologias**. Blumenau: Nova Letra, 2001.

RODRIGUES, Rosângela Hammes. **A relação entre gênero, enunciado e texto: uma leitura bakhtiniana**. In: Boletim da ABRALIN, V. 26, nº Especial, 2001.

SACCONI, Luiz Antônio. **Nossa gramática (teoria e prática)**. - 25ª ed. revista e atualizada. – São Paulo: Atual, 1999.

SIGNORINI, Inês. Construindo com a escrita “outras cenas de fala”. In: MARCUSCHI, L.A.[*et al*]; SIGNORINI, Inês (org.). **Investigando a relação oral/escrito e as teorias do letramento**. Campinas, SP: Mercado das Letras, 2001.

SINCLAIR, John M. **Corpus, concordance, collocation**. Oxford, Oxford University Press, 1991.

_____. Corpus evidence in language description. In: **Teaching and language**. London and New York: Longman, 1997.

STUBBS, Michael. **Text and corpus analysis: computer-assisted studies of language and culture**. Massachusetts: Oxford, 1998.

VOGT, Carlos. **O intervalo semântico**. São Paulo, Ática, 1977.