

VERA LÚCIA VASILEVSKI DUARTE

**USO DO ITEM LEXICAL 'TODO' E SUAS FLEXÕES EM
TEXTOS ESCRITOS: UMA ABORDAGEM A PARTIR DE *CORPUS***

Florianópolis
2003

VERA LÚCIA VASILEVSKI DUARTE

**USO DO ITEM LEXICAL ‘TODO’ E SUAS FLEXÕES EM
TEXTOS ESCRITOS: UMA ABORDAGEM A PARTIR DE *CORPUS***

Dissertação apresentada ao Curso de Pós-graduação em Lingüística da Universidade Federal de Santa Catarina (UFSC), como requisito para obtenção do grau de Mestre em Lingüística; área de concentração: Lingüística Aplicada; linha de pesquisa: *Corpus* e gênero: análise e aplicações.

Orientador: Prof. Dr. Marco Antônio Esteves da Rocha

Florianópolis
2003

Ficha Catalográfica (elaborada pela autora)

VASILEVSKI DUARTE, V. L. **Uso do item lexical ‘todo’ e suas flexões em textos escritos**: uma abordagem a partir de *corpus*. 2003. 120f. Dissertação (Mestrado em Linguística Aplicada) – Curso de Pós-graduação em Linguística. Universidade Federal de Santa Catarina (UFSC). Florianópolis.

1.*Corpus* lingüístico. 2.Gramática tradicional. 3.Língua escrita.

Organizado segundo as normas da Associação Brasileira de Normas Técnicas (ABNT).

UFSC/CCE/CPGLg

VERA LÚCIA VASILEVSKI DUARTE

**USO DO ITEM LEXICAL ‘TODO’ E SUAS FLEXÕES EM
TEXTOS ESCRITOS: UMA ABORDAGEM A PARTIR DE *CORPUS***

Esta dissertação foi aprovada como requisito para a obtenção do título de Mestre em Lingüística, pelo programa de Pós-graduação em Lingüística da Universidade Federal de Santa Catarina.

Prof. Dr. Heronides Maurílio de Melo Moura
Coordenador do Programa de Pós-graduação em Lingüística
da Universidade Federal de Santa Catarina

Banca examinadora:

Prof. Dr. Marco Antônio Esteves da Rocha
Orientador
Universidade Federal de Santa Catarina

Prof. Dr. Philipe Humblé
Universidade Federal de Santa Catarina

Prof. Dr. Antônio Paulo Berber Sardinha
Pontifícia Universidade Católica de
São Paulo

Florianópolis, agosto de 2003.

***A** minha querida mãe, Lúcia, e ao meu irmão Marcelo, meu Nenê, por tudo o que passamos e por nossas vitórias. Este trabalho representa mais uma... e pertence a nós três. Com proteção, amizade e cumplicidade, vocês me impulsionam. Amo vocês.*

AGRADECIMENTOS

Quando se pensa em agradecer, depois de um trabalho longo como uma dissertação, começam a surgir na mente as etapas da pesquisa e as pessoas que se vincularam a elas, por algum motivo. Quando faço isso, lembro-me das professoras **Lúcia Locatelli Flôres** e **Lúcia Maria Nassib Olímpio** e dos ensinamentos sempre permeados de discussões fundamentais e animadas que muito contribuíram para a qualidade formal e teórica deste trabalho, com elas, percebi que é mais gostoso aprender a teoria quando os motivos, as reflexões, que levaram teóricos a certas conjecturas são apresentados e debatidos; e do professor **Emílio Pagotto**, os materiais afins e também as longas e preciosas discussões que partilhamos.

Já a **Iza**, minha amiga de infância, apareceu na minha casa por um imprevisto do destino, ajudou nos arremates do trabalho e trouxe conforto com sua companhia, que lembra uma época mágica da minha vida; também, sem saber, revigorou meu ânimo de continuar firme até o final desta empreitada, com sua maneira intrépida de enfrentar problemas reais e dolorosos.

Também cabe agradecer aos professores **Kanavilill Rajagopalan** as suas considerações sobre o projeto deste trabalho, **Marco Rocha** ter sido meu orientador, **Philippe Humblé** e **Tony Berber Sardinha** terem aceitado compor a banca examinadora; à CAPES a bolsa concedida por 17 meses, sem a qual este estudo não seria possível; e à Secretaria da Pós-graduação em Lingüística da UFSC.

Embora a ciência construa-se com dados experimentais, da mesma forma que uma casa se constrói com tijolos, uma coleção de dados experimentais ainda não é ciência, da mesma forma que uma coleção de tijolos não é uma casa.

Poincaré

RESUMO

Pesquisa que se vale de ocorrências naturais de uso da língua combinadas a uma metodologia específica para investigar o comportamento dos vocábulos ‘todo’, ‘todos’, ‘toda’ e ‘todas’, bem como suas formas diminutivas, em textos escritos reconhecidos como pertencentes à norma culta da língua portuguesa do Brasil, a fim de construir um modelo que retrate o funcionamento desses itens lexicais. A base metodológica da análise dos dados é a lingüística de *corpus*, a lingüística computacional e a estatística – uma vez que se tratam os dados qualitativa e quantitativamente –, sobre as quais algumas considerações prévias são feitas. A averiguação apóia-se em uma amostra de 1.013 ocorrências, propriamente extraídas dentre uma população de mais de 56 mil. Em determinados momentos, a análise extrapola a amostra e leva em conta a população – com isso, tenciona-se aumentar a credibilidade e abrangência do estudo. Os padrões de ocorrência dos itens lexicais são obtidos a partir de informações coletadas exclusivamente dos dados, por meio da classificação deles – baseada na gramática tradicional –, emprego de fórmulas e testes matemáticos e gerenciamento eletrônico; e confrontados com o trato dado aos itens lexicais em questão por alguns teóricos da língua. Após discussões e reflexões, as informações são reunidas sob a forma de um modelo de funcionamento que visa a cobrir os usos dos itens de trabalho na língua portuguesa escrita no Brasil. Demonstra que o uso dos itens lexicais analisados diverge do que preconizam as teorias da língua.

Palavras-chave: ocorrências naturais, textos escritos, *corpus*, modelo de funcionamento.

ABSTRACT

This research combines natural occurrences of language usage to a specific handling methodology to investigate the behavior of the Portuguese words 'todo' (comparable to English 'all'), its inflexions – 'toda', 'todos', 'todas' –, and diminutive forms as well, in written texts published in Brazil, in order to build a functioning model for these lexical items. The methodological basis of the data analysis is *corpus* linguistics, computational linguistics, and statistics. Because of their importance to this study, some considerations about them are made previously. Data is treated quantitatively and qualitatively. The analysis takes on account a 1.013 data sample, properly collected among close to 56.000. At certain points, when it is possible, the analysis deals with all data, towards increasing research credibility. The patterns of occurrence for these lexical items are obtained from information gathered restrictively from data, by classifying them – according to traditional grammar nomenclature –, applying mathematical tests, and electronically manipulating them. Next, these patterns are discussed and compared to the treatment some language theories give these lexical items, and, then, the information is draw together to form a general functioning model of the words analyzed. The results show that the use of the lexical items analyzed is not totally covered by language theories.

Keywords: natural occurrences, written texts, functioning model, *corpus*.

SUMÁRIO

LISTA DE ILUSTRAÇÕES.....	11
ARGUMENTOS E COMPOSIÇÃO DO ESTUDO.....	13
1 CORPUS LINGÜÍSTICO E CORPUS DE TRABALHO.....	19
1.1 CORPUS LINGÜÍSTICO	19
1.1.1 Consolidação nos Estudos da Língua	19
1.1.2 Características	21
1.1.2.1 Representatividade	22
1.1.2.2 Considerações sobre o conteúdo	24
1.2 CORPUS DE TRABALHO	25
1.2.1 Impropropriedades	28
1.2.2 Corpus de Apoio	29
2 LINGÜÍSTICA COMPUTACIONAL, LINGÜÍSTICA DE CORPUS E ESTATÍSTICA: O MÉTODO	31
2.1 ABORDAGEM UTILIZADA	31
2.1.1 Lingüística de <i>Corpus</i>	32
2.2 LINGÜÍSTICA COMPUTACIONAL	35
2.3 ANÁLISES ESTATÍSTICAS	37
2.3.1 Freqüência	38
2.3.2 População e Amostra	39
2.4 CATEGORIZAÇÃO DOS DADOS	43
3 ANÁLISE	45
3.1 PREDEFINIÇÕES	45
3.2 PERFIL GENÉRICO DOS ITENS LEXICAIS	47
3.2.1 Freqüência Geral e Modelo Probabilístico	48
3.3 TODA	50
3.3.1 Estrutura, Função Sintática e Combinação Semântica	50
3.3.2 Significado	57
3.3.3 Alternância da Presença de Artigo e Ambigüidade	60
3.3.4 Locuções ou Frases Feitas	62
3.4 TODO	63
3.4.1 Estrutura, Função Sintática e Combinação Semântica	63
3.4.2 Significado.....	68
3.4.3 Alternância da Presença de Artigo e Ambigüidade.....	70
3.4.4 Locuções ou Frases Feitas.....	76
3.5 TODAS	78
3.5.1 Estrutura, Função Sintática e Combinação Semântica	78
3.5.2 Significado	82

3.5.3 Alternância da Presença de Artigo	83
3.5.4 Locuções ou Frases Feitas	84
3.6 TODOS	85
3.6.1 Estrutura, Função Sintática e Combinação Semântica	85
3.6.2 Significado	89
3.6.3 Alternância da Presença de Artigo	90
3.6.4 Mobilidade da Flexão	93
3.6.5 Locuções ou Frases Feitas	94
3.7 TODINHA, TODINHO, TODINHAS E TODINHOS	94
3.8 <i>CORPUS</i> DIACRÔNICO	97
4 PADRÕES DE OCORRÊNCIA	100
4.1 SIGNIFICADO E FUNÇÕES SINTÁTICAS	100
4.1.1 A Questão dos Artigos Sucedendo ‘tod*’	100
4.1.2 Valor Nominal e Valor Adverbial	102
4.1.3 Funções e Significados	106
4.2 RESPOSTAS	108
4.3 MODELO DE FUNCIONAMENTO	109
CONSIDERAÇÕES FINAIS	113
REFERÊNCIAS.....	117
APÊNDICE	119

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de aplicação de filtro aos dados e ‘anotação’ em planilha eletrônica	36
Gráfico 1 – Distribuição em % dos itens de trabalho	49
Quadro 1 – Nomenclatura utilizada na categorização dos dados	47
Quadro 2 – Adjacências de ‘toda’	51
Quadro 3 – Adjacências de ‘todo’	63
Quadro 4 – Adjacências de ‘todas’	78
Quadro 5 – Adjacências de ‘todos’	85
Quadro 6 – Comportamento dos itens lexicais estudados quanto ao significado	107
Quadro 7 – Modelo de funcionamento de ‘todo’ e ‘toda’	109
Quadro 8 – Modelo de funcionamento de ‘todos’ e ‘todas’	111
Tabela 1 – População: total de ocorrências dos itens pesquisados	40
Tabela 2 – Amostra: ocorrências selecionadas para análise	41
Tabela 3 – Quantidade de dados a ser selecionados para se tolerar erros amostrais de 2% e 3%	42
Tabela 4 – Formas diminutivas do item de trabalho no <i>corpus</i>	42
Tabela 5 – Percentual e ordem de ocorrência dos itens em estudo no <i>corpus</i>	47
Tabela 6 – Modelo probabilístico de ‘tod*’+formas diminutivas	49
Tabela 7 – Função do sintagma que contém ‘toda’	52
Tabela 8 – Função de ‘toda’ no sintagma em que está inserida	53
Tabela 9 – Trigramas notáveis de ‘toda’	55
Tabela 10 – Cotejo quantitativo entre as estruturas <i>toda+art+subst</i> e <i>art+subst+toda</i> com sentido equivalente	56
Tabela 11 – Noções semânticas da flexão ‘toda’ na amostra e percentual de ocorrência de cada uma	57
Tabela 12 – Função do sintagma que contém ‘todo’	64
Tabela 13 – Função de ‘todo’ no sintagma em que está inserida	65
Tabela 14 – Trigramas notáveis de ‘todo’	66
Tabela 15 – Cotejo quantitativo entre as estruturas <i>todo+art+subst</i> e <i>art+subst+todo</i> com sentido equivalente	67
Tabela 16 – Noções semânticas de ‘todo’ na amostra e percentual de ocorrência de cada uma	69
Tabela 17 – Cotejo quantitativo entre as estruturas ‘todo mundo’ e ‘todo o mundo’	72

Tabela 18 – Dados para o primeiro teste qui-quadrado	74
Tabela 19 – Freqüências observadas do primeiro teste qui-quadrado	75
Tabela 20 – Freqüências esperadas do primeiro teste qui-quadrado	75
Tabela 21 – Estatística do primeiro teste qui-quadrado	75
Tabela 22 – Função do sintagma que contém ‘todas’	79
Tabela 23 – Função de ‘todas’ no sintagma em que está inserida	79
Tabela 24 – Trigramas notáveis de ‘todas’	81
Tabela 25 – Noções semânticas de ‘todas’ na amostra e percentual de ocorrência de cada uma	82
Tabela 26 – Ocorrências de ‘todas’ em que há ou não artigo antes de pronome Possessivo	84
Tabela 27 – Função do sintagma que contém ‘todos’	86
Tabela 28 – Função de ‘todos’ no sintagma em que está inserida	87
Tabela 29 – Trigramas notáveis de ‘todos’	88
Tabela 30 – Noções semânticas de ‘todos’ na amostra e percentual de ocorrência de cada uma	90
Tabela 31 – Ocorrências de ‘todos’ em que há ou não artigo antes de pronome possessivo	91
Tabela 32 – Dados para o segundo teste qui-quadrado	92
Tabela 33 – Freqüências observadas do segundo teste qui-quadrado	92
Tabela 34 – Freqüências esperadas do segundo teste qui-quadrado	93
Tabela 35 – Estatística do segundo teste qui-quadrado	93
Tabela 36 – Adjacências das formas diminutivas e quantidades	95
Tabela 37 – Função do sintagma que contém as formas diminutivas	96
Tabela 38 – Função dos diminutivos no sintagma de que fazem parte	96
Tabela 39 – Quantias de ‘tod*’ no <i>corpus</i> diacrônico	98
Tabela 40 – Trigramas notáveis de ‘tod*’ no <i>corpus</i> diacrônico	98

ARGUMENTOS E COMPOSIÇÃO DO ESTUDO

Desde o início, na vida de estudante da pesquisadora, a língua portuguesa, sobretudo a gramática, foi a disciplina de maior destaque. Achava incrível a teoria e sempre queria mais do professor. Questionava. Com o avançar da escolarização, e depois muitas respostas rudes ou pouco esclarecedoras de seus professores, passou a ver a gramática como um conjunto de regras elaboradas para usar na prática. Perdeu a poesia de identificar nos ensinamentos do professor a explicação da magia que eram as combinações de palavras feitas para transmitir emoções ao mundo. Na oitava série, conheceu os períodos compostos e achou-os o máximo. Chegou a ficar decepcionada quando a professora disse que aquilo não era importante e não cairia em prova.

Depois não houve surpresas. Já na faculdade de Letras, percebeu que não atingia seu objetivo de examinar qualquer sentença da sua língua e teorizar, com nexos, as relações entre as palavras que a compunham. Antes mesmo de formada, em sala de aula, trabalhando com pessoas, entendeu que os professores que tivera outrora não compreendiam muito bem o que ensinavam – apenas se limitavam a repetir os dogmas dos teóricos que adotavam – e que seus questionamentos em sala, decorrentes da empolgação pela matéria, soavam como desafios a eles. Os alunos dela também perguntavam e, não raro, era necessário dizer que a resposta seria dada no dia seguinte, pois precisava pesquisar.

A sua sensação diante de uma dúvida dos alunos era de contentamento, pois eles manifestavam interesse pelo conteúdo e usavam sua criatividade e habilidade na língua materna ao questionar. Além disso, demonstravam gostar da matéria. Essas constatações foram as principais causas que a determinaram a perseguir aquele objetivo, para o que contribuíram algumas matérias em norma padrão escrita que professores corajosos decidiram ofertar na universidade, como disciplinas optativas, inacreditavelmente.

Não queria submeter um aluno seu ao que seus professores a submeteram, pois a rudeza e a falta de respostas são fatores que fazem com que as pessoas passem a repudiar o português na escola e transponham essa carga negativa para outros aspectos da língua. O resultado disso pode ser percebido todos os dias, na insegurança das pessoas ao escrever, nos empregos perdidos, porque, apesar do bom desempenho nos exames de inglês, informática e na parte técnica, a vaga foi para quem fez a melhor redação.

A partir do trabalho em sala de aula, constatou que essa insegurança se relaciona intimamente à gramática. Além disso, até mesmo lingüistas que são contra o ensino da norma culta já assumem que a função da escola é ensinar gramática padrão. Assim, é útil e natural desenvolver trabalhos que enfoquem essa área, para que o ensino aconteça de forma mais consciente e se incentive o aluno a refletir, a confiar no seu potencial nato de usar a sua língua materna. Para isso queria colaborar.

Com essa intenção em mente, já no mestrado, foi apresentada, pelo orientador, à Lingüística de *corpus*, a qual veio a apoiar seu intuito de explicar usos lingüísticos (pronunciamentos efetivos, reais, escritos ou falados) a partir da teoria gramatical, e com ela surgiu a chance de resgatar a poesia de outrora e contribuir para que outros alunos não perdessem a sua. Faltava o item de pesquisa. Eram tantos a escolher. Via em cada palavra um horizonte promissor para seu estudo e, em cada opção, a oportunidade de trabalhar com gramática.

Certa vez, em uma aula de semântica, o professor defendia que havia diferença de significado entre sentenças como *Toda mulher quer casar* e *Toda a mulher quer casar*. Imediatamente, veio-lhe à mente a frase tão repetida pelo padre nas missas de domingo que freqüentava quando era criança: *Todo o homem que crê em Deus será salvo*. Não percebeu, a princípio, por que haveria diferença de sentido entre as duas primeiras, até que um colega de classe explicou-lhe. Brotou nela uma inquietação. Começava a vislumbrar um objeto de estudo, o qual se consolidou quando uma amiga sua comentou que tinha ido *toda linda* a uma festa para impressionar um rapaz. Pensou nas relações entre aqueles termos. Cresceu uma excitação. E floresceu um tema que considerou perfeito.

Uma vez descrito o despontar do assunto a ser tratado neste trabalho, a seguir,

são expostos os principais esclarecimentos introdutórios da pesquisa, como suas justificativas, seus propósitos e a forma como se encontra organizada.

Sabe-se que, por muito tempo, preferiu-se trabalhar em pesquisa lingüística com dados criados especialmente para análise, conseqüentemente, artificiais, enquanto ocorrências naturais não eram valorizadas. Isso fragilizava os resultados encontrados por causa da distância entre eles e o uso efetivo da língua, ou seja, situações reais. Em razão dessas constatações, esta pesquisa privilegia a observação de dados naturais de uso da língua combinada a uma metodologia específica para analisar o comportamento dos vocábulos ‘todo’, ‘todos’, ‘toda’ e ‘todas’, bem como suas formas diminutivas, em textos escritos. A escolha desse fenômeno para observação deu-se por ‘todo’ e suas flexões serem, reconhecidamente, palavras de transição entre as chamadas classes dos pronomes, dos adjetivos e também dos advérbios. Por conta disso, tenciona-se verificar como esse trespasse aparente acontece na prática, a partir da investigação da estrutura das sentenças que os contêm e dos significados que tais termos assumem no discurso escrito.

Apesar de a metodologia adotada admitir que seja desenvolvida uma interpretação do fenômeno baseada na intuição do pesquisador para subseqüente confronto com os resultados da análise, optou-se por não utilizar predefinições dessa natureza, mesmo intuitivas, e sim, isento de qualquer expectativa, observar e classificar os dados propriamente selecionados para trabalho. É de fundamental importância que eles sejam autênticos e que o contexto seja preservado, uma vez que o propósito geral do estudo é averiguar o desempenho da língua em seu dinamismo natural. A partir da análise de um número significativo de dados, a pesquisadora definirá classificações para os tipos de ocorrências encontradas. É válido frisar que não se deseja a confirmação de concepções prévias, mas sim explorar o fenômeno lingüístico escolhido; embora as classificações tenham de ser baseadas em teorias lingüísticas.

Aqui cabe realçar que, ao trabalhar com dados naturais, em vários momentos, o pesquisador vê-se diante de elementos que desafiam as classificações sugeridas nas teorias – bem como mostram os limites delas – e assume a tarefa de propor categorizações mais coerentes. Essa incumbência é, não raro, angustiante, sobretudo

quando se trabalha com o significado assumido pelas palavras no uso, pois a subjetividade do usuário da língua está envolvida. Não obstante, a tarefa deve ser cumprida e as proposições do pesquisador melhoradas, se for o caso, a partir de discussões, justamente como feito aqui com as teorias. Por isso, ao longo da pesquisa, não se economizam exemplos que comprovem achados e suscitem comentários e reflexões.

Embora o estudo assente-se na descrição de exemplos de usos lingüísticos – e não na prescrição deles –, não haja comprometimento com postulados de nenhuma teoria, apenas com a metodologia, e os resultados estejam fundamentados no uso real da língua, isso não impede que, após as análises, algumas teorias sirvam de apoio para refinamento de resultados, conquanto não sirvam de base.

O aspecto quantitativo – que, neste caso, está relacionado à constância do uso de elementos da língua – recebe destaque na abordagem utilizada, de maneira que o tratamento de noções matemáticas como frequência e probabilidade deve ser feito a partir de técnicas de análise estatística e procedimentos computacionais. Por conseguinte, neste estudo, destaca-se o trio formado por as lingüísticas computacional e de *corpus* – um par adequado para investigações sobre a língua, que inaugura uma nova filosofia nessa área – e a estatística – por sua excelência no tratamento quantitativo e na testagem dos dados.

Na realidade, essas três ferramentas completam-se e propiciam uma metodologia científica, com a vantagem de os resultados obtidos poderem ser vistos e comprovados por todos. Integrar esses instrumentos neste tipo de pesquisa não constitui tarefa muito simples, dado que não foi encontrado material auxiliar que o fizesse em língua portuguesa – pelo menos não nos termos a que se propõe este trabalho – e pudesse servir de exemplo; aliás, a própria literatura utilizada é em língua estrangeira ou, quando em português, baseada nela.

A partir da investigação descritiva proposta, especificamente, pretende-se determinar os padrões de ocorrência de ‘todo’, ‘todos’, ‘toda’, ‘todas’ e suas formas diminutivas na língua portuguesa e, então, estabelecer um modelo de funcionamento sintático-semântico empírico desses vocábulos – um conjunto de elementos que demonstrem o seu uso –, definindo e testando procedimentos computacionais, à

medida que se elabora uma teoria para explicar a ocorrência desses padrões. Tal modelo, espera-se, poderá sugerir direcionamentos no trato dado a esses vocábulos pelas gramáticas tradicionais, pelos dicionários e suscitar reflexões sobre o tema, servindo de base para estudos vindouros em várias áreas da lingüística. Ainda, o tratamento eletrônico específico dado aos itens investigados neste trabalho poderá colaborar para o desenvolvimento da metodologia da lingüística de *corpus* no Brasil, já que aqui pouco se conhece sobre ela ainda, vindo a auxiliar pesquisas futuras nessa área, bem como ser aprimorado por elas. Em âmbito geral, espera-se contribuir para a compreensão da língua sob um ponto de vista empírico. Assim, demonstra-se a relevância desta pesquisa para estudos da língua.

Visando a atingir tais propósitos, esta dissertação está dividida em quatro capítulos principais. Primeiramente, expõem-se definições acerca de *corpus* lingüístico, ressaltando um pouco de sua história, e técnicas de composição a partir de autores reconhecidos, como Sinclair (1991) e Sardinha (2000); além disso, apresenta-se a massa lingüística utilizada na pesquisa – o *corpus* de trabalho e o de apoio, para revalidar os resultados obtidos – e dão-se informações mais detalhadas sobre seus componentes, à medida que se discutem suas virtudes e eventuais problemas.

A seguir, com apoio de autores pertinentes – como Leech (1992), Manning e Schütze (2000) e Barbeta (1998), dentre outros não menos relevantes –, o enfoque cabe à metodologia que norteia a pesquisa: a lingüística de *corpus* (da qual se destaca a trajetória histórica, apontam-se discussões referentes à sua utilização em pesquisas, bem como justificativas e aspectos de seu emprego), a lingüística computacional (sobretudo no que tange aos programas de computador utilizados) e a estatística (definições e aplicações); também se delineiam os procedimentos específicos do estudo, como os critérios de seleção dos dados, os dados selecionados, os testes prévios, a categorização dos dados e os autores consultados para tal; e, enfim, outros esclarecimentos acerca da análise.

O capítulo 3 refere-se à análise dos dados: expõem-se definições gerais e específicas sobre o item lexical em estudo, baseadas nos apontamentos dos programas eletrônicos, resultados numéricos, as classificações (com menção a autores citados no capítulo 2), cotejos (apresentados em quadros, tabelas e gráfico) e algumas

considerações da pesquisadora sobre os achados da análise. À medida que se torna necessário utilizar definições sobre a classificação dos dados, elas são transcritas e referidas.

A seguir, em meio a discussões e reflexões, demarcam-se os padrões de ocorrência dos itens lexicais analisados, oriundos dos dados, à luz de considerações de teóricos da língua – principalmente Neves (1999 e 2003), Almeida (1999) e Said Ali (1964) –, com o objetivo de comprovar sua eficácia e também de expor algum conhecimento histórico sobre os itens lexicais de trabalho, a fim de complementar os resultados da análise. Após isso, o modelo de funcionamento é proposto.

Finalmente, são expostas as considerações finais do estudo, concernentes ao trabalho com uma massa lingüística, à eficiência da metodologia adotada – com a discussão de suas facilidades e de seus entraves –, à análise feita, seus pontos mais importantes; em resumo, à utilização das técnicas e dos procedimentos englobados por este trabalho em pesquisas da língua.

1 *CORPUS* LINGÜÍSTICO E *CORPUS* DE TRABALHO

Estudar a língua a partir da análise de grandes quantidades de textos reais para descrevê-la não é uma prática nova; aliás, com limitações, já era exercida antes da chegada do computador. Entretanto, parece razoável afirmar que a novidade nos estudos recentes com base em *corpus* é pesquisar: critérios adequados de coletar, reunir e utilizar esses textos; que tipo de perguntas podem ser respondidas com base nesses dados; e as demarcações das respostas obtidas. Desse modo, explicações detalhadas acerca do uso de um *corpus* em pesquisa lingüística são necessárias, tendo em vista que aplicações atuais dessa natureza ainda são um tanto esporádicas. Outrossim, é essencial apresentar o *corpus* utilizado nesta pesquisa – a principal ferramenta de trabalho, o meio – e demonstrar o seu vínculo com estudos desenvolvidos na área.

1.1 *CORPUS* LINGÜÍSTICO

Primeiramente, expõem-se informações e considerações gerais concernentes a uma massa lingüística reunida para investigar a língua.

1.1.1 Consolidação nos Estudos da Língua

Conforme Sardinha (2000, p.325), desde a Antigüidade já se faziam conjuntos de documentos, como o *Corpus* Helenístico, na Grécia Antiga, e de citações da Bíblia, que também eram produzidos na Idade Média, por exemplo. Porém, vale ressaltar os estudos lingüísticos que remontam a 1897, quando Käding utilizou um *corpus* de 11

milhões de palavras – quantidade impressionante para a época – a fim de verificar a distribuição e a seqüência de letras em alemão. Como exemplo em língua portuguesa, pode-se citar o trabalho de Said Ali, *Gramática Histórica da Língua Portuguesa*, lançado em 1921, no qual ele revela – sem especificar quantidade – que fez pesquisas exaustivas em tantos escritos quanto pôde, de épocas o mais variadas possível, no intuito de vasculhar o uso de itens lexicais da língua portuguesa. Esses dois exemplos, nos quais obviamente não se tratava de textos eletrônicos, demonstram a importância dada à exploração da língua por meio de evidências empíricas.

No entanto, como expõem McEnery e Wilson (1997, p.4-5), nas décadas de 1950 e 60, Chomsky sugeriu que um *corpus* jamais poderia ser uma ferramenta válida para os lingüistas, que deveriam privilegiar a competência lingüística nos estudos e não o desempenho lingüístico. Chomsky defendia o racionalismo em detrimento do empirismo. Sobre isso, os primeiros autores salientam que uma teoria racionalista é baseada em dados artificiais e julgamentos introspectivos; em contrapartida, numa abordagem empírica, predomina a observação de dados lingüísticos naturais, efetivamente ocorridos. Dessa forma, uma abordagem empírica permite que se trabalhe com evidências, livres de contestação e especulações.

Um *corpus* é uma densa massa lingüística – quanto mais densa, melhor –, objeto de manejo da lingüística de *corpus*; seria, então, um meio de estudo lingüístico em massa. Por natureza, um *corpus* é um conjunto de elocuições, ou seja, expressões do pensamento; é desempenho lingüístico, língua externalizada. Essas elocuições não apenas traduzem a competência de um usuário da língua, como também inúmeros aspectos relacionados com ocasiões específicas de uso dela (p.5).

Contudo, isso não significa que as sentenças de uma língua sejam finitas e contáveis, mas sim que há como se trabalhar com uma amostra legítima de situações de uma língua. Obviamente, não há como todas as sentenças possíveis em uma língua estarem presentes em um *corpus*, mas as que aparecem nele são indiscutivelmente válidas, desde que os textos que o compõem sejam autênticos. Além disso, é normal que haja expectativa de que ocorram vastamente sentenças consideradas usuais, o que, caso não seja confirmado, suscita um filão de pesquisa, pois, se tais sentenças não são encontradas em um *corpus*, imediatamente surge uma discussão sobre a freqüência

delas (p.12). McEnery e Wilson (p.18) valem-se das palavras de Fillmore, que declara que sempre que teve oportunidade de examinar um *corpus*, mesmo pequeno, encontrou fatos que ele não poderia imaginar descobrir de nenhuma outra forma.

Menos distante, lingüísticamente, do Brasil, no citado trabalho de Said Ali, o autor confessa:

Sem a menor preocupação de descobrir novidades ou tratar questões lingüísticas melhor do que outros o haviam feito, não podia contudo deixar de ir diretamente às fontes buscar a solução dos problemas, porque a isto me obrigava a natureza do trabalho. Averigüei que certas teses sabidas em parte se confirmavam, em parte porém se tornavam insustentáveis. Além disso, o estudo comparado do ponto de vista evolutivo veio revelando, com grande surpresa minha, fatos lingüísticos cuja existência a princípio nem suspeitava (1964, p.7).

Assim, demonstra-se que o uso de *corpus* em estudos lingüísticos antecede o surgimento do computador, o que ratifica a validade de pesquisas em que se utilizou massa de dados. Apesar disso, é inegável que o recente aumento de estudos dessa natureza foi impulsionado pela tecnologia.

1.1.2 Características

Um *corpus* deve ser composto de dados coletados em situações reais de uso da língua para comunicação, de modo que não pode ter havido qualquer finalidade de pesquisa lingüística quando eles foram externalizados pelos usuários da língua. Outros quesitos essenciais de um *corpus* são: que seja eletrônico, permita aplicação dos recursos computacionais de armazenamento, busca, seleção e impressão de informações e esteja em formato que o possibilite ser lido pelos programas mais utilizados atualmente.

Como lembra Sinclair (1991, p.17), uma das motivações de um *corpus* é identificar o que é central e típico na língua. Para tal, esse autor aponta algumas características que devem ser então consideradas:

- é ideal que o conteúdo provenha de várias fontes e áreas e que seja de tipos variados

- para que não se privilegiem certos gêneros de dados em prejuízo de outros – portanto, o material deve ir do formal ao informal, do literário ao ordinário;
- os documentos devem ser inteiramente inseridos no *corpus* – assim, o contexto estará preservado e disponível para consulta, caso haja necessidade, e inúmeras pesquisas lingüísticas poderão ser feitas;
 - para estudar a língua encontrada na sociedade, não é necessário que todos os dados do *corpus* sejam atuais, pois um texto escrito permanece atual por um certo período e influencia a língua;
 - para observar o comportamento de palavras em textos, é necessário um grande volume de ocorrências, no mínimo alguns milhões de palavras – aliás, isso é necessário para que se possam aplicar análises estatísticas aos dados.

É possível complementar essas exigências em Sardinha (2000, p.338-339), que sustenta que um *corpus* que vise a cobrir genericamente a língua:

- deve ser composto de textos autênticos, naturais; e isso implica que
- deve conter textos criados por falantes nativos; e
- deve ser representativo da língua em estudo.

Vale destacar que, embora os dados de um *corpus* sejam naturais, a seleção dos textos que o compõem não o é, como se pode deduzir a partir dos quesitos acima; ao contrário, pois a finalidade de formá-lo é a pesquisa lingüística. Portanto, os critérios de escolha dos textos interferem nos resultados numéricos de análises com base em massa lingüística, ampliando sua abrangência ou reduzindo-a.

1.1.2.1 Representatividade

Já se sabe que *corpus* é uma extensa coletânea de dados lingüísticos, criteriosamente reunidos, teoricamente representativos de uma língua. Porém, há de se esclarecer o que concede a um *corpus* o caráter ‘representativo’.

De acordo com Sardinha (p.342-343), que utiliza explicações de Halliday, Sinclair e Leech, a representatividade está relacionada à probabilidade e à extensão do *corpus*. Sardinha esclarece que a linguagem é um sistema probabilístico, no qual certos traços são mais freqüentes do que outros, e que um *corpus* é uma amostra de uma população cujo tamanho não se conhece, – sem exagero, pode-se complementar – cujo tamanho é infinito, dado que essa população aumenta a cada instante. Ainda, Sardinha chama a atenção para a questão dos significados de um vocábulo, que podem não estar todos representados em um *corpus*. Também, em caso de *corpus* que vise a representar genericamente a língua, o autor levanta a questão do tipos textuais ou gêneros textuais, e frisa que “um número maior de textos de vários tipos permite uma maior abrangência do espectro genérico da língua” (p.345). Dessa forma, em havendo palavras com maior e menor freqüência, o volume de dados precisa ser realmente grande e variado, para que haja maior chance de ocorrer palavras e estruturas raras ou específicas de certos gêneros textuais, por exemplo.

Há de se destacar que a questão da representatividade não é simples. Rocha (2000) ressalta que a definição precisa de uma composição que garanta a representatividade de um *corpus* é seguramente uma ilusão, e que é fácil contestar a representatividade de um *corpus*, ao passo que é difícil assegurá-la. Porém, o autor fundamenta seu trabalho na crença de que essa dificuldade não impede o esforço de obter uma aproximação tão boa quanto possível de uma concepção representativa.

Assim, ao selecionar um *corpus* para trabalho, o pesquisador deve verificar se ele foi montado por uma entidade capacitada no ramo e se o seu tamanho é expressivo, uma vez que o critério da representatividade pode ser um obstáculo para a validade de pesquisas com base em massas de ocorrências naturais de uso da língua. Assim, há escalas de representatividade, de modo que um *corpus* pode ser mais ou menos representativo de uma língua e esse aspecto interfere no alcance dos resultados obtidos.

1.1.2.2 Considerações sobre o conteúdo

A maior credibilidade de um *corpus* é devida ao seu conteúdo refletir unicamente evidências lingüísticas. Assim, pesquisar a partir de exemplos inventados implica utilizar dados adaptados.

Por mais que se responda ‘sim’ a uma pergunta do tipo ‘Essa frase é possível (ou boa) na língua portuguesa?’, o exemplo continuará não sendo genuíno. Uma frase pode ser considerada boa e possível, mas ser irreal, pois os falantes podem preferir sempre outros recursos lingüísticos para expressar as idéias que ela encerra, de modo que tal exemplo possa raramente, ou até mesmo nunca, ocorrer. Essa é, indubitavelmente, uma argumentação plausível no que diz respeito a esse procedimento de estudo da língua. Em contrapartida, exemplos reais são inquestionáveis quanto a uso e ocorrência.

Sinclair (1991, p.5) afirma que proceder de tal maneira, ou seja, empregar frases inventadas em pesquisas, é trabalhar com simulações, e que os resultados podem ser extremamente improváveis. Para ilustrar seu ponto de vista, ele refere que não se pode dar conta da botânica, criando e estudando flores artificiais. Também vale chamar a atenção para a vagueza que norteia perguntas desse tipo. Uma frase pode ser ‘boa’ semanticamente, mas não sintaticamente, por exemplo; de modo que é compreensível, mas não seria escrita nem falada. Além disso, as frases objeto dessas perguntas são isoladas, isto é, não existe um contexto de uso, e isso implica erros e ambigüidades na interpretação de tais frases. Nesse sentido, Sinclair diz que um exemplo somente será completo se for um texto inteiro. Entretanto, não raro, mesmo a presença do contexto não exclui totalmente mal-entendidos e interpretações erradas; assim, a inexistência dele agrava bastante a situação. Em conseqüência, os resultados não são comprovadamente válidos.

Em virtude disso, quando se trabalha com textos escritos, como neste caso, é importante levar em conta que

[...] fazem parte do contexto não somente os elementos do texto, mas também características extratextuais como o gênero textual, o tipo de veículo em que o discurso foi publicado (jornal, revista, anais, [livro etc.]), o local e a época da divulgação; quanto a essa última, especialmente no que tange aos seus aspectos culturais (DUARTE, 2002).

Dessa forma, quanto mais informações sobre um texto puderem ser preservadas, mais fácil será compreendê-lo e mais conveniente utilizá-lo em pesquisas. Tomar esses cuidados aumenta a confiabilidade dos estudos baseados em *corpus*.

1.2 CORPUS DE TRABALHO

Corpus em língua portuguesa disponível para pesquisas é raro, por enquanto, mas já há material organizado dessa forma, o que possibilitou este estudo. O *corpus* de trabalho foi preparado pelo NILC¹ (Núcleo Interinstitucional de Linguística Computacional) e compreende, aproximadamente, 35 milhões de palavras, distribuídas em textos em prosa, escritos em português do Brasil, de variados conteúdos e tipos, coletados nos anos 90 – até 1996. O *corpus* é eletrônico e cada texto é um arquivo em formato TXT, que é específico para textos lineares e não aceita cores, figuras, tabelas e formatações mais elaboradas. Esse é o formato universal de arquivos texto, legível por qualquer editor eletrônico de textos, em qualquer sistema operacional.

O *corpus* foi organizado de acordo com o tipo de ferramenta computacional a ser desenvolvida a partir dele: um revisor gramatical. Os textos são divididos em corrigidos, não corrigidos e semicorrigidos. Essa classificação teve como critério a maior ou menor aproximação ao português escrito padrão, ou seja, o português estabelecido pelas Gramáticas Normativas (NILC, 2000).

Os textos não corrigidos totalizam 738 mil palavras e foram escritos por pessoas de nível médio de escolaridade e universitários. Essa parte do *corpus* é composta por 2.400 textos que incluem redações, monografias e textos de publicidade, por exemplo. Já os textos semicorrigidos são 238, com aproximadamente 1.150.000

¹ Esse *corpus* é citado por Sardinha (2000, p.331) como de destaque dentre os de língua portuguesa criados para pesquisas lingüísticas.

palavras, publicados para um pequeno número de leitores, ou não publicados; portanto, não passaram por correção rigorosa. Fazem parte deles contratos, relatórios, dissertações acadêmicas etc. Finalmente, a parte tocante aos textos corrigidos é composta por cerca de 4.300 textos, que totalizam 32.590.000 palavras e foram publicados para um grande número de leitores – livros, jornais, revistas etc. –, de modo que foram previamente avaliados e considerados propícios para publicação.

A seleção dessa parte diversifica-se da seguinte forma:

- 1) Constituição da República Federativa do Brasil;
- 2) Livros didáticos de primeiro e segundo graus, das matérias ciências (primeiro grau), biologia, química e física (segundo grau), história e geografia (primeiro e segundo graus);
- 3) Cartas comerciais;
- 4) Textos de enciclopédias;
- 5) Artigos de jornais e revistas sobre esporte;
- 6) Artigos de jornal sobre futebol;
- 7) Textos de vários cadernos do Jornal do Brasil e da Folha de S. Paulo, de várias datas; textos traduzidos para o português, do New York Times; e textos do Jornal da USP;
- 8) Textos jurídicos: Código de Processo Civil, Código de Processo Penal, decretos, leis e portarias;
- 9) Textos de literatura infantil e literatura brasileira;
- 10) Livros sobre arte, um manual de geologia e uma tese de doutorado sobre Albert Camus e Ernesto Sábato;
- 11) Partes de livros editados pela Edusp;
- 12) Vários artigos das revistas Isto é, Globo Rural e Veja.

Essa é a parte do *corpus* utilizada nesta pesquisa; assim, trabalhou-se com a variação culta e escrita da língua portuguesa do Brasil.

Dessa forma, todos os textos do *corpus* apresentam-se sem formatação,

gráficos, tabelas, ilustrações e fórmulas. Cada texto, integral ou parcial, constitui um documento que faz sentido ao leitor, ou seja, aborda temas e apresenta informações sobre eles; não sendo, portanto, apenas conjuntos de frases. Merece destaque o fato de os textos jornalísticos aparecerem em grande quantidade no *corpus*. Sobre isso, é relevante expor as palavras de Rocha (2001):

Não obstante, parece seguro afirmar que a linguagem dos meios de comunicação, por seu caráter de massa e conseqüente impacto sobre comunidades lingüísticas inteiras, deve figurar em proporção significativa em um *corpus* que busca ser representativo de uma determinada língua. [...]

Vale lembrar que a análise concentrou-se em contextos imediatos, ou contextos, mas foi possível recorrer aos textos de onde eles provêm, em caso de dúvida sobre a classificação ou significado de uma ocorrência. Ainda, puderam-se obter dados como o tipo de texto, o nome do jornal em que foram publicados, ano, mês e a seção (dos textos jornalísticos); série escolar a que se destinam e a matéria (textos didáticos), autores etc. Assim, preservou-se bastante contexto, o que, como dito, é importante para validar os resultados da pesquisa. Na transcrição de exemplos do *corpus*, alterações textuais mínimas foram feitas, no intuito de clarificar a compreensão, mas sem modificar o contexto dos excertos.

Por ser de língua escrita, o *corpus* utilizado revela uma vantagem a mais, pois isso contribui para a validade da pesquisa, uma vez que a intenção do usuário está presente nos dados. Essa circunstância não ocorre com *corpus* de língua falada, em que se pode perder a intenção do falante, por fatos como a entonação e a gesticulação, por exemplo, desaparecerem na transcrição.

Neste ponto, vale resgatar a explanação anterior sobre a representatividade de um *corpus*. Pouco mais de 32 milhões de palavras não conferem caráter representativo por si a uma massa lingüística, mesmo porque, com os recursos tecnológicos atuais, esse número poderia ser muito maior. Contudo, devem-se levar em conta alguns pontos. Como esclarecido, não há muito material organizado dessa forma disponível para pesquisas, o que dificulta a iniciativa de pesquisadores que querem trabalhar com ocorrências reais da língua escrita. Há material jornalístico, por exemplo, reunido em

larga escala, mas apenas tamanho não é suficiente; é necessária diversificação. A partir disso, argumenta-se aqui, o *corpus* do NILC, embora não seja muito grande, é suficientemente amplo e diversificado o bastante, além de ser organizado por uma entidade competente na área, para que, se não puder ser considerado representativo, pelo menos, pode ser considerado adequado para estudos lingüísticos e, de qualquer forma, é uma amostra da língua portuguesa escrita destes tempos.

1.2.1 Impropriedades

Ao se fazer um teste prévio com a massa lingüística, observou-se que poderia haver algumas ocorrências repetidas na seleção dos dados. Isso aconteceu nos textos jornalísticos e por haver coincidência entre os títulos ou as chamadas das reportagens e alguma sentença no corpo do texto; em casos mais raros, por tratar-se de tabelas em forma de texto, que repetiam sentenças inteiras, alterando somente um nome (de uma rua, por exemplo); e, em situações mais incomuns ainda, porque o texto estava duplicado.

Vale registrar que a maior quantidade de duplicações ocorrida na amostra foi quatro, e que o fato de haver algumas repetições no *corpus* do NILC não comprometeu os procedimentos estatísticos, pois, matematicamente, essa situação pode ser considerada excepcional, dado o extenso volume de material constante no *corpus*.

Também perceberam-se alguns problemas provavelmente causados durante a transcrição, como em:

- 01 a) Diferentemente do futebol gringo, do beisebol [...] nos quais a atuação do técnico é ativa durante **todo a** partida, no futebol, os jogadores têm uma autonomia maior dentro do campo.
- b) **Todo os** dias, o consumidor norte-americano faz 400 milhões de transações.

Como não se conhecia ao certo o que originou esse tipo de impropriedade, as ocorrências que a apresentaram não foram corrigidas e reaproveitadas, mas sim propriamente substituídas por outras. Essa falha não prejudicou o estudo.

Também chamou a atenção a ausência de artigo na seguinte ocorrência:

- 02 a) Ficar sócio de uma locadora de CDs para ter à disposição um acervo de até 9.000 títulos, com **todos gêneros** musicais – inclusive as novidades – e uma taxa de locação de 5% do valor do CD pode durar só 120 dias.

Porém, decidiu-se analisá-la, tendo em vista que ela não foi a única do tipo encontrada – sem que isso descartasse a hipótese de erro de digitação, dentre outras.

1.2.2 *Corpus* de Apoio

Com o objetivo de ilustrar um pouco a trajetória de uso do item lexical em questão e compará-la com os dados do *corpus* de trabalho, bem como expandir a abrangência do modelo de funcionamento a ser desenvolvido para os itens lexicais em estudo, utilizaram-se duas seleções de textos referentes a épocas distintas.

- *Corpus* Histórico

O código *Ofícios do Conselho Supremo Militar ao Governo da Capitania/Presidente da Província* faz parte do acervo do Arquivo Público de Santa Catarina e constitui-se de 144 documentos, distribuídos em 146 folhas, originais dos anos de 1821 a 1827. Essa documentação compõe-se de ofícios, decretos e alguns anexos (tabelas, planos e desenho de medalhas), sendo poucos deles impressos – na “Impressão Regia” ou na “Imprensa Nacional” – e a grande maioria manuscrita. Os documentos correspondem ao Diário Oficial da época.

Muitos deles tratam de pedidos de aumento de salário dos militares, de elevação/extinção de patentes, criação de novos postos na Armada Naval e nos exércitos, concessão e regulamentação de distintivos, julgamentos de réus militares, perdão a desertores etc. Em 2000, o código foi integralmente transcrito para meio eletrônico por Gréggio, Merkle e Duarte, o que possibilitou sua utilização como fonte diacrônica de dados nesta pesquisa. Assim, a partir do arquivo eletrônico desse material, rastream-se e analisaram-se as ocorrências dos itens em questão constantes

nos documentos que o compõem. Os dados foram rodados no computador e geraram-se listagens variadas.

- *Corpus* Sincrônico

Utilizou-se a rede mundial como *corpus*, quando necessário, para aplicar testes estatísticos que confrontassem duas massas lingüísticas distintas. Fizeram-se pesquisas a partir do buscador AltaVista®, nas quais se condicionou que a busca fosse feita em servidores do Brasil, apenas. Esse procedimento engloba páginas eletrônicas de todos os tipos, diversificando assuntos e áreas, e tem-se, portanto, uma amostra aleatória da língua em seu pleno uso na forma escrita, porém, com erros eventuais de digitação e outras impropriedades. Contudo, deve-se relevar que os textos eletrônicos foram considerados adequados, por seus autores, para ser publicados. À medida que os dados chegavam pela rede, algumas páginas eram verificadas para conferir se os itens estavam de acordo com o esperado, e desejado, em termos de significado. Essa conferência foi essencial para que se pudesse utilizar a rede. Em um dos testes aplicados, em se esperando casos de uso de ‘todos’ como em *Ele usará todos seus recursos*, verificou-se se as ocorrências buscadas pela rede não representavam o uso como em *Os discos eram todos seus*, pois o que se queria era investigar se a presença de artigo sucedendo ‘todos’ é obrigatória.

A explanação feita neste capítulo permite observar que, com a possibilidade de reunião de uma consistente e volumosa massa lingüística, podem-se conhecer efetivamente e explorar muitas das facetas da língua. Com a utilização de *corpus*, o estudo da língua tornou-se mais prático, voltado para as expressões que realmente chegam aos falantes e por eles são proferidas, em contextos de vida. Além disso, constata-se que um *corpus* planejado propriamente é uma reveladora fonte de informações sobre a língua falada e escrita; e que o *corpus* de trabalho é adequado para a pesquisa – mesmo tendo em vista a polêmica que envolve a questão da representatividade – dadas as suas características e as referências feitas a ele por autores renomados na área.

2 LINGÜÍSTICA COMPUTACIONAL, LINGÜÍSTICA DE *CORPUS* E ESTATÍSTICA: O MÉTODO

Como exposto no capítulo anterior, primeiro utilizou-se massa lingüística, muito tempo depois é que se começou a desenvolver a forma adequada de tratá-la: a lingüística de *corpus*, que envolve a lingüística computacional (manuseio virtual de dados) e a estatística, uma vez que sempre se trabalhará com uma amostra da língua. O método utilizado é fundamental para os resultados deste trabalho, de forma que é necessário, a partir de considerações de autores pertinentes, esclarecê-lo propriamente e justificá-lo; e ainda detalhar a pesquisa desenvolvida, em virtude da teoria que a norteia.

2.1 ABORDAGEM UTILIZADA

Há autores que consideram a lingüística de *corpus* uma abordagem, como Sardinha (2000), e outros uma metodologia, como Rocha (2001). Leech (1992, p.106) refere-se a ela como não apenas uma nova metodologia, mas também uma nova abordagem filosófica. De uma ou outra forma, a base metodológica que possui a lingüística de *corpus* é incontestável.

Assim, esta pesquisa insere-se no tipo de abordagem de fenômenos lingüísticos que recebe o nome de lingüística de *corpus* – aqui também referida como lingüística de massa. Trata-se de um método caracterizado pela utilização de um *corpus* (massa) lingüístico. A partir da observação dos dados do *corpus*, desenvolve-se uma interpretação do fenômeno da língua estudado. Essa observação é feita com o auxílio indispensável de programas de computador que manuseiam dados textuais e de recursos matemáticos que permitem interpretar dados e extrair significados para um

todo a partir de uma amostra.

A fim de ilustrar a validade de estudos lingüísticos a partir de *corpus*, Manning e Schütze (2000, p.8) apresentam duas indagações básicas a que a lingüística deveria responder:

- a) Que tipo de coisas as pessoas dizem?
- b) O que essas coisas dizem/perguntam sobre o mundo e o que requerem dele?

Os autores consideram que a questão ‘a’ envolve todos os aspectos estruturais da língua, enquanto a questão ‘b’ diz respeito a semântica, pragmática e discurso, ou seja, a como conectar enunciações com o mundo. A questão ‘a’ é a base da lingüística de *corpus*, mas os padrões de uso de uma palavra podem atuar como substitutos para a compreensão profunda e, conseqüentemente, possibilitar que a segunda questão seja abordada a partir de técnicas baseadas em *corpus* (p.8).

2.1.1 Lingüística de *Corpus*

Lingüística de *corpus* é uma metodologia relativamente nova para estudos da língua e pode ser aplicada a muitas áreas lingüísticas – é a sistemática que envolve a pesquisa com base em *corpus*. Ela não é considerada um ramo de estudo da lingüística, porque não explica aspectos da língua, ou seja, não se refere a um campo específico de estudo, mas sim serve de base metodológica para estudos da língua. Por isso, combina-se facilmente a várias áreas de pesquisa dessa natureza, sendo possível estudar sintaxe, semântica, fonética e sociolingüística, dentre outras, a partir da lingüística de *corpus*. Trata-se de uma metodologia que é ainda pouco utilizada no Brasil, apesar de já não ser novidade em vários outros países.

Embora estudos lingüísticos envolvendo *corpus* datem de 1897, conforme mencionado no capítulo 1, e alguns lingüistas, inclusive estruturalistas, tenham utilizado um método com base em *corpus* em pesquisas, nos anos 1940 (McENERY e WILSON, 1997, p.2), o termo lingüística de *corpus* é novo. De acordo com Leech (1992, p.105), ele apareceu pela primeira vez em 1984 como título de um livro de

Aarts e Meijs. Entretanto, em épocas anteriores, trabalhar com um grande volume de dados era impraticável por causa das dificuldades de manipulação, que exigia muito tempo, pessoal e dinheiro. Dessa forma, desacreditados, métodos com base em *corpus* não foram muito utilizados por algumas décadas. Devido a essa inviabilidade, Abercrombie, em 1965, chamou o uso de *corpus* em pesquisas lingüísticas de pseudotécnica.

Porém, a tecnologia tornou esse argumento obsoleto. O advento do computador veio a resolver os problemas de gerenciamento e estocagem de dados, viabilizando a lingüística de massa e tornando-a instrumento aplicável a quase todos os campos de pesquisa em língua. Assim, intimamente associada à tecnologia, Leech (1992, p.106) vê a lingüística de *corpus* como um meio essencial para um novo tipo de conhecimento e como um portal para um novo modo de pensar sobre a língua.

Além de acelerar o manuseio de enormes quantidades de dados (busca, armazenamento, classificação e disposição, dentre outros), o computador pôs a matemática a favor da lingüística, de forma que não somente cálculos complexos importantes em pesquisas puderam ser realizados em milissegundos, mas também cálculos simples, cuja aplicação a enormes quantidades de dados era, antes, trabalhosa e suscetível a erros. Em consequência, a estatística passou a ser aliada da lingüística de *corpus* e concedeu mais cientificidade a essa metodologia. A partir daí, modelos matemáticos de comportamento lingüístico puderam ser desenvolvidos (p.106).

Apesar de o volume de dados não ser o principal aspecto referente à lingüística de *corpus* (mas sim o estudo da língua por meio de textos autênticos), ele é a razão imediata mais compreensível para dizer que técnicas com base em *corpus* vêm crescendo e se tornando cada vez mais significativas desde os anos 50 (p.107). O foco desse tipo de abordagem está na descrição do comportamento lingüístico, sob a forma do discurso natural escrito ou falado – isso é o que um *corpus* deve conter (p.107). Logo, é o desempenho lingüístico que é estudado, e não, pelo menos diretamente, a competência.

Todavia, Leech argumenta que

[...] performance itself is not a unitary phenomenon, but contains its own dichotomy. We can study performance either as *process* or as *product*. When we study performance as process, we take account of the psychological process involved in the production of the text (spoken or written) and in its interpretation. This means that performance qua process is in part physical but in part psychological. But CCL [computer *corpus* linguistics] studies performance as product: a *corpus* consists of spoken or written texts in themselves, the physical manifestations of language, independent (in principle) of the mental processes of their addressers and addressees (p.108).

O autor explica que considera que essa independência seja *a princípio* porque não há como garantir uma interpretação totalmente objetiva de um texto em um *corpus* eletrônico, devido ao processo de transferência dos textos para o computador implicar certas decisões e, sobretudo no caso de textos de língua falada, interpretações subjetivas. Mesmo assim, os dados transferidos são menos abstratos e podem ser diretamente observados.

Desse modo, a metodologia releva a experiência, de forma que a observação contribui para a teoria mais do que a teoria para a observação. Portanto, os dados de um *corpus* independem dos preceitos das teorias que serão utilizadas na pesquisa, embora elas determinem a classificação e interpretação desses dados (p.111). Nesta pesquisa, para a classificação dos dados, utiliza-se a nomenclatura da gramática tradicional, conforme exposto na subseção 2.4.

Ao comentar o paradigma de pesquisa empírica da lingüística de *corpus*, Leech (p.111-112) atenta para alguns aspectos que devem ser levados em conta quando se trabalha com esse tipo de abordagem:

- por maior que seja um *corpus*, ele é apenas uma amostra da língua em uso;
- os dados a analisar não devem ser escolhidos de acordo com as preferências do pesquisador, e sim aleatoriamente, e nenhum deles pode ser considerado irrelevante para a pesquisa;
- teorias ou modelos podem ser criados para explicar os dados encontrados (a partir da intuição ou experiência do investigador, por exemplo), mas os valores quantitativos do modelo devem ser obtidos dos dados do *corpus*;
- a precisão do modelo pode ser testada em outro *corpus*;

- a princípio, a qualidade de um modelo pode ser medida e comparada com a de outros modelos (essa interação é importante para que os modelos de desempenho lingüístico sejam progressivamente aperfeiçoados) e diferentes modelos podem ser testados com o mesmo *corpus*, de forma que a superioridade de um modelo em relação a outro possa ser demonstrada.

Leech (1992) defende que, do ponto de vista científico, lingüística de *corpus* é uma poderosa metodologia, por possibilitar a verificação objetiva de resultados. Assim, não se trabalha com o que é teoricamente possível, e sim com o que é realmente possível.

2.2 LINGÜÍSTICA COMPUTACIONAL

A lingüística computacional é bem mais antiga do que a lingüística de *corpus* e insere-se no âmbito da inteligência artificial. Portanto, configura uma complexa discussão técnico-científica, que não cabe ser pormenorizada neste estudo, no qual a lingüística computacional é entendida, de modo restrito, como o desenvolvimento e a aplicação de recursos eletrônicos para manusear propriamente dados lingüísticos. Há programas informatizados específicos para tal, embora não sejam muitos os de fácil acesso aos pesquisadores, porque o idioma em que trabalham é estrangeiro, por não estarem disponíveis nos ambientes de pesquisa e por não ser tão simples utilizá-los.

Esses programas enxergam cada palavra dentro de um arquivo de texto e, em certa medida, permitem que o pesquisador refine as buscas por palavras ou expressões, de forma que é possível fazer uma varredura rápida em muitos milhões de palavras, procurando expressões e itens lexicais específicos, sem que seja necessário ler cada texto, cada linha de texto, cada palavra, à cata desses dados.

Mesmo assim, não raro o pesquisador precisa fazer uma seleção que o programa não permite – isso não é problema exclusivo das ferramentas eletrônicas para estudos da língua, ocorre com todos os programas de computador, independentemente da sua aplicação. Assim, a fim de suprir as necessidades desta pesquisa, foram aproveitados recursos de programas não específicos para estudos

lingüísticos, pois, rodando em um sistema operacional multitarefa, os programas comunicam-se, de forma que é possível transportar dados e lê-los em vários formatos.

Os programas em ambiente Windows® utilizados para tratar os dados da análise foram o WordSmith® – coocorrência, seleções preliminares dos itens de trabalho, seleções específicas posteriores e alguns dados numéricos; a planilha Excel® – cálculos, gráficos e leitura das listagens; e o próprio editor de textos Word® – buscas por item e destaque de itens. O intercâmbio de dados possível entre esses programas contribuiu para a eficiência e agilidade das leituras. O WordSmith® pode ser considerado um exemplo, deveras muito simples e limitado, de aplicação da lingüística computacional. Contudo, nem mesmo ele é comum em pesquisas acadêmicas, pois seu funcionamento ainda parece obscuro para muitos pesquisadores, além de estar em língua inglesa.

Não se utilizou anotação automática (etiquetagem), porque se optou por trabalhar com amostra, cujos dados foram analisados um a um pelo pesquisador. Uma espécie de anotação foi feita em células adjacentes às dos co-textos, na planilha eletrônica. A partir disso, foi possível aplicar filtros e cálculos diversos – ou seja, fazer ensaios – diretamente nas células contendo as ‘anotações’, bem como gerar e testar gráficos, o que facilitou a preparação da análise.

	A	B	C	D	E	F	G
1	N	Dados	SBJ	ADN			reforço
7	16	claramente colocada em evidência nas máquinas de					
10	25	A <i>biomédica se passa quase todo</i> num apartamento em	SBJ	ADN	ADV		interesse
13	36	Seja por uma necessidade de expansão, seja porque a	SBJ	ADN			conjunto
15	46	Cinco anos atrás, as pessoas diziam "eu conheço uma	SBJ	ADN			reforço

Figura 1 – Exemplo de aplicação de filtro aos dados e ‘anotação’ em planilha eletrônica

2.3 ANÁLISES ESTATÍSTICAS

Mais do que descrever parte da metodologia, esta seção também traz algumas aplicações, haja vista que vários dos recursos aqui expostos são empregados também preliminarmente à análise.

McEnery e Wilson, Leech, Biber, Manning e Schütze são unânimes em corroborar a afinidade entre a lingüística de *corpus* e a Estatística. Por definição, Estatística é o ramo da matemática que trata da coleta, análise, interpretação e apresentação de grandes quantidades de dados numéricos, geralmente com o propósito de inferir as proporções em um todo, a partir das proporções em uma amostra representativa, ou seja, estimar parâmetros (características descritivas dos elementos da população), por meio de estatísticas (operações com dados da amostra), conforme Barbetta (1998, p.159); e a probabilística é a parte dela que lida com a perspectiva favorável de fatos acontecerem.

Sobre a língua ser um sistema probabilístico, Manning e Schütze (2000, p.15) esclarecem que o mais radical argumento a favor da probabilidade como parte de uma compreensão científica da língua é que a cognição humana é probabilística e então a língua também deve ser, uma vez que é parte da cognição. A justificativa para se considerar que a cognição é um sistema probabilístico fundamenta-se no fato de que as pessoas vivem em um mundo onde há muitas incertezas e informações incompletas; e, para elas interagirem satisfatoriamente com esse mundo, precisam ser capazes de lidar com esse tipo de informação (p.15). Nessa lida, elas devem absorver, compreender, avaliar e completar essas informações com êxito, bem como compará-las, para então tomar uma decisão, que variará de acordo com a pessoa.

Biber (1998, p.4-8) enfatiza que uma das características da abordagem com base em *corpus* é que ela depende de técnicas de análise quantitativas e qualitativas (que podem ser chamadas de funcionais), uma vez que interpretações funcionais dos dados quantitativos são essenciais para explicar as razões por que há certos padrões de ocorrência de uma palavra, por exemplo. Tanto quantidade como qualidade são englobadas pela metodologia desta pesquisa. Análises quantitativas estão intimamente ligadas a estudos com base em *corpus*, pois, havendo bastantes dados, é possível obter

frequências e aplicar testes estatísticos, bem como dispor os resultados em gráficos e fazer várias comparações. Como aponta Leech (1992, p.110), não há como utilizar a intuição do falante nativo em testes probabilísticos.

Vale lembrar que outras ciências há muito tempo utilizam análises quantitativas e a matemática em pesquisas. Por outro lado, dados não quantitativos também podem ser tratados, pois, por exemplo, há como extrair apenas os tipos de ocorrências existentes sem se preocupar com a quantidade de cada uma. Isso é essencial para a criação de modelos de ocorrência de uma palavra e verificação dos significados assumidos por um vocábulo ou por uma expressão.

2.3.1 Frequência

A percepção da língua como sistema probabilístico pressupõe que os traços lingüísticos não ocorrem uniformemente, ou seja, com a mesma frequência. Assim, a frequência de verbos em textos não é a mesma que a de adjetivos, bem como as estruturas possíveis não ocorrem todas com a mesma frequência, por exemplo. Sardinha (2000, p.350-351) chama a atenção para o fato de essas frequências não serem aleatórias e fundamenta essa concepção nos estudos de Biber, que evidenciaram que “conjuntos de traços lingüísticos variam sistematicamente com relação a textos típicos de contextos comunicativos específicos”. Dessa forma, há certos padrões na língua que somente podem ser delineados a partir da observação empírica da frequência do emprego feito pelos usuários, em contextos determinados (p.352).

Já se sabe que o modo mais simples de encontrar combinações ou frases fixas em um *corpus* é mediante contagem. Então, se a taxa de frequência de duas palavras adjacentes (bigrama) é reconhecidamente alta – em comparação à de outras palavras do *corpus* –, tem-se, a princípio, uma evidência de que elas têm uma função especial que não é explicada como simples resultado da combinação das duas palavras (MANNING e SCHÜTZE, 2000, p.153).

Como previsto, esta pesquisa iniciou-se a partir da observação e, à medida que se desenvolvia, as hipóteses passíveis de verificação que surgiam eram

matematicamente confirmadas. Optou-se por utilizar o teste qui-quadrado (χ^2) – um dos mais utilizados em pesquisa social e que permite testar a significância da associação entre duas variáveis qualitativas ou comparar duas amostras (BARBETTA, 1998, p.222-223) – para verificar a dependência entre o item lexical em análise e as palavras que o cercam. Considera-se propício que maior detalhamento sobre esse teste seja fornecido no momento de sua aplicação, para que o seu funcionamento seja melhor compreendido.

No entanto, Manning e Schütze (2000, p.162) referem que alta frequência e baixa variância podem ser apenas acaso; de forma que, se a ocorrência de dois vocábulos é muito frequente, então, espera-se que ocorram juntos muitas vezes sem que haja dependência. Dado isso, o que se investiga, por meio da aplicação desse teste, é se tais palavras ocorrem juntas mais frequentemente do que seria admissível para que fosse apenas por coincidência.

Barbetta (1998, p.14) ressalta que em pesquisas empíricas é fundamental o teste das hipóteses formuladas, pois, uma vez comprovadas estatisticamente, podem servir de suporte a outras pesquisas, encadeando conhecimentos.

Nesta pesquisa, decidiu-se trabalhar, sobretudo, com grupos de três palavras adjacentes (trigramas) – dado que bigramas são mais inconsistentes, mas também foram feitos testes com outros n-gramas, pois são automaticamente gerados pelo WordSmith e isso facilita o trabalho com eles. Manning e Schütze (2000, p.156) alertam para a simplicidade desse procedimento, mas chamam a atenção para a surpreendente meticulosidade dos resultados.

Também se decidiu que se trabalharia somente com os trigramas cuja frequência fosse expressiva no *corpus*, a qual se arbitrou que seria maior do que 100.

2.3.2 População e Amostra

Por população entende-se o conjunto de elementos abordados pelo estudo, para os quais se deseja que as conclusões obtidas sejam válidas; enquanto amostragem é a seleção de parte da população (amostra) para ser observada (BARBETTA, 1998,

p.19-20). O principal cuidado ao se trabalhar com uma amostra é que ela seja representativa ou, pelo menos, adequada, tendo em vista a impossibilidade de garantia de representatividade, discutida anteriormente. Contudo, há como garantir que uma amostra originada a partir do *corpus* de trabalho seja representativa.

Para efeito desta pesquisa, o *corpus* de trabalho – apesar de ser uma amostra da língua, que é infinita – é a população acessível, isto é, observável (p.19), e a amostra é a seleção de um percentual dessa população, feita para que se pudesse observar um a um alguns dados.

Em termos de dados disponíveis no *corpus*, buscas a partir do WordSmith indicaram o seguinte:

Tabela 1 – População: total de ocorrências dos itens pesquisados

Termo	Ocorrências
Todos	19.785
Todo	16.089
Todas	11.218
Toda	9.614
<i>Total</i>	56.706

As listagens originadas a partir de agrupamento de dados nos programas de computador utilizados permitiram visar a população. Porém, é inviável trabalhar com mais de 56 mil dados, de forma que foi utilizada uma amostra proporcional ao número de ocorrências encontrado para cada item, representativa da população em estudo.

Ao defender a viabilidade do uso de amostragem quando se fazem levantamentos de grandes populações, Barbetta (p.39) comenta quatro razões: economia (já que se releva apenas parte da população), tempo (obtêm-se resultados mais rapidamente), confiabilidade dos dados (quando se pesquisa um número reduzido de dados, pode-se dar mais atenção a casos individuais, evitando erro nas respostas) e operacionalidade (por ser mais fácil realizar operações em pequena escala).

Optou-se por pesquisar mais detalhadamente 1.000 ocorrências, porque, apesar de ser um número relativamente grande de dados, a pesquisadora considerou possível observar cada uma, o que seria complicado fazer com um número maior de dados, e, com menos dados, haveria risco de a análise ser superficial. Assim,

considerou-se que 1.000 dados é um tamanho intermediário para uma amostra ser adequada à realização deste estudo. Isso decidido, foi necessário descobrir qual seria o valor percentual proporcional de cada conjunto de itens de trabalho, a fim de obter uma amostra com essa quantia exata de dados. Esse método é chamado amostragem aleatória proporcional estratificada (WALSH, 1990, p.80). Para esse procedimento, utilizou-se a planilha eletrônica Excel®. Cálculos indicaram que 1.000 ocorrências equivaliam a aproximadamente 1,764% da população, o que acarretou os seguintes números de dados por termo de trabalho:

Tabela 2 – Amostra: ocorrências selecionadas para análise

Termo	Itens para trabalho
Todos	349
Todo	284
Todas	198
Toda	170
<i>Total</i>	1.000

Os dados da amostra foram sorteados pelo computador e cada ocorrência teve igual chance de ser escolhida. Para a seleção, utilizou-se um pequeno programa que gerou números aleatórios que podiam variar de 1 ao número total de ocorrências de cada termo pesquisado, sem permitir repetições, desenvolvido especialmente para este estudo.

Como já se sabia que poderia haver repetições e erros de digitação, tendo em vista contornar essa casualidade, para cada item a analisar, foram selecionados cinco números a mais do que o necessário, e por eles foram substituídas as ocorrências repetidas. Em quatro etapas, foram gerados e ordenados 289, 354, 175 e 203 números aleatórios, e as ocorrências correspondentes a eles no *corpus* – numeradas pelo WordSmith – foram transferidas para a planilha eletrônica Excel.

Quando se toma por base uma amostra, os resultados representam uma estimativa, de modo que se deve calcular o erro amostral, em virtude de não se pesquisar toda a população, mas se aplicar os resultados a ela. Neste caso, o cálculo feito apontou que a pesquisa tolera um erro de 2%; porém, para efeito de maior

segurança, considerou-se um erro de 3%. Saber o erro permite conhecer o tamanho mínimo da amostra, para que se possam obter os valores desejados. Para que sejam tolerados tais percentuais de erro, a amostra deve ter os seguintes tamanhos, a partir dos quais se percebe que a amostra da pesquisa é adequada²:

Tabela 3 – Quantidade de dados a ser selecionados para se tolerar erros amostrais de 2% e 3%

Item	Erros amostrais e dados	
	3%	2%
Todos	266	306
Todo	226	255
Todas	168	183
Toda	147	159

Referenciando a asserção de que, quando se trabalha com *corpus*, obtêm-se achados inesperados, nas verificações preliminares no *corpus* de trabalho, observou-se que havia uma falha no projeto inicial quanto às flexões do item lexical ‘todo’, pois a forma diminutiva ‘todinho’ e flexões não haviam sido previstas. Então, a fim de tornar a pesquisa mais completa, procuraram-se esses itens no *corpus* de trabalho para inclusão na análise. Os dados obtidos revelam que as formas diminutivas do item lexical em estudo encontradas no *corpus* correspondem a 0,00023%, ou 23^{-5} , do total de ocorrências desse item. A busca indicou o seguinte:

Tabela 4 – Formas diminutivas do item de trabalho no *corpus*

Termo	Itens de trabalho
Todinho	6
Todinha	5
Todinhos	1
Todinhas	1
<i>Total</i>	13

² Esse cálculo foi feito de acordo com as explicações de Barbetta (1998, p.58) e leva em conta o número de elementos da população, o tamanho da amostra com que se pretende trabalhar e o percentual de erro que se admitirá, sob um nível de confiança de, no mínimo, 95%.

2.4 CATEGORIZAÇÃO DOS DADOS

Neste tipo de estudo, o papel das teorias da língua é auxiliar na categorização e a interpretação dos dados. Nesta pesquisa, para a classificação dos dados, utilizou-se, principalmente, a nomenclatura da gramática tradicional, além de algumas designações consideradas gerais; e trabalhou-se com a morfologia e a sintaxe, sobretudo para investigar a função dos sintagmas de que os itens lexicais estudados participam e a função dos itens nesses sintagmas; e com a semântica, uma vez que se pretendia descobrir as noções de sentido agregadas ao item lexical pesquisado. Considera-se que dominar a teoria escolhida é importante, pois, se a classificação for inadequada – ou seja, se não for sustentada por estudos específicos –, os resultados ficarão comprometidos.

Escolheu-se a teoria tradicional, porque a pesquisadora tem conhecimento adequado sobre ela, inclusive de seus problemas e limitações, e já desenvolveu outros estudos em que a empregou. Nesse sentido, uma vez que categorizar os dados é necessário, utilizar outra teoria poderia dificultar e retardar o trabalho, bem como aumentar o risco de erros. Além disso, em dois dos livros citados neste trabalho – Manning e Schütze (2000) e Sinclair (1991) – os autores utilizam a nomenclatura gramatical ou parte dela.

Os teóricos que serviram de suporte para essa etapa da pesquisa foram o filólogo renomado Said Ali (1953), Almeida (1998), que trabalha com gramática tradicional, e Neves (2000), que utiliza a observação para descrever os usos da língua. Contudo, a sua participação restringiu-se a nomenclatura e base para discussões, mediante confronto com os resultados do padrão de uso a ser elaborado, de forma que, como dito, não haverá preocupação em encaixar os dados nas teorias, mas apenas vislumbrar os limites de conformidade delas frente aos dados pesquisados, mesmo porque também é intenção, neste estudo, contribuir para a veracidade da teoria da língua.

Estudos anteriores em lingüística de *corpus* comprovam que é comum deparar ocorrências que fogem à classificação. Segundo Manning e Schütze (2000, p.15), um argumento freqüente nesses casos é que esse fenômeno é marginal e raro. Nesta

pesquisa, em havendo tais ocorrências, optou-se por reportá-las e discuti-las. Exemplo desse problema são as frases feitas, expressões idiomáticas e locuções adverbiais, de cujo grupo, já se sabe, participam os itens lexicais em análise.

A classificação começou a ser desenvolvida um ano antes da conclusão da pesquisa, para que houvesse oportunidade de esclarecer dúvidas a respeito dela com profissionais que trabalham com a gramática padrão e também com usuários da língua não estudiosos dela. Até o resultado final, a classificação teve de ser revista várias vezes e amadurecida; em várias ocasiões, foi preciso recorrer ao texto integral para avaliar o contexto e retificá-la. Apesar desses cuidados, não se pode assegurar que há isenção total de erros, mas sim que houve esforços no sentido de reduzi-los ao máximo.

A partir da exposição realizada neste capítulo, percebe-se que a tecnologia possibilitou a realização, em projetos viáveis, da metodologia da lingüística de *corpus* e incluiu, definitivamente, a matemática nos estudos da língua, conferindo maior credibilidade a esses estudos, facilitando-os e agilizando-os. Tornou-se possível, então, fazer comparações a partir de grandes volumes de dados e assim descobrir e pesquisar o uso real – enfocando o desempenho – que o falante faz da sua competência lingüística. Desse modo, esta pesquisa usufrui, sobretudo, a tecnologia e a matemática para investigar evidências de uso da língua, a partir de uma massa lingüística.

3 ANÁLISE

Uma vez cumpridos os pré-requisitos necessários, é tempo de iniciar a análise com base em *corpus*. Trata-se de uma investigação que visa a propor questões sobre o item lexical de trabalho, suas flexões e formas diminutivas e buscar respostas para elas, por meio: da decomposição dos sintagmas que contêm as flexões; de verificações qualitativas e quantitativas deles e das combinações semânticas possíveis com os itens a estudar; de cotejos entre estruturas; e de discussão dos fatos encontrados.

3.1 PREDEFINIÇÕES

Primeiramente, considera-se importante para a análise conhecer a origem do vocábulo ‘todo’. Segundo Ferreira (1999, p.1969) ‘todo’ advém do latim *totu*, em que era adjetivo e significava ‘inteiro, completo, total’. Portanto, sabe-se que seu valor fundamental era modificar a compreensão de um substantivo. Said Ali (1964, p.118) explica que o vocábulo ‘todo’, variável em gênero e número, pertence à língua desde os tempos mais remotos, e o invariável ‘tudo’ data da literatura quinhentista e substituiu o antigo ‘todo’ empregado no sentido de ‘toda a coisa’.

Na teoria tradicional, ‘todo’ é considerado pronome indefinido, que é como “se chama a palavra que determina o substantivo de modo vago, de maneira imprecisa”, conforme Almeida (1999, p.192); já Neves (2000, p.533) postula que vocábulos dessa natureza são, a princípio, “palavras **não-fóricas**, isto é, não constituem itens com função de instruir a busca de recuperação semântica na situação ou no texto”, o traço que as une é a indefinição semântica; e Said Ali (1953, p.66) diz que “Dá-se o nome de PRONOMES INDEFINIDOS a uma série de nomes aplicáveis à 3.^a pessoa do discurso quando ela tem sentido vago e indeterminado” – salienta esse

autor que alguns pronomes indefinidos podem ser adjuntos.

Sobre outros aspectos, predefine-se que: doravante, utilizar-se-á a expressão ‘tod*’ quando se quiser indicar o item lexical em estudo e suas flexões (‘todo’, ‘todos’, ‘toda’, ‘todas’); as formas diminutivas do item lexical serão assim designadas; os dados numéricos das tabelas apresentadas referem-se à amostra, salvo quando for indicado o contrário; são feitas investigações tanto a partir da listagem completa dos dados como da amostra, sendo sempre informados os dados que as determinaram; quando for indicado *corpus*, entenda-se que se trata da população; também, a classificação semântica denominada ‘reforço’ foi elaborada a partir da percepção da pesquisadora, tendo em vista a análise dos dados, e não adotada de um autor; as demais foram obtidas em Almeida (1999) e Said Ali (1964).

Além disso, fica subentendido que, embora nem sempre seja transcrito o adjetivo pertencente a uma estrutura, há possibilidade de haver um adjetivo anterior ou posterior a qualquer substantivo, sendo que omiti-lo não desqualifica o sintagma que o contém; ainda, é considerado que após qualquer artigo pode vir um substantivo, valendo a mesma postura, e que é facultativo o artigo antes de pronome possessivo.

Dado que se trabalha com morfossintaxe, os sintagmas – aqui entendidos como partes das sentença, ou seja, unidades formadas por uma ou mais palavras que, juntas, desempenham uma função na frase (NILC, 2003 e UFRJ, 2003) – recebem atenção especial, sobretudo, no tocante à sua função na sentença, e não ao seu tipo; já os adjuntos são utilizados, respeitando-se, a princípio, o que prediz a gramática tradicional, isto é, são modificadores de substantivos, verbos, advérbios e adjetivos (ALMEIDA, 1999, p.430-431). Portanto, consideram-se os componentes da sentença para fins de classificação, para a qual se adota a seguinte nomenclatura – quando não se transcrevem integralmente as designações utilizadas –, a partir das definições da gramática tradicional:

Quadro 1 – Nomenclatura utilizada na categorização dos dados

Código	Significado
adj	adjetivo ou locução adjetiva
adv	advérbio
art	artigo
conj	conjunção
dem	pronome demonstrativo
indef	pronome indefinido
part	particípio
poss	pronome possessivo
prel	pronome relativo
prep	preposição
pro	pronome reto
subst	substantivo
verb	verbo
verbau	verbo auxiliar
verbi	verbo intransitivo
verblig	verbo de ligação
verbtđ	verbo transitivo direto
verbti	verbo transitivo indireto

3.2 PERFIL GENÉRICO DOS ITENS LEXICAIS

A partir de uma listagem do WordSmith por frequência, pôde-se observar a situação de ‘todo’ e suas flexões, em relação às palavras do *corpus*, e fazer algumas comparações, bem como comentários. Já se sabe que o total de ‘tod*’ e suas formas diminutivas no *corpus* é 56.719. No que diz respeito às palavras que compõem os textos da massa lingüística, esse expressivo valor corresponde aos seguintes percentuais de participação:

Tabela 5 – Percentual e ordem de ocorrência dos itens em estudo no *corpus*

Termo	%	Ordem
Todos	0,07	101
Todo	0,05	184
Todas	0,03	259
Toda	0,03	317
Formas diminutivas	<0,01	>88.000

Assim, o item lexical e suas flexões correspondem a aproximadamente 0,18% do total de palavras do *corpus*. Com base na listagem, constatou-se que: ‘todos’ ocorre na mesma proporção que palavras como ‘qualquer’ e ‘tempo’ e em menor proporção do que ‘contra’ e ‘mesmo’; ‘todo’ é mais freqüente do que ‘alguns’ e ‘quase’, por exemplo; há mais ‘todas’ do que ‘mulher’ e ‘enquanto’; e ‘toda’ aparece mais do que ‘nós’ e ‘muitas’.

No *corpus*, observa-se que, em comparação com a palavra mais comum em textos de língua portuguesa, ou seja, a preposição ‘de’, na freqüência, ‘tod*’ distancia-se dela mais de um milhão de ocorrências (aliás, as aglutinações *de+art* também são mais freqüentes do que ‘todos’). Sozinha, ‘de’ corresponde a 4,65% das palavras do *corpus*. Mesmo assim, a posição de ‘tod*’, sobretudo a de ‘todos’, alcança bastante destaque, pois as listagens geradas pelo WordSmith revelam que as formas masculinas estão em meio às 260 palavras mais freqüentes da língua portuguesa escrita e as formas femininas estão em meio às 320. Porém, isso não vale para as formas diminutivas, dado que palavras pouco conhecidas e raramente empregadas, como ‘nióbio’, podem ser mais freqüentes do que elas. Por meio de um cotejo com outro *corpus*, poder-se-ia refinar essas constatações, bem como referendá-las.

3.2.1 Freqüência Geral e Modelo Probabilístico

Um importante e primordial aspecto a ser verificado é a freqüência de ‘tod*’. Na coocorrência desses itens lexicais, observou-se que, em número, o plural supera o singular em 5% e, em gênero, o masculino supera o feminino em 26%. Notável também é a freqüência de ‘todos’, que se distancia 8.567 (15%) ocorrências de ‘todas’ e 3.696 (7%) ocorrências de ‘todo’; e de ‘todo’, que se distancia 6.475 (11%) ocorrências de ‘toda’. O gráfico seguinte mostra os percentuais de freqüência obtidos:

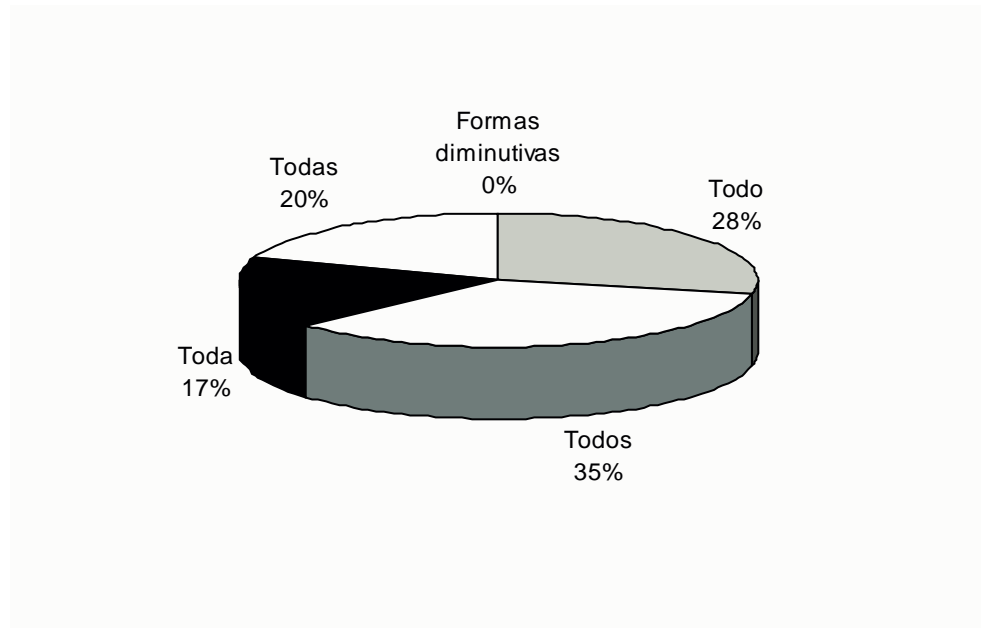


Gráfico 1 – Distribuição em % dos itens de trabalho

Os números do gráfico representam a probabilidade de encontrar esses itens lexicais em um texto. Em virtude disso, estima-se que, quando um leitor deparar algum desses vocábulos em um texto semelhante aos do *corpus*, há 35% de chance de ele ser ‘todos’, por exemplo, 55% de chance de ele ser plural e praticamente nenhuma chance de ser qualquer uma das formas diminutivas, dado que elas aparecem em quantidade muito inferior a 1% no *corpus*. Percebe-se também que o gênero masculino representa 63% das ocorrências; que o feminino plural supera o singular em 3%; e o masculino plural supera o singular em 7%. Disso decorre que, em termos de número, o uso é mais ou menos equilibrado, mas não o é em termos de gênero.

Dessa forma, o modelo probabilístico obtido a partir dos dados é o seguinte:

Tabela 6 – Modelo probabilístico de ‘tod*’+formas diminutivas

Resultado	Probab.
Todos	0,35
Todo	0,28
Todas	0,20
Toda	0,17
Formas diminutivas	0,00

Assim, põem-se algumas questões pertinentes para a pesquisa: Há distinção entre o uso do item no singular e seu uso no plural? E entre seu uso no masculino e no feminino? Em caso afirmativo, outra pergunta assume posição: Como essa distinção é feita? E, ainda, cabem três outras questões: Por que o masculino plural é bem mais freqüente do que o feminino plural?; Por que ‘todo’ é bem mais freqüente do que ‘toda’?; Por que as formas diminutivas são tão escassas no *corpus*?

As informações numéricas relatadas nesta subseção revelam que o fenômeno de trabalho tem um alto potencial para suscitar questões lingüísticas e que as causas dos percentuais correspondentes merecem ser investigadas, de forma que se espera que, por meio da análise do uso de ‘todo’ e de cada uma das suas flexões, seja respondido às perguntas postas.

3.3 TODA

3.3.1 Estrutura, Função Sintática e Combinação Semântica

No exame qualitativo³ de uma listagem do WordSmith por agrupamento da população do *corpus*, levando em conta os tipos de palavras que podem estar adjacentes a ‘toda’, verificou-se que ela ocorre em meio às seguintes formas:

³ Essa verificação não é quantitativa, porque se torna inviável trabalhar com quantidade nesse caso, uma vez que os programas de computador utilizados não fazem distinção estrutural automaticamente. Na amostra, a diversidade de acompanhamentos de ‘tod*’ é limitada, de modo que os números obtidos unicamente a partir dela não expressariam a realidade. O quadro foi montado a partir de intensas observações da pesquisadora, mas não esgota as possibilidades de acompanhamento do item lexical estudado.

Quadro 2 – Adjacências de ‘toda’

Estrutura do sintagma	Exemplo
ϕ + toda + art	toda a (sua força)
ϕ + toda + dem	toda essa (conceituação)
ϕ + toda + indef	toda uma (geração)
ϕ + toda + poss	toda nossa (história)
ϕ + toda + pro	toda ela (estava)
ϕ + toda + subst	toda vez (que)
adv + toda + art	quase toda a (partida)
conj + toda + poss	e toda nossa (tradição)
prep + toda + art	de toda a (temporada)
prep + toda + poss	de toda nossa (literatura)
prep + toda + subst	de toda espécie
pro + toda + verblig	ela toda estava
pro + verblig + toda	ela está toda
subst + toda + ϕ	(essa) porcariada toda
subst + toda + adj	sandália toda no tom
subst + toda + prep	túnica toda em (renda)
verb + toda + art	passar toda a
verb + toda + poss	marcou toda nossa
verbau + toda + part	foi toda inspirada

Observa-se que ‘toda’ está estreitamente associada à presença de substantivo na mesma sentença – uma vez que artigos, pronomes demonstrativos, indefinidos e possessivos precedem substantivos – e não raro adjacente a esse substantivo. O artigo também se destaca como acompanhante dessa flexão; de forma que se poderia dizer que a estrutura básica em que ela ocorre é *toda(+art)+subst*. Podem anteceder ‘toda’ advérbios (apenas os advérbios ‘quase’ e ‘praticamente’ aparecem no *corpus* antecedendo esse vocábulo), conjunções, preposições, pronomes retos, substantivos e verbos; bem como a sucedem verbos, substantivos, pronomes retos, demonstrativos e possessivos, mas, nessa posição, não ocorrem advérbios e somente nela aparecem artigos e adjetivos.

Ainda, percebe-se que ‘toda’ pode iniciar períodos e finalizá-los; não ocupa posição de núcleo do sintagma a que pertence; e que apenas duas estruturas apresentam a seqüência *subst+toda*, o que poderia insinuar que essa seqüência não alcança expressiva freqüência na língua escrita; porém, isso deve ser confirmado por um levantamento quantitativo.

Passando a uma análise quantitativa, observaram-se os sintagmas que ‘toda’

ajuda a compor, no que tange às funções sintáticas deles na sentença. As funções desses sintagmas, bem como o percentual de ocorrência de cada um na amostra são os seguintes:

Tabela 7 – Função do sintagma que contém ‘toda’

Função	%	Exemplo
Adjunto adverbial	38	[Toda vez] que a amava, no areal, sonhava com o corpo de Lindinalva.
Complemento de verbt _d	26,3	Reparou [que toda notícia no Plano...]
Sujeito	11,7	[Toda literatura] digna desse nome caracteriza-se por uma quantidade indefinida de significações potenciais, que se atualizam na relação dialógica entre texto e leitor.
Adjunto adnominal	7,6	Essa pode ser a prova [mais surpreendente de toda a temporada].
Predicativo (antecedendo um adjetivo ou uma locução adjetiva)	5,9	Sou corintiano, minha família é [toda corintiana], nasci no Corinthians, jogo aqui desde os nove anos, fui bicampeão pelo Corinthians.
Complemento nominal	3,5	Ao destruí-los, põe sobre os ombros de um só deus a responsabilidade [de toda esta bagunça que grassa no universo].
Locução verbal (antecedendo o particípio na voz passiva analítica)	2,3	Sua nova coleção [foi toda inspirada] nos mais tradicionais sapatos norte-americanos.
Complemento de verbt _i	2,3	Raí participou ativamente [de toda a partida].
Agente da passiva	1,2	Isso se forem de fato confirmadas as informações divulgadas [por toda a mídia].
Adjunto adverbial de oração	1,2	Protesta, esperneia – [com toda a razão] – contra um novo imposto de 0,25%, que vai retirar alguns trocados de sua renda.

Nota-se que ‘toda’ está presente em vários tipos de sintagmas com função nominal, pois aparece inserida em oito diferentes sintagmas que desempenham essa função (incluindo os iniciados por preposição); apesar disso, sua ocorrência em sintagmas com função adverbial é a mais alta, quase 40%, e o restante, aproximadamente 60%, está distribuído nos outros nove itens, com destaque maior para os com função de complemento de verbo transitivo direto e menor para os com função de agente da passiva e adjunto adverbial de oração⁴.

⁴ Essa expressão é utilizada por Bechara (1964, p.95) para designar um advérbio, ou adjunto adverbial, que se refere não apenas a um termo da oração, mas ao conteúdo de uma oração inteira, à declaração total. Observa-se que se trata de uma opinião do autor da sentença sobre o fato que nela trata.

Dessa forma, entende-se que as conjunções, preposições e outros tipos de palavras que possam iniciar esses sintagmas podem anteceder ‘toda’. Vale destacar que, na maioria das sentenças em que há verbo de ligação, ‘toda’ está no sujeito; mas antecede o predicativo nos casos em que tal verbo não está explícito. Na amostra, quando a voz verbal é passiva analítica, ‘toda’ posiciona-se entre o verbo auxiliar e o principal ou, em apenas um caso, insere-se no sujeito da sentença. Nos três casos em que há a voz passiva sintética, ‘toda’ está no sujeito, e, em um deles, não poderia ser movida, por conta do sentido que assume (ver exemplo de Sujeito na Tabela 7). Esse caso será abordado adiante.

Quanto à função do vocábulo em estudo dentro do sintagma em que está inserido, obteve-se o seguinte, levando em conta a posição em que esse item se encontra na sentença:

Tabela 8 – Função de ‘toda’ no sintagma em que está inserida

Função	%	Exemplo
Adjunto adnominal	92,4	Um jovem acompanha as gravações do programa durante toda uma semana e, no final, comenta o que achou.
		Seja porque a música e a poesia suavizem a dor, toda a criatura triste acha no canto um supremo consolo.
Adjunto adverbial	7,6	Ao chegarem à varanda, Ana Rosa, já em trajes de passeio, os esperava para sair toda debruçada no parapeito da janela.
		Clodovil disse que a casa foi toda revirada e que notou a falta de um aparelho de videocassete e de algumas jóias.

Indubitavelmente, a forma ‘toda’ é um adjunto na sentença, e esse é o uso regular desse termo. O fato de ser adnominal em mais de 90% dos casos revela sua especialização em acompanhar um nome. Já a sua possível ocorrência como adjunto adverbial – ignorada pela maioria dos teóricos da língua – está relacionada à presença de um verbo de ligação, ou de um verbo na voz passiva, ou, ainda, de um adjetivo (predicativo), em casos em que não há verbo de ligação explícito na sentença. Vale lembrar que, nessa situação, os exemplos encontrados mostram que essa palavra varia, concordando em gênero com o nome ao qual está diretamente vinculada. Essa classificação é possível e não definitiva, e será refinada posteriormente.

Contudo, muitas vezes, essa distinção entre os tipos de adjunto parece ser apenas estrutural, tendo em vista que o sentido se mantém. Por exemplo, uma comparação entre as sentenças 3a e 3b, abaixo, mostra que, a princípio, não há diferença semântica em relação a ‘toda’ em ambas, embora no caso 3a ela esteja adjacente ao nome – seria então adjunto adnominal (‘inteira’) – e em 3b esteja adjacente ao verbo – o que caracterizaria um adjunto adverbial (‘inteiramente’).

Já em 3c, apesar de estar adjacente ao verbo, está claro que ‘toda’ desempenha função de adjunto adnominal – pois faz parte do complemento do verbo – e não há como considerá-la adjunto adverbial. Contudo, em 3d, o valor desse vocábulo, pode-se defender, seria de adjunto adverbial (‘totalmente’, ‘muito’), nesse caso, modificador de um adjetivo que, com ele, formaria o predicativo do sujeito. Retomar-se-á essa discussão mais adiante.

- 03
- a) **Toda** ela estava nos filhos, e agora especialmente na carta e no discurso.
 - b) Se existe ironia neste filme, ela está **toda** no título.
 - c) A companhia abastece **toda** a região Nordeste, com exceção do Maranhão.
 - d) Estava eu **toda** prosa em meu carro novo [...] quando um senhor me abordou com um revólver.

Quanto à frequência de combinação semântica, obtiveram-se, em quantidades expressivas (maiores do que 100) no *corpus*, os seguintes trigramas:

Tabela 9 – Trigramas notáveis de ‘toda’

Ordem	Trigrama	Quantia
1	de toda a	754
2	em toda a	504
3	toda a sua	344
4	durante toda a	337
5	por toda a	297
6	toda vez que	272
7	com toda a	250
8	para toda a	189
9	quase toda a	154
10	que toda a	148
11	toda a sociedade	135
12	e toda a	126
13	toda a história	125
14	toda a família	121
15	em toda parte	113
16	a toda a	110
17	toda a população	109
18	toda a parte	106
19	toda a cidade	102
20	toda a vida	102

A relação entre ‘toda’ e o artigo repete-se em 18 das 20 expressões mais frequentes que envolvem esse termo. A expressão mais comum do tipo *toda(+art)+subst* é ‘toda a sociedade’ – a expressão ‘a sociedade toda’ ocorre apenas três vezes no *corpus* –, o que confirma a estreiteza das relações de ‘toda’ com um artigo feminino no singular e um substantivo feminino e no singular, nessa ordem, e a preferência do usuário da língua escrita por essa seqüência.

O artigo presente no trigrama 1 indica que um substantivo completa a seqüência, e isso assinala que o substantivo que sucede tal estrutura varia acentuadamente, mas há sempre um nome após o trigrama. Assim, mais forte é a relação entre ‘toda’ e o substantivo, pois ele está sempre presente, haja artigo ou não, e isso explica que ocorram, entre ‘toda’ e o substantivo, palavras que antecedem um nome (como pronomes demonstrativos e possessivos). Isso também revela que a posição desse vocábulo é relativamente fixa nas sentenças; e esses fatos complementam a discussão sobre a estrutura sentencial em que ‘toda’ está presente (Quadro 2). A partir daí, outra questão põe-se, concernente à alternância da presença

do artigo posterior a ‘toda’. Essa questão suscita outra, tratada nas seções seguintes.

A estrutura *art+subst+toda* realmente não é preferida pelo usuário da língua escrita, mas marca sua presença, embora geralmente modesta. Acerca disso, vejam-se os cotejos abaixo, nos quais se observa que os casos em que as duas estruturas são, de certa forma, expressivas são o 6, o 7 e o 8, sobretudo os dois últimos, embora os quatro trigramas correspondentes a eles não sejam tão comuns no *corpus*:

Tabela 10 – Cotejo quantitativo entre as estruturas *toda+art+subst* e *art+subst+toda* com sentido equivalente

	Estrutura 1	Q.	Estrutura 2	Q.
1	toda a sociedade	135	a sociedade toda	3
2	toda a história	125	a história toda	5
3	toda a família	121	a família toda	15
4	toda a população	109	a população toda	4
5	toda a cidade	102	a cidade toda	7
6	toda a vida	102	a vida toda	37
7	toda a noite	46	a noite toda	22
8	toda a casa	16	a casa toda	8

Voltando à Tabela 9, nota-se que a maior frequência de ‘toda’ está associada a preposições; juntas, as frequências expressivas dessa natureza totalizam 2.331 trigramas, dos 4.398 selecionados. Os trigramas das demais flexões devem confirmar esse fato e fornecer mais dados para uma discussão posterior.

Há ainda de se destacar que a ocorrência insistente de certas combinações de palavras caracteriza frases feitas da língua (melhor explicadas na seção 3.3.4). Em se tratando de ‘toda’, parece ser, de certa forma, raro esse tipo de combinação, pois o trigrama mais frequente, e que caracteriza um sintagma completo – não apenas parte de um – chega a 135 somente (Tabela 9, item 11), de forma que, em tese, não é necessário aplicar um teste estatístico para comprovar tal raridade⁵.

⁵ Com esse item lexical, dantes constituía-se uma locução, ‘toda via’, que significou primeiro “em todo o caminho, completamente”; depois, “sempre, constantemente, a cada passo”; então, tomou valor adversativo e tanto apareceram juntas essas palavras que acabaram soldando-se em uma conjunção (NASCENTES, 1966, p.735). No *corpus*, ‘todavia’ aparece 841 vezes. Tendo em vista que se distanciou das funções de ‘tod*’, ela não será analisada neste trabalho.

3.3.2 Significado

As noções semânticas⁶ de ‘toda’ observadas nas sentenças da amostra são: reforço (quando a palavra a que ‘toda’ se associa diretamente já representa um todo por si), intensificação (quando se agrega a um adjetivo ou, em alguns casos, a um verbo no particípio), conjunto (quando está associada a coletividade) – nesse caso, há situações em que pode ser interpretada como partição (sinônimo de ‘qualquer’) –, distribuição (com valor de ‘cada’) e inteireza (representa completude). Vale salientar que a classificação ‘reforço’ de forma alguma manifesta que o vocábulo em questão seria dispensável na sentença, mas que elucida uma idéia, sem modificá-la, apenas fortalecendo-a. Essas noções são atribuídas, a princípio, pelo termo a que ‘toda’ está vinculada, ou até pela subjetividade do leitor, e nem sempre classificar é uma tarefa fácil, devido ao limite dos sentidos não ser sempre claro e a interpretação das palavras pelo usuário não ser somente uma.

Da amostra, originou-se a seguinte tabela:

Tabela 11 – Noções semânticas da flexão ‘toda’ na amostra e percentual de ocorrência de cada uma

Noção	%	Exemplo
Reforço	45,3	Criou-se toda uma nova doutrina econômica para combater, no Primeiro Mundo, o potencial de crescimento das indústrias dos países emergentes.
Conjunto	15,9	Conheceu aventureiros de toda espécie. Durante a campanha, sofremos um grave dano, porque toda a propaganda sugeria nosso nome para “constituente”.
Sinônimo de qualquer	3,5	Em toda novela tem mulher com peito de fora na abertura da novela.
Distribuição	11,8	E tem um amigo meu que toda noite chega bêbado a essa roda viva de festas. Caminhões da prefeitura passam toda semana nos PEVs, recolhem o material da coleta seletiva e transportam o lixo.
Inteireza	15,9	Então, quem assume o poder [...] é o outro príncipe [...] cuja sombra se projeta, simbolicamente, sobre toda a narrativa. Existem inclusive casos extremos de anfíbios que permanecem com essas estruturas por toda a vida.
Intensificação	7,6	A crista já toda despencada.

⁶ As noções ‘conjunto’, ‘inteireza’, ‘partição’ e ‘distribuição’ foram obtidas em Said Ali (1964) e Almeida (1999); ‘intensificação’ baseia-se em Almeida (1999), mas foi refinada pela pesquisadora; e ‘reforço’, como dito, foi elaborada pela pesquisadora a partir de observações do comportamento de ‘tod*’ no *corpus*.

Dessa forma, os – aqui funcionalmente considerados – adjuntos adverbiais em potencial encaixam-se na noção de intensificação⁷, e os adjuntos adnominais distribuem-se pelas demais noções, conforme a carga semântica do nome a que se referem.

Os casos em que ‘toda’ transmite a idéia de ‘qualquer’ (exemplo: ‘toda novela’) não foram englobados pela noção de conjunto, porque ‘qualquer’ está associada a uma individualização e não a uma coletividade. Contudo, a diferença de significado é sensível, para o usuário da língua, e a classificação instável, cabendo, em muitas situações, defesa para ambos esses sentidos. A noção ‘distribuição’ abrange as circunstâncias em que ‘toda’ remete à expressão ‘cada’, também havendo individualização de algo. Em se tratando de ‘distribuição’, a grande maioria das ocorrências da amostra refere-se a períodos de tempo (dia, semana, mês, ano, hora). As locuções ‘em/por toda (a) parte’ foram classificadas como conjunto, por passarem a idéia de universalidade.

Nessas duas situações, ‘toda’, cuja denotação primordial é inteireza, é usada para separar algo dentre outros do mesmo grupo. Assim, quando se diz ‘toda semana’, faz-se referência não ao conjunto de sete dias, mas a algum desses dias; e a expressão ‘toda novela’ pode aludir ao conjunto de tudo o que for novela, sem idéia de especificação, ou a qualquer das que estejam nesse conjunto, então havendo especificação. Não obstante, a dificuldade de classificação de alguns casos fez com que se optasse por não determinar um limite rígido para as ocorrências de cada uma, devido ao risco de haver erro.

Ainda, chama a atenção o fato de a presença do artigo não bloquear a noção semântica de conjunto, como se observa no segundo exemplo dessa noção na tabela anterior. Como esse caso pode configurar uma polêmica com a área semântica, o exemplo merece ser melhor avaliado, juntamente com outros do mesmo tipo.

⁷ Na Tabela 11, no exemplo de ‘intensificação’, a flexão intensifica um adjetivo, ou o predicativo, como ocorre no exemplo 3d, em que há verbo de ligação.

A Tabela 11 mostra a maior frequência do uso de ‘toda’ como ‘reforço’, que, muitas vezes, é estilístico. Entretanto, os demais percentuais, dado que não são demasiado baixos, mostram que o uso de ‘toda’ não é, em certa medida, concentrado, em termos de significado.

Observou-se que, em alguns casos, apesar de ‘toda’ acompanhar o sujeito, ela pode ser deslocada na sentença para junto do verbo, sem prejuízo da idéia expressa. Nessas situações, embora varie em gênero, o comportamento do item poderia ser considerado adverbial (‘inteiramente’, ‘completamente’), uma vez que há, na amostra, sentenças estruturalmente similares que não admitem a moção desse termo (quando distributivo; partitivo; quando expressa conjunto e não há artigo após a flexão – a menos que haja pronome possessivo, que dispensa artigo; e quando um artigo indefinido vem anteposto ao substantivo – conforme mostra o excerto 4a). Assim, há que se decidir entre os significados ‘inteiro’ e ‘inteiramente’ para proceder uma classificação, conforme ilustrado no grupo de excertos seguinte, no qual os exemplos 4c e 4e foram modificados para ilustração.

- 04
- a) Durante seu reinado, **toda uma geração** de nobres e burgueses esclarecidos influiu no governo.
 - b) **Toda** a nossa história fundava-se sobre uma homogeneidade radical.
 - c) *A nossa história fundava-se **toda** sobre uma homogeneidade radical.*
 - d) **Quase toda** sua obra se perdeu no confisco de bens.
 - e) *Sua obra se perdeu **quase toda** no confisco de bens.*

Como a diferença na mensagem é sutil, uma opção parece ser avaliar a estrutura da sentença e classificar a flexão de acordo com o termo ao qual ela está adjacente (assim, em sentenças como 4b e 4d, a flexão seria adjunto adnominal e, em casos como 4c e 4e, seria adjunto adverbial). Ainda, observa-se que, quando há um adjetivo posposto ao item em questão – desde que o sujeito não represente um conjunto de entes – parece ser ele intensificado pelo item, que então assume valor adverbial (como em: ‘Uma linda casa, *toda alva* e louçã...’); mas, embora não tenham aparecido na amostra, observaram-se casos no *corpus* em que o sentido das palavras da sentença é que define o valor sintático desse item. Porém, isso vale para o singular

feminino. É preciso amadurecer esses raciocínios a partir da verificação do comportamento de ‘todo’ e das outras flexões.

3.3.3 Alternância da Presença de Artigo e Ambigüidade

Como é de se esperar quando se trabalha com dados autênticos da língua, há algumas ocorrências cuja interpretação é, pelo menos aparentemente, mais complexa. Observa-se que, quando no singular, o item lexical em questão, às vezes, provoca uma ambigüidade, configurada na duplicidade de significado. Isso ocorre quando a noção semântica que norteia o sintagma em que ‘toda’ está inserida denota, normalmente, ‘inteireza’, mas poderia ser interpretada como ‘conjunto’, mediante a presença de artigo posposto ao item lexical em análise. Assim, esse caso dá-se quando os limites do significado não estão claros. Há várias situações como essa no *corpus*, mas nem sempre existe ambigüidade, porque, por motivos delineados a seguir, não se admite dupla interpretação, embora o artigo esteja presente.

Desse modo, em tese, há duas interpretações para a sentença seguinte, extraída de um texto de uma coluna de variedades de um jornal:

05 **Toda a** pintura é abstrata no sentido de que realizamos uma abstração.

Nesse caso, a noção de conjunto ficaria por conta de se tratar de todas as pinturas existentes, e a noção de inteireza teria vez se a interpretação correspondesse a uma única pintura que é abstrata. Entretanto, o contexto deixa claro que se trata de conjunto; aliás, a sentença consegue transmitir essa idéia por si. Poder-se-ia até mesmo argumentar que a presença do artigo é em virtude de erro de digitação; mas tal hipótese perde força devido a essa sentença não ser um caso isolado na amostra, apesar de não representar a forma mais comumente usada para expressar tal circunstância.

Esse caso pode até mesmo ocorrer quando ‘toda’ está posposta ao substantivo núcleo do sintagma – construções desse tipo são, como visto, comparativamente raras

no *corpus*. Mas, em geral, o sentido é delimitado (6a), de modo que a noção semântica preservada é apenas a de inteireza, e sempre há artigo:

- 06 a) O viajante, em vez de pagar diárias, compra o apartamento do hotel e garante sua estada por 20, 30 anos ou para **a vida toda**.
 b) **A obra toda** de Joyce é “eu vi o mundo, ele começava em Dublin”.

Observa-se que, em 6b, a interpretação semântica é de conjunto, mas é atribuída pelo uso metonímico de ‘obra’ e não por ‘toda’, pois, se ‘toda’ fosse removida, a idéia de conjunto seria preservada. Além disso, ‘obra’ vem determinada, de forma que não se trata de uma verdade aplicável a tudo o que for obra, apenas à de Joyce, diferentemente do que ocorre no excerto 5.

As situações potencialmente ambíguas encontradas no *corpus* levam a crer que, apesar da duplicidade de interpretação e da clara preferência por não se utilizar determinante, o artigo é facultativo, sem prejuízo da idéia que se quer exprimir. Nesse aspecto, merecem destaque, para comparação, os seguintes co-textos, extraídos da população:

- 07 a) “Nesse sentido, elas têm **toda a razão**”, diz Sarkovas.
 b) Em seus livros e discursos, ele não poupou (e com **toda razão**), a oligarquia nordestina, habituada a seculares mamatas e nutrida por odiosos esquemas de manipulação política.
- 08 a) Desde que Humanitas [...] é o princípio da vida e reside em **toda a parte**, existe também no cão, e este pode assim receber um nome de gente, seja cristão ou muçulmano.
 b) “Eles estão em **toda parte** e nos matar é um prazer”, diz o carpinteiro de barcos.
- 09 a) O telégrafo, poucos dias antes de assustar **toda a gente** com a notícia de uma grande revolução socialista na Europa inteira, dera-nos a triste notícia de que o ex-imperador...
 b) E o meu amor ficava na conversa de **toda gente**.

Observa-se que o sentido empregado é o mesmo em cada par, independentemente do artigo, e que, sem dúvida, o leitor entenderia os co-textos, sem que a presença ou ausência de artigo causasse estranheza ou uma parada na leitura para reflexão sobre tais fatos. Ainda, a alternância de artigo, a princípio, não está

relacionada ao tipo de texto, dado que representantes de ambas as opções foram encontrados, por exemplo, em textos literários e jornalísticos. No entanto, essas considerações somente podem ser completas após a observação do comportamento de ‘todo’ e das formas plurais.

3.3.4 Locuções ou Frases Feitas

Neste trabalho, entende-se por locução ou frase: a reunião de termos que juntos expressam uma idéia, sem nada afirmar ou negar, expressão do pensamento (ALMEIDA, 1999, p.18). Assim, uma frase feita é um grupo de palavras que encerra um significado próprio e ocorrem juntas com certa frequência. Já expressão idiomática é entendida como toda dicção que não se analisa ou está em conflito com os princípios gerais da gramática, conforme expõe Said Ali (1964, p.227).

Em virtude disso, alguns trigramas podem ser frases feitas, quando se encaixam nessas definições, o que parece não ocorrer com os trigramas notáveis de ‘toda’, expressos na Tabela 9, pois nenhum deles encerra um significado próprio. Porém, isso não significa que não haja frases feitas formadas por outros n-gramas, cujo uso seja específico em casos não muito frequentes na língua escrita. Por exemplo, na amostra, há ‘toda a gente’, que designa coletividade humana – um significado único, engendrado pela combinação dos vocábulos –, e é, portanto, uma frase feita.

Nesse sentido, as locuções adverbiais que envolvem ‘toda’ e aparecem na amostra totalizam oito e são: ‘a toda’ (uma vez), que é uma locução adverbial, denota intensidade e parece ser uma redução de ‘a toda a força’ ou ‘a toda a pressa’, de forma que, nela, ‘toda’ seria um adjunto adnominal de um nome implícito; ‘em toda parte’ (três vezes); ‘em toda a parte’ (uma ocorrência); e ‘por toda parte’ (três vezes) – ver excertos 8a e 8b.

3.4 TODO

3.4.1 Estrutura, Função Sintática e Combinação Semântica

Pelo mesmo processo utilizado anteriormente, revelou-se que ‘todo’ ocorre em meio às seguintes adjacências:

Quadro 3 – Adjacências de ‘todo’

Estrutura do sintagma	Exemplo
ϕ + todo + art	todo o nosso (conhecimento)
ϕ + todo + dem	todo esse (volume)
ϕ + todo + indef	todo um (sistema)
ϕ + todo + pro	todo ele (se transformou)
ϕ + todo + subst	todo título (que)
adv + todo + art	quase todo o (papel)
conj + todo + art	como todo o (time)
conj + todo + poss	que todo nosso (imaginário)
prep + todo + art	de todo o Brasil
prep + todo + poss	com todo seu (gênio)
prep + todo + subst	em todo amistoso
subst + todo + ϕ	(esse) tempo todo
verb + todo + adj	caminhava todo atrapalhado
verb + todo + art	recebe todo o
verbau + todo + part	foi todo escrito
verblig + todo + poss	é todo nosso
verblig + todo + subst	era todo felicidade

Tal como ocorre com o item lexical anterior, ‘todo’ está fortemente associada à presença de substantivo no mesmo sintagma – uma vez que artigos, pronomes demonstrativos, indefinidos e possessivos precedem substantivos – e não raro adjacente a esse substantivo. O artigo destaca-se como acompanhante desse item lexical; assim, poder-se-ia dizer que a adjacência mais comum a ‘todo’ é *todo(+art)+subst*, de forma que a relação entre ele e o substantivo é a mais presente. Chama atenção o fato de não ocorrer a estrutura *pro+todo*, diferentemente do que acontece com ‘toda’. No *corpus*, os advérbios que antecedem ‘todo’ são ‘nem’, ‘praticamente’ e ‘quase’. Assim, podem anteceder esse item lexical: advérbios,

conjunções, preposições, substantivos e verbos; bem como sucedê-lo: artigos, substantivos, adjetivos, pronomes retos, possessivos, demonstrativos e verbos. Esse item também pode começar períodos e terminá-los.

Já as funções sintáticas dos sintagmas que contêm o item em questão e o percentual de cada um deles encontrado na amostra são os seguintes:

Tabela 12 – Função do sintagma que contém ‘todo’

Função	%	Exemplo
Adjunto adverbial	51,5	A demanda por cursos de especialização e globalização do profissional cresce [no mundo todo].
Sujeito	17,6	[Todo esse volume], no entanto, foi superior somente ao movimento financeiro do vencimento de dezembro...
Complemento de verbtd	15,1	...no fim de contas, fico conhecendo [todo o norte do Brasil]
Adjunto adnominal	7,7	Os visitantes [...] se retiram, os filhos da casa, com a roupa [de todo o dia], caem no samba.
Predicativo (antecedendo adj. ou loc. adj.)	2,1	O mérito é [todo nosso].
Complemento nominal	2,1	A escravidão ou o desespero desta Europa provocaria a desapareição de valores indispensáveis [a todo homem digno deste nome].
Locução verbal (antecedendo o particípio na voz passiva analítica)	2,1	Supondo que o salário de 2.000 URVs [seja todo pago] no final do mês e a URV tenha uma variação de 40% até lá, a empresa descontará, a título de IR, o mesmo valor de CR\$ 207.065.
Complemento de verbti	1,1	Os usuários que dispõem de antena parabólica [...] poderão assistir [a todo o movimento da feira].
Aposto	0,7	[Todo -poderoso do J.P. Morgan no Brasil], Arthur Kalita assistiu ao jogo Brasil e Estados Unidos com um olho no gol e outro na calculadora.

Percebe-se que a maior participação de ‘todo’ na língua escrita é em sintagmas com função de adjunto adverbial, que comportam quase 52% das ocorrências desse termo, e as menores participações são como complemento verbal indireto e aposto, que juntas equivalem a 1,8% das funções sintáticas. Na grande maioria das sentenças em que há verbo de ligação, ‘todo’ está no sujeito. Quando a voz verbal é passiva analítica, em 11 casos na amostra, ‘todo’ posiciona-se seis vezes entre o verbo auxiliar e o principal, e insere-se no sujeito cinco vezes; na voz passiva sintética, ‘todo’ está no sujeito nas duas únicas ocorrências da amostra, e poderia ser movida, sem prejuízo do sentido da sentença.

Nota-se que somente não se repetem, nos dois primeiros itens lexicais estudados, sintagmas cuja frequência é bastante baixa; então, essa condição permite entender que é possível que esses itens ocorram também nesses sintagmas, uma vez que tais ocorrências são raras. De maneira geral, os sintagmas e dados percentuais equivalem-se nas tabelas correspondentes de ambos os itens no singular.

Quanto à função do vocábulo dentro do sintagma, obteve-se o seguinte:

Tabela 13 – Função de ‘todo’ no sintagma em que está inserida

Função	%	Exemplo
Adjunto adnominal	88,7	Assim, as cidades com mais de 500 mil habitantes não terminariam todo o processo no mesmo dia.
		Tinha amigos no Bronx, no Brooklin, em todo lugar.
Núcleo nominal	6,7	Ao todo , oito pessoas morreram no desastre.
		É preciso propor novas direções para nosso próprio futuro e para o futuro da humanidade como um todo .
Adjunto adverbial	4,6	O ex-governador também afirmou que seu programa de governo é todo embasado no desenvolvimento.
		Todo branco, com recheio de nozes e baba de moça, lembrava um bolo de noiva.

Novamente, constata-se que a função do item lexical em estudo, nas sentenças, é sobretudo adjunto; mas, no presente caso, também pode ser núcleo nominal, a partir de sua existência como substantivo. A classificação ‘Núcleo nominal’ está associada, principalmente, às locuções adverbiais que contêm ‘todo’. Comparando-a à função de adjunto adverbial – que, vale lembrar, é potencial e será ajustada na discussão sobre a análise – nota-se que é expressiva a quantidade de tais expressões, em se tratando de ‘todo’. Isso será abordado adiante.

Quanto à frequência de combinação semântica, os destaques são os seguintes trigramas:

Tabela 14 – Trigramas notáveis de ‘todo’

Ordem	Trigrama	Quantia
1	em todo o	1.586
2	de todo o	1.049
3	todo o país	1.005
4	todo o mundo	768
5	Durante todo o	468
6	o tempo todo	397
7	como um todo	362
8	por todo o	257
9	todo o dia	255
10	todo o estado	246
11	de todo modo	209
12	para todo o	208
13	todo tipo de	183
14	em todo caso	182
15	quase todo o	181
16	todo o seu	172
17	com todo o	169
18	todo o Brasil	166
19	que todo mundo	155
20	que todo o	142
21	todo o tempo	136
22	todo o processo	131
23	no mundo todo	129
24	todo mundo sabe	121
25	a todo o	119
26	todo o ano	111
27	todo o sistema	107

Neste caso, os trigramas que envolvem preposição são 4.376 dentre 9.014 e, na forma de substantivo, destaca-se somente a expressão ‘como um todo’, que aparece 362 vezes no *corpus* e é uma frase feita.

Diferentemente do que ocorre com ‘toda’, há as seguintes frases feitas: ‘todo o mundo’ e ‘como um todo’. Uma vez que a palavra ‘mundo’ ocorre em quatro dos 27 trigramas selecionados, uma investigação de sua relação com ‘todo’ merecia ser feita. Uma pesquisa a partir de outros n-gramas mostrou que a expressão ‘todo mundo’ ocorre 1.745 vezes no *corpus*. Como não aparece nos principais trigramas, deduz-se que ela se combina com palavras diferentes, não havendo preferências significativas. Parece conveniente considerar que essa expressão constitui uma frase feita utilizada para expressar coletividade, normalmente, humana; nela, o ‘mundo’ não representaria

lugar – que estaria associado à noção semântica de inteiro – mas sim conjunto. Nesse caso, faria diferença no sentido a presença de artigo anterior a ‘mundo’. Todavia, retomar-se-á esse assunto na seção sobre o significado, porque uma verificação mais criteriosa nesses dados revelou que isso nem sempre é verdade, como ilustra o exemplo seguinte – aliás, ‘em todo mundo’ ocorre, neste sentido, 30 vezes no *corpus*:

- 11
- a) É um novo sistema que facilita o acesso de corporações à Internet – a rede mundial de mais de 40 milhões de computadores em **todo mundo**.
 - b) A canção “Don’t Cry For Me Argentina” se torna sucesso em **todo mundo**.
 - c) Este movimento provocou queda em cascata dos mercados acionários em **todo mundo**.

Outro achado é que o quadrigrama ‘em todo o mundo’ ocorre 455 vezes e ‘de todo o mundo’ ocorre 246, o que indica que apenas 67 das expressões ‘todo o mundo’ não são acompanhadas pela preposição ‘em’ ou ‘de’. Ainda, observa-se que a expressão ‘no mundo todo’, equivalente à anterior, mas sem artigo antecedendo ‘todo’, aparece 138 vezes no *corpus* e ‘do mundo todo’ outras 85. Esse fato reforça a idéia da aptidão de ‘todo’ para ser adjunto, pois a expressão iniciada por ‘em’ tem características adverbiais e a iniciada por ‘de’ tem características adnominais, o que confirma os dados da Tabela 13. Há também o fato de a maioria dos trigramas que representam sintagmas nominais indicar tempo ou lugar. Essa análise será complementada com o cotejo entre trigramas a seguir:

Tabela 15 – Cotejo quantitativo entre as estruturas *todo+art+subst* e *art+subst+todo* com sentido equivalente

	Estrutura 1	Q.	Estrutura 2	Q.
1	todo o país	1.005	o país todo	8
2	todo o mundo	768	o mundo todo	38
3	todo o dia	255	o dia todo	91
4	todo o estado	246	o estado todo	0
5	todo o Brasil	166	o Brasil todo	12
6	todo o tempo	136	o tempo todo	397
7	todo o processo	131	o processo todo	7
8	todo o ano	111	o ano todo	65

A disparidade no uso das duas estruturas é patente, pois, das 3.436 ocorrências da Tabela 15, as referentes à estrutura 1 somam 2.818 (em torno de 82%) e as correspondentes à estrutura 2 totalizam 618 (quase 18%). Nota-se que o item 6 surpreende, por a preferência do usuário da língua ser pela estrutura 2 (aproximadamente, 75% das ocorrências) e estar na contramão dos demais; porém, a estrutura 1 desse item também é utilizada (representa, aproximadamente, 25% das ocorrências). Chamam a atenção, ainda, a diferença extrema entre as duas estruturas do item 1, pois a estrutura 1 equivale a 99,2% das ocorrências dessa expressão, e a ausência de exemplos da estrutura 2 do item 4. Além disso, nota-se que o item 8 é o único que sugere um certo equilíbrio no uso – ou pelo menos uma boa taxa de ocorrência de ambas as estruturas. Finalmente, percebe-se a clara preferência pela colocação ‘todo o mundo’ em detrimento de ‘o mundo todo’; mesmo não esquecendo que são equivalentes as expressões ‘em todo o mundo’ e ‘no mundo todo’, a preferência confirma-se.

3.4.2 Significado

As noções semânticas de ‘todo’ observadas nas sentenças da amostra são: reforço, intensificação, conjunto, partição, inteireza e distribuição. Os casos em que ‘todo’ é sinônimo de ‘qualquer’ foram classificados conforme feito com ‘toda’. Os sentidos são atribuídos, a princípio, pelo termo a que ‘todo’ está vinculada e nem sempre é fácil distingui-los, devido ao limite dessas noções ser sutil, como no caso da flexão anterior.

Tabela 16 – Noções semânticas de ‘todo’ na amostra e percentual de ocorrência de cada uma

Noção	%	Exemplo
Reforço	44,7	São suspeitas que, enquanto durarem, constituem motivo de vergonha para todo o país.
		Coloquei todo o meu esquerdismo no bolso e me diverti na rua.
Conjunto	21,5	“Está todo mundo esperando pelo novo Gol”.
		Sempre me pareceu absurdamente injusta a generalização de achar que todo político não presta.
Sinônimo de qualquer	7,4	E, de todo modo, a elegância no filmar permanece intacta.
Distribuição	4,9	Todo dia surge uma notícia que evidencia a violência que o pedetista não conteve...
		E lá, onde as coisas andam relativamente bem, a folha cresce cerca de 1% todo mês.
Inteira	14,1	Na área do Butantã, uma obra fez com que o abastecimento fosse cortado o dia todo .
		O problema não ocorreu em todo o período previsto no projeto.
Intensificação	7,4	– Vamos à capela, antes que anoiteça de todo .
		O ex-governador também afirmou que seu programa de governo é todo embasado no desenvolvimento.

Observa-se que, dessa vez, a noção ‘intensificação’ foi incrementada com a presença da locução ‘de todo’, que não simplesmente reforça um termo, mas modifica um verbo ou um adjetivo. Mesmo assim, seu percentual é parecido com o de ‘toda’, como também ocorre com ‘reforço’. Em se tratando de ‘distribuição’, nas ocorrências da amostra repetem-se as referências a períodos de tempo (momento, dia, mês e ano).

A exemplo do que acontece com ‘toda’, observa-se que, em alguns casos, apesar de o vocábulo ‘todo’ acompanhar o sujeito – a mais polêmica das participações do item em análise no singular –, ele pode ser deslocado na sentença para junto do verbo, sem que a idéia que exprime seja prejudicada, assumindo comportamento que se poderia considerar adverbial (‘inteiramente’, ‘completamente’). Porém, na amostra, o sujeito das outras sentenças estruturalmente similares que não admitem a moção desse termo é ‘todo mundo’ (‘conjunto’), que, como dito, é uma frase feita. Também com ‘todo’ os significados a escolher são ‘inteiro’ ou ‘inteiramente’ para proceder a uma classificação, e uma opção parece ser classificar esse item lexical levando em conta o termo ao qual ele está adjacente na estrutura da sentença; como ilustram os excertos seguintes, em que 12b e 12d foram alterados para fins de análise. Vale a afirmação anterior de que um adjetivo posposto ao item no singular – desde que o

sujeito não represente um conjunto de seres – é sempre intensificado por esse item, que então assume valor adverbial.

- 12
- a) Desde então a sua vida mudou radicalmente; **todo** ele se transformou nos seus modos de ver e julgar.
 - b) *Desde então a sua vida mudou radicalmente; ele se transformou **todo** nos seus modos de ver e julgar.*
 - c) O problema **todo** está no óleo de coco usado por 70% dos cinemas norte-americanos para estourar as pipocas.
 - d) *O problema está **todo** no óleo de coco usado por 70% dos cinemas norte-americanos para estourar as pipocas.*

Após a discussão, nesse âmbito, sobre as duas formas do singular, percebe-se que, em determinadas circunstâncias, os limites da distinção entre o adjetivo e o advérbio podem ser imprecisos, dado que há mais de uma possibilidade de interpretação, ou seja, o mesmo item lexical, nesse caso, pode estar relacionado diretamente a diferentes tipos de termos da oração, embora varie em gênero.

3.4.3 Alternância da Presença de Artigo e Ambigüidade

O caso da alternância de artigo sem alternância de sentido repete-se com ‘todo’, conforme se observa nestes exemplos:

- 13
- a) Fez-lhe sentir que querendo **a todo o custo** naquela noite segurar o Teotônio, temia que os de casa desconfiassem disso.
 - b) Isso constituía um objetivo político e ideológico mais presente do que negar, **a todo custo**, a nobreza das paixões (e das falhas) humanas.

Poder-se-ia questionar a validade do exemplo, a partir do argumento de que um deles é literatura romântica. Todavia, quando se trabalha com *corpus*, não se admite que ocorrências comprovadamente corretas sejam ignoradas, o que inviabiliza tal argumento. Além disso, o texto literário sofreu atualizações, para que pudesse ser compreendido hodiernamente, a partir do que se conclui que se julgou haver sentido em manter o artigo.

- 14 a) Viveu em Lisboa alguns meses, “viajando **todo tempo**”.
 b) Dicionários incorporam **todo o tempo** novas palavras e expressões.

Nas sentenças acima, o que há é uma locução adverbial, em que o artigo ou a ausência dele não interfere no significado. A mesma circunstância dá-se com os exemplos abaixo, nos quais ‘todo’ participa de sintagma nominal.

- 15 a) Isso segundo Leone, que está sempre contra **todo o tipo** de inovação.
 b) Os termos dessas relações podem ser submetidos a **todo tipo** de transformação

Também nos co-textos seguintes, merece atenção o artigo presente.

- 16 a) Criar a gente uma rapariga com **todo o carinho**, e no fim ter aquela recompensa!
 b) “Neste período, **quase todo o pêssego** vendido no Brasil fica concentrado na nossa região”, afirma o gerente.
 c) Os convidados, após assistir à festa do santo, uma vez comidos e bebidos, se retiram, os filhos da casa, com a roupa de **todo o dia**, caem no samba.
 d) O SBT começa a exibir o “Resumo da Copa”, **todo o final de noite** após o “Jô Onze e Meia”.

Nota-se que é indiferente a presença do artigo na sentença 16a; mas não em 16b, pois, supostamente, há margem para uma ambigüidade. Contudo, o contexto esclarece que não se trata de um pêssego inteiro, mas sim do total de pêssegos vendidos no Brasil – aliás, parece não haver prejuízo da idéia central se for movida a expressão ‘quase todo’ para depois de ‘fica’, pois ‘o pêssego’ garante a idéia de conjunto por si. Já no co-texto 16c, há duplicidade de interpretação, pois a presença de artigo permite que se entenda que ‘todo’ se refira à roupa usada durante um dia inteiro – inteireza – ou à roupa que se usa cada dia – distribuição (que não causa ambigüidade na amostra de ‘toda’). Isso também ocorre em 16d, em que se pode entender que a tal transmissão dura o final de noite inteiro ou que é feita todas as noites – que é o significado mais provável.

Ainda sobre a presença de artigo, vale retomar o assunto iniciado com os n-gramas: ‘todo mundo’ e ‘todo o mundo’. Observaram-se os seguintes dados:

Tabela 17 – Cotejo quantitativo entre as estruturas ‘todo mundo’ e ‘todo o mundo’

	Trigrama	Q.	Quadrigrama	Q.
1	em todo mundo	30	em todo o mundo	455
2	todo mundo sabe	133	todo o mundo sabe	10
3	todo mundo que	74	todo o mundo que	12

Esses dados mostram que, apesar de não ser preferido pelo usuário da língua, supostamente, há uso em que é indiferente a presença de artigo, pois ambos os n-gramas podem funcionar como adjunto adverbial de lugar e como sujeito de um verbo pessoal (‘saber’ é o mais expressivo, mas não é o único que acompanha o trigrama no *corpus*); e o item 3 também indica uso pessoal dos dois n-gramas. No que concerne a isso, vejam-se os excertos a seguir:

- 17
- a) Este movimento provocou queda em cascata dos mercados acionários **em todo mundo**.
 - b) Não tenho dúvidas de que ele vai representar melhora no padrão de vida **em todo o mundo**.
 - c) **Todo mundo sabe** que um está a fim do outro.
 - d) Mas **todo o mundo sabe** também que essa não é a prioridade da elite que domina as finanças...
 - e) Quando saio só com meninas, as mães nem ficam preocupadas, achando que eu posso defender **todo mundo**.
 - f) A expressão mais espantosa desta ânsia de beijar é o beijoqueiro, aquele cara que beija **todo o mundo**.

Um exame na seleção ‘todo o mundo’ revelou que, em 62 sentenças, essa expressão está empregada de uma forma em que, normalmente, esperar-se-ia ‘todo mundo’, como em:

- 18
- a) Houve um tumulto dentro do plenário e **todo o mundo** achou que aprovando a reeleição teria que aprovar a desincompatibilização para candidatar-se a...
 - b) Pegue-se o caso das mensalidades escolares. **Todo o mundo** ficou condoído com a situação das escolas particulares, cujo futuro estaria ameaçado.

Ainda, há casos em que a coletividade expressa é determinada e, então, sempre há artigo – por exemplo: ‘todo o mundo científico’ –, o que reforça a idéia de que ‘todo o mundo’ não necessariamente está relacionada a lugar. Também, nota-se que alguns fatores impedem que, por causa da questão do artigo, a sentença seja ambígua. A preposição que acompanha o n-grama é o principal deles. Dentre elas, destacam-se as mencionadas ‘em’ e ‘de’, e, em alguns casos, também ‘por’; já as preposições ‘para’ e ‘com’ não ajudam nesse aspecto, pois não impedem o duplo sentido. Outro fator é a semântica da sentença em si, que ativa o conhecimento do usuário da língua, de modo que se bloqueia a ambigüidade de sentido. Assim, entende-se que ‘todo o mundo’ pode funcionar como ‘lugar’ e ‘coletividade – nos co-textos analisados – humana’, não causando problemas de interpretação. Isso significa que o falante distingue os casos que seriam ambíguos e lida perfeitamente com as duas formas. Aliás, vale lembrar que os tipos de textos em que se alterna a presença de artigo, nesse sentido, são variados.

A partir disso, calcula-se que, quando ocorre a expressão ‘todo o mundo’, há aproximadamente 8% de chance de ela ser empregada no sentido em que ‘todo mundo’ é mais usual. Com os dados obtidos no *corpus* e mais uma amostra do mesmo fenômeno em outro *corpus*, pode-se aplicar o teste qui-quadrado (nos moldes explicados no capítulo 2), no qual as amostras são confrontadas, e que dirá se tal uso é significativo. Vale destacar que, dada a dificuldade de separar qualitativamente, ou seja, por significado, a expressão ‘todo o mundo’, posto que envolve seleção manual, esse teste fica limitado.

Em virtude disso, os n-gramas que servem de base são ‘todo o mundo sabe’ e ‘todo mundo sabe’, pois não há como eles serem utilizados em outro sentido que não seja ‘coletividade de seres’. A segunda amostra (7.330 dados) foi obtida na rede mundial, em conformidade com o exposto no capítulo 1.

Os dados de trabalho formam a seguinte tabela:

Tabela 18 – Dados para o primeiro teste qui-quadrado

Expressão	Quantia	
	NILC	Internet
	Amostra 1	Amostra 2
Todo o mundo sabe	10	331
Todo mundo sabe	133	6.999
Total	143	7.330

Esse teste trabalha com duas hipóteses, a hipótese nula ou de trabalho (H_0) e a hipótese alternativa (H_1). A primeira deve ser, basicamente, uma negação daquilo que o pesquisador quer provar; assim, ela admite que as diferenças nos dados são casuais, isto é, que não há dependência. Quando os dados mostram que essa hipótese é falsa, aceita-se a alternativa, que, em geral, reflete aquilo que o pesquisador quer provar, ou seja, que há dependência entre as variáveis (BARBETTA, 1998, p.185).

As hipóteses assumidas são, tendo em vista as quantidades de cada caso existentes nas duas amostras:

- H_0 : A expressão ‘todo o mundo’ usada para designar ‘coletividade de seres’ não configura um uso na língua escrita.
- H_1 : A expressão ‘todo o mundo’ usada para designar ‘coletividade de seres’ configura um uso na língua escrita.

O próximo passo é obter as frequências observadas e esperadas, uma vez que esse teste consiste em compará-las⁸.

⁸ O objetivo deste trabalho é aplicar o teste qui-quadrado e não explicar as fórmulas utilizadas para os cálculos. No Apêndice, expõem-se algumas dessas fórmulas. Essas informações estão detalhadas em Barbetta (1998, p.222-229).

Tabela 19 – Frequências observadas do primeiro teste qui-quadrado

Distribuição das Frequências Observadas
(%)

Expressão	Amostral	Amostra2	Total
Todo o mundo sabe	10	331	341
(%)	(7,0)	(4,5)	(4,6)
Todo mundo sabe	133	6.999	7.132
(%)	(93,0)	(95,5)	(95,4)
Total	143	7.330	7.473

Tabela 20 – Frequências esperadas do primeiro teste qui-quadrado

Frequências Esperadas
%

Expressão	Amostral	Amostra2
Todo o mundo sabe	7	334
Todo mundo sabe	136	6.996

A partir desses dados, faz-se o cálculo das parcelas que compõem a estatística do qui-quadrado, também chamadas de ‘contribuições do χ^2 ’. Essa estatística é uma espécie de medida de distância entre as frequências observadas e calculadas em cada casela, na suposição de as variáveis serem independentes (BARBETTA, 1998, p. 224). Quando as variáveis são independentes, a diferença entre as frequências observadas e esperadas tende a ser pequena, decorrente de variações casuais, e o valor do teste é baixo. Em virtude disso, um valor baixo na estatística indica que as variáveis podem ser independentes (p.226).

Tabela 21 – Estatística do primeiro teste qui-quadrado

Contribuições do χ^2

Expressão	Amostral	Amostra2
Todo o mundo sabe	1,85	0,04
Todo mundo sabe	0,09	0,00

A soma desses números indica o valor do qui-quadrado. Após o cálculo de χ^2 , obtém-se a probabilidade de significância (p), um valor tabelado, que representa a probabilidade de o teste acusar “um resultado tão ou mais distante do esperado, quanto

ao resultado da amostra observada” (p.189). O nível de significância (α) utilizado é 5%, valor arbitrário usual, conforme recomendado por Barbetta (p.228) e o grau de liberdade (gl) é 1, dado que se trabalha com duas categorias (uma tabela 2×2)⁹. Quando dos dados observados resulta um valor alto na estatística χ^2 , considerando-se os valores tabelados para esse teste, rejeita-se H_0 .

Neste caso, obteve-se $\chi^2 = 1,98$, então, deve-se rejeitar H_1 ¹⁰; portanto, a partir dos dados utilizados, não se pode afirmar que o emprego da expressão ‘todo o mundo’ para designar coletividade configura um uso na língua escrita.

Contudo, uma das exigências para que esse teste seja aplicado com eficácia é que a amostra seja grande. Neste caso, dada a impossibilidade de seleção dos dados e a baixa ocorrência da expressão escolhida para representar o fenômeno, considera-se que o resultado é discutível e somente aceitável se a amostra for aumentada e englobar mais situações. O resultado vale apenas para a expressão analisada; assim, estatisticamente, as ocorrências da expressão ‘todo o mundo sabe’, no sentido estudado, podem representar erros, frutos do acaso, dada sua baixa frequência. Todavia, os 8% de casos dessa ordem mencionados anteriormente indicam que é um uso a expressão ‘todo o mundo’ no sentido de coletividade de seres, e isso merece uma averiguação mais detalhada, que não foi possível desenvolver com base nos dados disponíveis, nesta oportunidade.

3.4.4 Locuções ou Frases Feitas

É relativamente alta a ocorrência de locuções adverbiais que envolvem ‘todo’ na amostra – há 19. A principal delas é ‘ao todo’ (sete ocorrências); seguida de ‘de todo’ (três ocorrências); ‘de todo modo’ (três ocorrências); ‘em todo caso’ (três ocorrências); ‘a todo vapor’, ‘a todo o custo’ e ‘a todo custo’, que aparecem uma vez cada. Fica claro que as locuções valem-se do sentido de ‘todo’ como ‘qualquer’, como adjetivo (‘máximo’) e como substantivo (‘total’), na sua formação.

⁹ $gl = (\text{linhas}-1) \cdot (\text{colunas}-1) = 1$

¹⁰ De acordo com a tabela indicada: $0,25 > p > 0,1$, ou seja, $p > \alpha$, ou seja, o teste aceita H_0 .

- 19 a) **Ao todo**, 26 morreram no incidente da semana passada, quando os aviões dispararam por engano...
- b) Continuo, **de todo modo**, avalizando as notícias que publiquei na última segunda-feira.
- c) Como o Corinthians precisa ganhar (embora o empate não seja **de todo** ruim, no primeiro jogo), é provável que entre com dois volantes.
- d) Claude Montana está **a todo vapor**.

Além disso, percebe-se que, quando um advérbio – como ‘quase’ ou ‘praticamente’, por exemplo – antecede ‘todo’, o item lexical em estudo não é somente reforço ou intensificação, mas parte de um recurso para expressar algo único, com sentido próprio, pois não há ocorrência de locuções como *quase o resto*, por exemplo.

Como discutido, ‘todo mundo’ encaixa-se na condição de frase feita. Ainda, como dito, a expressão invariável ‘como um todo’, que aparece quatro vezes na amostra, também é uma frase feita que reforça uma idéia de totalidade. Comprova-se isso com o fato de em três das sentenças em que ela ocorre já haver um *todo* expreso, sem necessidade de adicionar vocábulos para indicar essa totalidade; também, poder-se-ia fazer o reforço com o adjetivo ‘todo’ (no sentido de inteiro) simplesmente, mas prefere-se a locução, de forma que seu uso configura, principalmente, um estilo:

- 19 e) Talvez seja o momento de nos darmos conta de que cabe a nós, herdeiros de civilizações milenares, propor novas direções para nosso próprio futuro e para o futuro da humanidade **como um todo**.
- f) Esses fatos são: **como um todo**, os asiáticos têm QI superior aos brancos e os brancos têm QI superior aos negros.
- g) O Mercosul é bom para todos nós, porque nos abre os mercados dos outros países e tende a aumentar a renda nacional **como um todo**.

Entretanto, em um exemplo, o único encontrado em que há um adjetivo posposto à expressão, a noção semântica é de inteireza:

- 19 h) Ao mesmo tempo, porém, o jogo da homofonia faz desta pedra o símbolo do romance **como um todo** estruturado – a forma no sentido aristotélico do termo: talhada segundo as leis da refração...

3.5 TODAS

3.5.1 Estrutura, Função Sintática e Combinação Semântica

Na sentença, ‘todas’ ocorre em meio às seguintes formas:

Quadro 4 – Adjacências de ‘todas’

Estrutura do sintagma	Exemplo
ϕ + todas + art	todas as nossas (oportunidades)
ϕ + todas + dem	todas essas (normas)
ϕ + todas + part	todas paridas (recentemente)
ϕ + todas + poss	todas nossas (leis)
ϕ + todas + pro	todas elas
ϕ + todas + subst	todas etapas (do WQS)
adv + todas + art	nem todas as (províncias)
conj + todas + art	com todas as (perdas)
prep + todas + art	de todas as (circunstâncias)
prep + todas + dem	com todas aquelas (grinaldas)
prep + todas + poss	em todas suas (lojas)
prep + todas + subst	em todas atrações
pro + todas + verb	elas todas são
subst + todas + ϕ	(as suas) fichas todas
verb + todas + art	concentrou todas as (suas ações)
verb + todas + poss	colocar todas suas (expectativas)
verblig + todas + adj	estavam todas preparadas

A relação estreita com artigo e substantivo repete-se com a flexão ‘todas’; à primeira vista, surpreende a aparição da adjacência *todas+poss(+subst)*. O fato de não ocorrer *todas+indef(+subst)* já era esperado, visto que parece tratar-se de uma construção, a princípio, impossível. Antecedem essa flexão, que também inicia e finaliza períodos: advérbios, conjunções, preposições, pronomes retos, substantivos e verbos; bem como a sucedem: artigos, substantivos, pronomes possessivos, demonstrativos e retos, verbos e adjetivos.

As funções dos sintagmas que contêm o item em questão são as seguintes:

Tabela 22 – Função do sintagma que contém ‘todas’

Função	%	Exemplo
Adjunto adverbial	38	Deixa sua mulher em casa [todas as noites], mas considera sua rotina “normal”.
Complemento de verbtd	23,9	Oliver tomou [todas as drogas], matou no Vietnã...
Sujeito	21,8	[Todas as suas missões] foram no Oriente Médio...
Adjunto adnominal	5,6	Um grande painel de azulejos pintados mostra um mapa [com todas as rodovias existentes no Estado em 1920].
Complemento nominal	5,1	As empresas deveriam garantir, em seus contratos, o atendimento [a todas as enfermidades relacionadas no código internacional de doenças da OMS].
Complemento de verbti	3,6	Fay diz que só olhou [para todas as cicatrizes] no espelho há três dias, depois de já ter deixado a prisão.
Predicativo (posterior a verbo de ligação)	1,0	Essa discussão é [a mais urgente de todas].
Aposto	1,0	Contra a fome que, [de todas as nossas várias vergonhas é a mais séria e deplorável], os festeiros se engajaram.

Nota-se que permanece a constância dos primeiros quatro tipos de sintagmas em relação aos itens lexicais anteriores. Contudo, chama a atenção o fato de haver menos verbos de ligação envolvidos com essa flexão e o seu alto percentual de participação em sujeitos. Quando o verbo está na voz passiva, ‘todas’ está no sujeito ou é o sujeito; nos três casos da amostra em que o verbo ‘ser’ é intransitivo, ‘todas’ está deslocada do sujeito, posterior a esse verbo. Observa-se que os valores do Adjunto adverbial e do Complemento de verbo transitivo direto são bem parecidos com os correspondentes a ‘toda’.

Quanto à função do vocábulo dentro do sintagma, obteve-se o seguinte:

Tabela 23 – Função de ‘todas’ no sintagma em que está inserida

Função	%	Exemplo
Adjunto adnominal	83,8	Em Alcântara tivera uma capela de Santa Bárbara e obrigava a sua escravatura a rezar aí todas as noites. “Nós levamos todas essas notícias a sério”, disse Kamm.
(Fórico)		
Sujeito	8,6	Todas moram em casas construídas no interior da fazenda.
Aposto	7,6	Os casos foram registrados nas praias de Pitangueiras, Astúrias, Pernambuco e Enseada, todas no Guarujá.

A vocação de ‘tod*’ para ser adjunto é mais uma vez confirmada, porém, o item em questão não ocorre na função de adjunto adverbial, apenas adnominal. A princípio, isso não causa surpresa, dado que, a grosso modo, os advérbios são invariáveis. Desse fato, deduz-se que o usuário da língua sempre utiliza, em lugar de ‘todas’, outro recurso para expressar as circunstâncias de totalidade verificadas nas flexões anteriores. Resta investigar qual é esse recurso – ou esses recursos. Prosseguir-se-á nessa questão após o exame do comportamento de ‘todos’.

Ainda, observa-se que a possibilidade de ‘todas’ ser empregada como advérbio é bloqueada pelo fortíssimo sentido nominal que essa palavra encerra – mesmo porque é forma pluralizada – e que, conseqüentemente, fá-la referir-se a um substantivo e não a um adjetivo. Assim, o uso, ao que parece, é sempre pronominal.

Essa discussão serve de prelúdio à maior novidade desta primeira flexão do plural, que ficou por conta do seu escape à exclusividade da função de adjunto, mediante as funções que assume quando é fórica¹¹, as quais não ocorrem com as formas do singular. Quanto às funções correspondentes, elas são sujeito e aposto resumitivo – o qual iguala, por meio de uma qualidade comum, uma cadeia de nomes femininos, numa mesma oração e sem a presença de um verbo. Isso indica que a carga semântica dessa flexão, ou seja, as interpretações possíveis para ela, é maior do que a das formas do singular de tod*, pois somente assim pode valer por um sintagma nominal, atributo de pronomes substantivos.

Quanto à frequência de combinação semântica, obtiveram-se, em quantidades expressivas, os seguintes trigramas:

¹¹ Com base em Neves (2000, p.449), toma-se por ‘fórico’ um termo que tem como traço categorial a capacidade de fazer referência a outro termo mencionado ou a mencionar no discurso; e ‘não fórico’ um termo que não tenha relações de referência a outros termos do discurso.

Tabela 24 – Trigramas notáveis de ‘todas’

Ordem	Trigrama	Quantia
1	de todas as	1.452
2	em todas as	1.079
3	que todas as	404
4	com todas as	365
5	todas as suas	344
6	a todas as	330
7	quase todas as	261
8	para todas as	259
9	todas as pessoas	227
10	e todas as	179
11	todas as outras	155
12	por todas as	144
13	todas as partes	126
14	todas as informações	121
15	todas as regiões	112
16	todas as noites	108
17	todas as coisas	105
18	todas as empresas	104
19	todas as áreas	104

Os trigramas revelam que, na grande maioria das vezes em que ocorre, ‘todas’ está às voltas com um artigo feminino no plural, já que ele participa de todos os trigramas dessa flexão. A preposição marca presença em seis desses trigramas, num total de 3.629 ocorrências. O único possessivo que se destaca é ‘suas’ e o único advérbio é ‘quase’. Ainda, nota-se que há um trigrama cuja combinação de palavras expressa um sentido único de coletividade humana: ‘todas as pessoas’, cuja ocorrência é manifesta; portanto, trata-se de uma frase feita. Destaca-se apenas um trigrama usado, geralmente, em função adverbial: ‘todas as noites’.

Observa-se que a estrutura *art+subst+todas* não está nos trigramas mais comuns, igualmente como ocorre com a forma singular dessa flexão. Por meio de uma busca na população, constatou-se que não há *as pessoas todas*, *as outras todas* e nem *as partes todas*, que seriam estruturas similares aos 9^o, 11^o e 13^o trigramas nominais mais comuns dessa flexão.

3.5.2 Significado

As noções semânticas de ‘todas’ observadas nas sentenças da amostra são as seguintes:

Tabela 25 – Noções semânticas de ‘todas’ na amostra e percentual de ocorrência de cada uma

Noção	%	Exemplo
Conjunto	99	Todas as construções são no estilo vitoriano, que influenciou a arquitetura do Japão.
		Todas essas normas tratam da atualização monetária das demonstrações financeiras.
Reforço	1	As partidas da fase final do Mundial serão todas no ginásio do Ibirapuera, em São Paulo, a partir de sexta-feira.
		As imagens de nossa tragédia estão todas no palco.

Observa-se que o fato de ‘todas’ vir sempre acompanhada de artigo implica partes individualizadas que formam um conjunto e não um todo único – o que impede as noções de inteireza, distribuição, partição e intensificação. Por isso, dizer que a noção de conjunto ou de reforço engloba as ocorrências dessa flexão quando não pronominal (quando ‘todas’ é pronome substantivo, possui a semântica do pronome ‘elas’) é defensável, pois, em se tratando de plural, sempre haverá reforço para o conjunto que os outros termos da sentença, estando também no plural, expressam. Até mesmo a idéia de distribuição dos dois itens do singular passa a ser de conjunto ou reforço, embora todas as ocorrências dessa natureza, como acontece com os casos do singular, continuem referindo-se a períodos de tempo e seus contextos sejam similares aos anteriores – pode-se até defender a noção de distribuição, mas o plural parece fortificar a idéia de conjunto.

- 20
- a) **Todas as noites** sonho que peço um autógrafo ao Michael Jackson.
 - b) E tem um amigo meu que **toda noite** chega bêbado nessa roda viva de festas.

Porém, a princípio, parece haver uma distinção possível entre a presença de ‘todas’ no predicativo e as demais funções. Quando no predicativo, ‘todas’ pode ser

associada ao verbo de ligação, e então reforça o sujeito, resgatando-o. Nesse caso, somam-se as características de reforço e a pronominal – o que é preservado, portanto, é a função de adjunto. A idéia de ‘reforço’ alude ao fato de o deslocamento do adjunto adnominal para após o verbo obrigar a retomada do sujeito da sentença. Contudo, há poucos exemplos com predicativos na amostra, para uma análise mais profunda.

3.5.3 Alternância da Presença de Artigo

Foram encontradas ocorrências em que não há artigo entre ‘todas’ e o substantivo:

- 21
- a) Gosto de pensar que **todas nossas** atividades são espirituais, que tudo que vemos está envolvido por algo.
 - b) **Todas nossas** leis sociais têm a mesma estrutura.
 - c) No Parque da Mônica, você brinca em **todas atrações**.
 - d) O SBT estima gastar cerca de US\$ 25 milhões para, em um prazo de três anos, transferir **todas suas** instalações para a nova sede.
 - e) Antes de colocar **todas suas** expectativas no micro para iniciar o ano com disposição a mudanças, é bom checar qual o programa mais indicado.
 - f) O Hang Loose, assim como **todas etapas** do WQS no Brasil, conta pontos para o circuito doméstico junto com os campeonatos regionais.

A certa raridade de casos como esses e a estranheza que podem causar à primeira vista dão permissão à hipótese de que, no caso dos textos do *corpus*, trata-se de erro de digitação. Contudo, a repetição da ausência de artigo diante de pronome possessivo indica que, nessa circunstância, o artigo é opcional, embora a grande maioria das ocorrências desse caráter tragam artigo, o que configuraria a ausência de artigo sucedendo ‘todas’ diante de pronome possessivo um uso pouco privilegiado na língua escrita. Entende-se que é necessário verificar essa situação em relação a ‘todos’ para, então, realizar procedimentos estatísticos. No *corpus*, a proporção é a seguinte:

Tabela 26 – Ocorrências de ‘todas’ em que há ou não artigo antes de pronome possessivo

Expressão	Quantia
Todas as nossas	20
Todas nossas	4
Todas as suas	344
Todas suas	43
Todas as minhas	32
Todas minhas	5

3.5.4 Locuções ou Frases Feitas

Há apenas uma frase feita que envolve ‘todas’, na amostra, uma locução adverbial e uma sentença em que uso desse vocábulo é peculiar, apresentadas respectivamente a seguir.

- 22
- a) “**De uma vez por todas**, nossa greve não é política, é por salário”.
 - b) Andréa Beltrão **está em todas**.
 - c) Fulgêncio Batista **tomou todas** no Carnaval

Uma verificação do uso de ‘todas’ como em 22c no *corpus* revelou a sentença: ‘...aprontar todas e mais um pouco com a pantera cor-de-rosa’. A partir disso, pode-se refletir sobre a existência de um novo sentido para essa flexão, o de adjunto adverbial (comparável a ‘muito’), intensificador de um verbo intransitivo, ou sobre uma propriedade sua de formar uma expressão idiomática com alguns verbos.

Neste ponto, cabe uma emenda. Na discussão apresentada na página 80, observou-se que ‘todas’ encerra um forte sentido nominal que a faz referir-se a um substantivo, e não a um adjetivo, o que bloquearia um sentido adverbial. Agora, ‘todas’ surpreende ao referir-se a um verbo, e assim se encaixar na função primitiva que se atribuía a um advérbio.

3.6 TODOS

3.6.1 Estrutura, Função Sintática e Combinação Semântica

Nas sentenças, ‘todos’ pode ser rodeada pelos seguintes tipos de palavras:

Quadro 5 – Adjacências de ‘todos’

Estrutura do sintagma	Exemplo
ϕ + todos + art	todos os (seus bens)
ϕ + todos + dem	todos aqueles (cheiros)
ϕ + todos + poss	todos nossos (trunfos)
ϕ + todos + pro	todos eles
adv + todos + subst	quase todos membros (anafórico)
prep + todos + ϕ	(contar) com todos
prep + todos + art	com todos os (escritores)
prep + todos + poss	a todos esses
prep + todos + subst	com todos gêneros (musicais)
pro + todos + verb	eles todos sabem
subst + todos + ϕ	(seus) personagens todos
verb + todos + art	cancelou todos os (seus compromissos)
verb + todos + poss	promete todos seus (sucessos)
verbau + todos + part	estavam todos sentados
verblig + todos + adj	permaneceram todos concentrados

Não há surpresas em relação à flexão anterior. Repete-se a constância das relações da flexão com substantivos e artigos, principalmente, e pronomes – o que reafirma a aptidão para ser adjunto que ‘tod*’ tem –; e ‘todos’ também pode introduzir e terminar períodos. Podem anteceder ‘todos’: advérbios, preposições, pronomes retos, substantivos e verbos; e sucedê-la: artigos, substantivos, pronomes retos, possessivos e demonstrativos, verbos e adjetivos.

As funções dos sintagmas que contêm o item em questão são as seguintes:

Tabela 27 – Função do sintagma que contém ‘todos’

Função	%	Exemplo
Sujeito	38,4	[Todos os aparelhos usados no Brasil] permitem essa adaptação.
Adjunto adverbial	29,5	Cinco policiais foram destacados para vigiar o andar [durante todos os períodos].
Complemento de verbtd	14,3	A comissão já está ouvindo [todos os envolvidos com assassinatos no sertão sergipano].
Adjunto adnominal	10,3	Há paixões [de todos os tipos] no decorrer do romance...
Predicativo (posterior a verbo de ligação)	2,6	Os títulos a seguir são [todos inéditos no Brasil], mas podem ser encomendados nas livrarias importadoras.
Complemento nominal	2,6	Sarney já conta com o apoio [de praticamente todos os futuros senadores do norte do país].
Complemento de verbti	1,4	Será preciso testar opções táticas para quando não contar [com todos eles].
Verbo na voz passiva analítica	0,6	Fora o pouco caso com o roteiro do <i>show</i> – os <i>hits</i> dançantes [estão todos reunidos] no começo.
Aposto	0,3	O motivo é simples: entre todos os equipamentos utilizados [...] a antena é de longe o mais comentado.

Constata-se que, pela primeira vez, a função adjunto adverbial perdeu seu primeiro lugar em ocorrências para a função sujeito – que já havia crescido em participação com ‘todas’ –, mas ainda fica com a segunda posição e mantém um percentual elevado. O predicativo apresenta valor maior em relação à tabela de ‘todas’, mas que não chega a ser expressivo, e, na maioria dos casos em que há verbo de ligação, ‘todos’ está no sujeito; a participação em sintagmas cuja função é adjunto adnominal também cresceu, embora somente cinco pontos percentuais, em relação a ‘todas’; e as outras funções, de maneira geral, continuam uniformes. Na amostra, há duas ocorrências de ‘todos’ entre o verbo auxiliar e o principal da voz passiva; e, nessa voz verbal, há 13 casos em que ‘todos’ está no sujeito; ainda, há duas ocorrências de voz passiva sintética, nas quais ‘todos’ está igualmente no sujeito.

Quanto à função do vocábulo dentro do sintagma, obteve-se o seguinte:

Tabela 28 – Função de ‘todos’ no sintagma em que está inserida

Função		%	Exemplo
Adjunto adnominal		68,8	Em 13 de maio de 1888 a princesa Isabel assinou a Lei Áurea, libertando todos os escravos de nossa Pátria.
(Fórico)	Sujeito	17,2	Todos , ao pisar no Brasil, perdiam as suas raízes.
	Aposto	3,7	Nos três hospitais da cidade, há apenas oito leitos destinados a doentes de Aids, todos no Hospital Santa Isabel de Clínicas.
(Afórico)	Sujeito	10,3	Daqui para frente todos vão conhecer esses novos ídolos do esporte.
			Todos sabem que nos processos de estabilização há uma quase inevitável tendência à sobrevalorização lenta...

Desta vez, há diferença em comparação com a Tabela 23, configurada em uma nova função – característica de nome – em que não há valor fórico, mas referência a algo no mundo. Essa função, que expressa ‘coletividade’, equivale à segunda expressão nominal mais comum na estrutura *todo+art+subst*, que é ‘todo o mundo’, quando não antecedida por preposição e também a ‘todo mundo’, que é muito mais utilizada do que a anterior. Entretanto, neste caso, não há combinação com outras palavras, ou seja, não se trata de uma frase feita, pois o vocábulo ‘todos’ sozinho possui a carga de significado necessária para expressar uma coletividade.

Nessa função, ‘todos’ refere-se ao conjunto de elementos que o contexto abrange, sendo ora não delimitado – por ser inestimável – ora delimitado, em certa medida, mas sempre sem valor numérico, e, então, ‘todos’ assume gênero neutro; conforme ilustram, respectivamente, os seguintes exemplos:

- 22
- a) Hoje, a astrologia está ao alcance de **todos** e pode nos ensinar muito.
 - b) O trabalho na nova organização baseia-se no envolvimento e participação de **todos** nos planos e objetivos da empresa.

No caso da sentença 22a, devido à amplitude do contexto, poder-se-ia dizer que ‘todos’ significa ‘qualquer pessoa’ e, no segundo, devido a restrições de contexto, que ela se refere a quem se encaixa nas condições expressas na sentença.

Essa flexão não assume características adverbiais na sentença; supõe-se então que outros recursos sejam utilizados para intensificar adjetivos. Os excertos seguintes

– extraídos da população via busca no WordSmith – permitem uma noção do que pode ser parte da resposta, ao mesmo tempo em que corroboram a característica adverbial, ou seja, intensificadora, de ‘toda’ e ‘todo’ perante um adjetivo, visto que os contextos em que esses dois itens lexicais admitem esse aspecto sintático são comparáveis aos que seguem:

- 23
- a) Pelo menos uma pessoa morreu, três ficaram feridas e dez casas foram **totalmente** destruídas.
 - b) As laterais da alça ficaram **completamente** danificadas.
 - c) “Quero todos **totalmente** recuperados para as quartas de final”, afirmou o treinador corintiano.
 - d) Alguns se ornavam de penas e colares de ossos; outros **completamente** nus tinham o corpo untado de óleo por causa dos insetos.

Quanto aos trigramas, os destaques foram os seguintes:

Tabela 29 – Trigramas notáveis de ‘todos’

Ordem	Trigramas	Quantia
1	de todos os	2.057
2	em todos os	1.149
3	todos os dias	602
4	que todos os	428
5	a todos os	495
6	para todos os	435
7	com todos os	389
8	todos os seus	308
9	quase todos os	348
10	por todos os	338
11	todos os tempos	244
12	todos os outros	245
13	e todos os	245
14	todos os lados	212
15	todos os países	162
16	todos os jogadores	155
17	todos os anos	129
18	todos os tipos	125
19	todos os candidatos	124
20	todos os que	114
21	como todos os	134
22	todos os jogos	105

Nota-se que se repetem os dois primeiros trigramas da flexão anterior e com valores extremamente altos em relação aos demais. Destacam-se trigramas, geralmente, adverbiais, como o 3°, o 11° e o 17°, e nominais, como o 12°, o 15° e o 16°. Uma preposição está presente em seis dos primeiros dez trigramas, totalizando 4.863 aparições. Também se destaca o advérbio ‘quase’, que é o único dessa classe a participar dos trigramas com frequência maior do que 100. Dos possessivos, aparece apenas ‘seus’. Também, observa-se que não há trigrama notável que designe genericamente a coletividade humana, ao contrário do que ocorre com ‘todas’. A princípio, isso ocorre porque, como mencionado, o equivalente a ‘coletividade’ de ‘todos’ é uma palavra somente e não uma combinação. Cabe salientar que a participação do artigo ocorre em 100% dos trigramas.

Mais uma vez, a estrutura *art+subst+todos* não aparece nos trigramas. Uma busca na população pelo WordSmith indicou que não há *os dias todos* nem *os tempos todos* e que *os outros todos* ocorre três vezes e *os jogadores todos* quatro vezes.

3.6.2 Significado

As noções semânticas de ‘todos’ observadas nas sentenças da amostra são: reforço e conjunto. Esses sentidos são atribuídos, ao que parece, pelo termo a que ‘todos’ está vinculada. Os demais significados foram bloqueados pela noção de conjunto, como ocorre com a flexão anterior.

Essas informações geraram a seguinte tabela:

Tabela 30 – Noções semânticas de ‘todos’ na amostra e percentual de ocorrência de cada uma

Noção	%	Exemplo
Conjunto	96,8	Todos em Nova Atenas imaginaram que Tertuliano “seria o que quisesse” no novo governo.
		A sua sabedoria a tal respeito estava reduzida às batalhas do Paraguai, para ele a maior e a mais extraordinária guerra de todos os tempos.
Reforço	3,2	Os criminosos e os pobres eram todos arrebanhados nos mesmos locais de confinamento, tratados (ou, antes, maltratados) da mesma forma, exibidos como objetos...
		Seus habitantes, divididos em várias espécies, têm em comum o fato de serem todos nobres sustentados por gentios – vulgarmente conhecidos como bugrada ou negra da.

A exemplo da flexão anterior, observa-se que o fato de ‘todos’ vir sempre acompanhada de artigo impede as noções de inteireza, distribuição e partição. Porém, vale a mesma distinção possível entre a presença de ‘todos’ no predicativo e nos demais sintagmas: ‘reforço’, em que se preserva a função de adjunto. Todavia, novamente, há poucos exemplos com predicativos na amostra, apesar de haver mais neste caso, provavelmente por causa da maior quantidade, do que no item lexical anterior.

3.6.3 Alternância da Presença de Artigo

Foram encontradas na população ocorrências em que não há artigo entre ‘todos’ e o substantivo:

- 24
- a) A nova divisão vai apurar **todos** crimes de trânsito ocorridos no Estado.
 - b) Ele hoje defende muito mais amor como a cura para **todos nossos** males.
 - c) **Todos nossos** homens têm experiência de combate.
 - d) Assim como **todos seus** escritórios.
 - e) o “Dream Team” original venceu **todos seus** adversários nos Jogos Olímpicos de Barcelona, em 92.

Embora esses casos repitam-se em vários textos do *corpus*, ainda vale a possibilidade de que se trata de erro de digitação, sobretudo no primeiro exemplo. Contudo, a ausência de artigo diante de pronome possessivo novamente indica que, também neste caso, o artigo é opcional, o que configuraria a ausência de artigo sucedendo ‘todos’ diante de pronome possessivo um uso na língua escrita. No *corpus*, tem-se a seguinte proporção:

Tabela 31 – Ocorrências de ‘todos’ em que há ou não artigo antes de pronome possessivo

Expressão	Quantia
Todos os nossos	43
Todos nossos	13
Todos os seus	422
Todos seus	89
Todos os meus	54
Todos meus	0

De posse dessas quantias e considerando-se que pode haver casualidade nos dados, deve-se aplicar um teste de hipóteses que dirá se se trata de erro ou se pode ser considerado uso, ou seja, se é significativa a ausência do artigo ou não. O teste a ser aplicado é, novamente, o qui-quadrado, no qual duas amostras são confrontadas – a segunda amostra (68.673 dados) foi obtida na rede mundial. Pretende-se estudar se a presença de artigo posposto a possessivos no plural pode ser considerada acaso, isto é, erro, ou não. As hipóteses assumidas são as seguintes, lembrando que H_1 representa o que o pesquisador quer provar, ou seja, que há associação entre as variáveis:

- H_0 : *A ausência de artigo antes de possessivos no plural foi causada por algum erro.*
- H_1 : *A ausência de artigo antes de possessivos no plural configura um uso.*

Os dados utilizados no teste são:

Tabela 32 – Dados para o segundo teste qui-quadrado

Expressão	Quantia	
	NILC Amostra 1	Internet Amostra 2
Todas as nossas	20	8.094
Todas nossas	4	1.153
Todas as suas	344	47.672
Todas suas	43	6.137
Todas as minhas	32	4.962
Todas minhas	5	655
Total	448	68.673
Todos os nossos	43	13.719
Todos nossos	13	2.337
Todos os seus	422	60.997
Todos seus	89	9.246
Todos os meus	54	9.423
Todos meus	0	1.916
Total	621	97.638
Total geral	1069	166.311

As tabelas das freqüências a ser comparadas – observadas e esperadas – ficaram como segue:

Tabela 33 – Freqüências observadas do segundo teste qui-quadrado

Distribuição das Freqüências Observadas (%)			
Artigo*	Amostr1	Amostra2	Total
SIM	915	144.867	145.782
(%)	(85,6)	(87,1)	(87,1)
Não	154	28.951	29.105
(%)	(14,4)	(17,4)	(17,4)
Total	1.069	166.311	167.380
*: Artigo sucedendo possessivos quando a flexão está no plural			

Tabela 34 – Freqüências esperadas do segundo teste qui-quadrado

Freqüências Esperadas		
%		
Artigo*	Amostr1	Amostr2
SIM	931	144.851
NÃO	186	28.919

Após o cálculo de χ^2 , obtém-se a probabilidade de significância (p). O nível de significância (α) utilizado é 5% e o grau de liberdade (gl) é 1, dado que se trabalha com duas categorias. Já se sabe que o teste rejeita H_0 se o valor encontrado for grande, tendo em vista os padrões tabelados para esse teste. Neste caso, obteve-se $\chi^2 = 5,78$, que é um valor elevado. Em consequência, H_1 é aceita¹²; portanto, a ausência de artigo sucedendo as flexões plurais de ‘todo’ não decorre do acaso, é um uso na língua portuguesa do Brasil.

A estatística do teste configura-se da seguinte forma:

Tabela 35 – Estatística do segundo teste qui-quadrado

Contribuições do χ^2		
	Amostr1	Amostr2
SIM	0,28	0,00
NÃO	5,47	0,04

3.6.4 Mobilidade da Flexão

Observa-se mais uma vez que a posição das duas flexões do plural na estrutura varia. Esse fato não é atípico na língua, dado que outras categorias gramaticais também se organizam nas sentenças de diferentes formas, alterando suas posições, mas mantendo suas relações inter-termos; por exemplo, é o caso do sujeito e dos adjuntos. Dessa forma, uma vez que ‘tod*’ assume funções cujas posições dos termos mudam, é

¹² De acordo com a tabela mencionada: $0,025 > p > 0,01$, ou seja, $p < \alpha$.

natural que haja variação estrutural com o item em análise. Alguns exemplos ilustram e complementam esse aspecto do item lexical em estudo.

- 25
- a) Desta vez, os três revezaram-se na liderança e, até o finalzinho, **todos estavam** no páreo.
 - b) Tudo foi para a periferia do fato mais importante do dia, e **estavam todos** de radinho metafórico no ouvido, ligado em São Paulo.
 - c) Gosto do trabalho **deles todos**.
 - d) Má-fé, desconhecimento, falta de assunto ou entropia da informação? Provavelmente uma mistura, em doses variáveis, **de todos eles**.

Como exposto nos comentários da Tabela 27, em construções com o verbo ‘ser’, a flexão pode aparecer posposta a esse verbo, ou seja, deslocada do sujeito, a menos que haja o advérbio ‘nem’ ou que, obviamente, ‘todos’ seja o sujeito, pois, então, a posição posposta ao verbo representa opção pela seqüência *verbo+sujeito*, comum em português:

- 26
- a) Os aparelhos das academias **são quase todos** fabricados pela empresa norte-americana Life Fitness.
 - b) Nossa revolução se faz de forma tão pacífica e silenciosa, **que nem todos são** capazes de notá-la.
 - c) Os portugueses, que no início povoaram o Brasil, viviam somente no litoral, porque aí **estavam todas as riquezas que podiam ser vendidas na Europa**.

3.6.5 Locuções ou Frases Feitas

Foram constatadas apenas três locuções adverbiais com ‘todos’ na amostra: ‘de todos os tempos’.

3.7 TODINHA, TODINHO, TODINHAS E TODINHOS

As formas diminutivas do item lexical em estudo distribuem-se estruturalmente da seguinte forma, nas orações dos documentos do *corpus*, sempre em

textos jornalísticos, em colunas sobre cotidiano, especiais e cadernos de variedades ou em transcrições de diálogos de entrevistas¹³:

Tabela 36 – Adjacências das formas diminutivas e quantidades

Estruturas	Quantia	Exemplo
verb + (poss +) subst + todinho + adj	1	...mataram esse pessoal todinho da favela
verb + pro + verb + todinho	1	...sem tê-lo conhecido todinho .
verb + pro + todinho	1	Dá vontade de comer você todinho .
verblig + todinho + adj	2	...é todinho de plástico.
verb + art + subst + todinho	1	...ler o relatório todinho ...
verb + subst + todinha	1	...derrubar a MP todinha .
verb + todinha + adv	2	...é feita todinha em cima de “Ponteio”.
verblig + todinha + adj	2	...foi todinha de Mané Garrincha.
verb + todinhas + part	1	...foram todinhas vendidas.
verb + pro + todinhos	1	...quem manda aqui sou eu e eu vou ferrar vocês todinhos .

A partir da tabela acima, observa-se que as formas diminutivas de ‘tod*’ podem relacionar-se: a um verbo de ligação – e, nessa condição, posicionam-se entre o verbo e o predicativo, enquanto outras posições (como antecedendo ou sucedendo o sujeito, por exemplo) não são constatadas; a um verbo transitivo direto – então, ordinariamente, pospõem-se ao núcleo do complemento verbal, contudo, quando se trata de locução verbal, o núcleo do complemento pode ser atraído pelo verbo auxiliar, mas isso não altera a essência dessa ordem; e a um verbo intransitivo, situação em que se pospõem ao verbo. Não há ocasião em que essas formas coocorram adjacentes a artigos – diferentemente do que se verificou com ‘tod*’ – nem ao sujeito da sentença.

Já as funções dos sintagmas que contêm as formas diminutivas são as seguintes:

¹³ Para confirmação dos resultados, dado que as formas diminutivas são exíguas, observou-se se havia mais exemplos na parte não corrigida do *corpus* do NILC e constataram-se mais duas: em uma redação de vestibular e em um artigo de um jornal universitário sobre o vestibular, no qual pertencia ao depoimento de um aluno.

Tabela 37 – Função do sintagma que contém as formas diminutivas

Função	Q.	Exemplo
Complemento de vtd	5	Quanto mais rápido eles matarem [esse pessoal todinho da favela], mais rápido eles eliminam a possibilidade de que eles digam quem são os reais chefes disso dentro da polícia.
Predicativo (posterior a verbo de ligação)	4	À exceção das roupas, o mundo da Barbie é [todinho de plástico]. A série “Merveilles du Passé” [...] é [todinha recomendável] com seu “jazz africano” do final dos anos 50.
Adjunto adverbial	3	Ela se juntou [todinha] ao contato do ar frio e Battleship, se rindo, borrifou um pouco de água no corpinho escuro, fizeram as pazes.
Adjunto adnominal	1	Cheio de maus pressentimentos, também trocou o terno cinza que usava antontem por outro – [todinho preto].

Pode-se considerar que esses resultados eram esperados, uma vez que a estrutura das sentenças (Tabela 36) indica a relação estreita das formas diminutivas com verbo, substantivo (ou pronome substantivo) – na posição de complemento verbal – e adjetivo. Assim, estima-se também a função desses vocábulos dentro do sintagma:

Tabela 38 – Função dos diminutivos no sintagma de que fazem parte

Função	Q.	Exemplo
Adjunto adnominal	7	Vamos ler o relatório [todinho] e já marcamos uma reunião com Luís Eduardo para amanhã (hoje), para dizer o que está errado. “Mas tá errado, quem manda aqui sou eu e eu vou ferrar vocês [todinhos].”
Adjunto adverbial	6	Apesar de não estar creditada corretamente, a faixa “Beatitude”, do Jazz Brother é feita [todinha] em cima de “Ponteio”. Entrou Amarildo e a Copa foi [todinha] de Mané Garrincha.

Observa-se que, para as formas diminutivas, repetem-se as constatações válidas para os itens lexicais anteriores: o singular admite função adverbial e nominal; o plural admite somente função nominal, dada a noção de conjunto que encerra – apesar de o diminutivo plural intensificar a idéia de reforço; o fato de não haver forma diminutiva participando de sujeitos fortalece essa idéia. Mais uma vez, trabalha-se com os limites de significado entre um adjetivo (‘inteiro’) e um advérbio (‘inteiramente’, ‘totalmente’, ‘muito’). A título de comparação, vejam-se os excertos

seguintes, que ilustram esse caso e mostram que a flexão no plural, quando deslocada do sujeito e próxima ao verbo, reforça a idéia expressa na sentença, ainda mais se estiver no diminutivo, mas não chega a ter o potencial intensificador de um advérbio:

- 27
- a) As mil vagas **foram todinhas vendidas**.
 - b) As primeiras 70 unidades fabricadas, apresentadas em um pré-lançamento ocorrido em novembro, em Fortaleza, já **foram todas vendidas**.
 - c) Se o imóvel estiver situado [...] em um edifício cujas unidades já **foram totalmente vendidas** pela construtora, o melhor caminho é tentar revender a unidade.

A relativa mobilidade das formas diminutivas constatada na sentença – adjacência ao nome ou ao verbo – e o fato de acarretarem noção de superlativo – denotam ênfase –, porém, representando um reforço que não é indispensável para se transmitir a idéia central da mensagem, são indícios que conferem características próprias de adjunto a essas formas. Quanto aos sentidos observados para elas, eles são: conjunto, inteireza, reforço e intensificação, e não constam distribuição nem partição.

Conforme verificado já na seleção dos dados, há apenas uma ocorrência de cada uma das formas plurais no *corpus*. Entretanto, não se pode afirmar que elas são utilizadas mais raramente do que as outras duas, dado que a quantidade de ocorrências encontrada é extremamente pequena, embora o *corpus* seja robusto.

Não obstante, considerando o tamanho do *corpus* – justamente por isso as constatações sobre essas formas são relevantes –, é pertinente afirmar que na linguagem padrão escrita esses diminutivos são escassos. Essa asserção torna-se mais evidente ao levar-se em conta que, em duas ocorrências, trata-se da transcrição da fala de pessoas entrevistadas. Ainda, essas constatações permitem deduzir que esses itens são próprios da linguagem coloquial e mais comuns na língua falada.

3.8 CORPUS DIACRÔNICO

No código de ofícios do Conselho Supremo Militar ao presidente da Província de Santa Catarina, há as seguintes quantidades de ‘tod*’, dentre aproximadamente 33.400 palavras:

Tabela 39 – Quantias de ‘tod*’ no *corpus* diacrônico

Termo	Quantia
Todos	42
Todas	20
Todo	7
Toda	2
Formas diminutivas	0
<i>Total</i>	71

Observa-se que as flexões participam da massa diacrônica na mesma ordem de ocorrência constatada na amostra do *corpus* de trabalho, resguardadas as devidas proporções. A palavra ‘todos’ apresenta frequência equivalente a 0,13%, ‘todas’ 0,06%, ‘todo’ 0,02% e ‘toda’ menor do que 0,01%, sendo a última colocação nesse *corpus* a 329^a. Sem dúvida, a participação do item é expressiva no códice.

Os trigramas com ‘tod*’ são os seguintes:

Tabela 40 – Trigramas notáveis de ‘tod*’ no *corpus* diacrônico

Ordem	Trigrama	Quantia
1	a todos os	16
2	todos os officiaes	8
3	todos os que	6
4	todos os individuos	5
5	saber a todos	4
6	de todos os	3
7	que todos os	3
8	em todo o	2

Deve-se esclarecer que a seleção dos trigramas ficou prejudicada, porque separações silábicas e sinais gráficos entre palavras foram transcritos tal como estavam no original, o que desestrutura algumas expressões. Como se pode perceber, não há trigramas com a flexão feminina no singular e a quantidade do primeiro trigrama é expressiva em relação à dos demais. As preposições estão presentes, mas, diferentemente do que ocorre no *corpus* de trabalho, a preposição ‘a’ destaca-se mais

do que a ‘de’ e ‘em’ é a menor em aparições.

Quanto às estruturas e combinações dos itens lexicais em questão, constatou-se que não há surpresas em relação ao que foi visto no *corpus* de trabalho. A flexão ‘todos’ ocorre em sintagmas cuja função é nominal ou adverbial (sempre acompanhada de artigo ou pronome demonstrativo) e como sujeito fórico; para ‘todas’, valem os mesmos tipos de sintagmas e acompanhamentos; ‘todo’ também participa dos mesmos tipos de sintagmas; e ‘toda’ ocorre em um complemento verbal e em um adjunto adverbial. A função de ‘tod*’ no sintagma é sempre adjunto adnominal.

Embora haja pouca massa lingüística, ainda assim, alguns fatos merecem ser exemplificados:

- 28
- a) ...e pondo-se nos respectivos assentos as necessarias / clarezas, para **a todo tempo** constar esta Minha Imperial Determinação.
 - b) ...e / recolhendo-se ao Archivo do Corpo este documento para **a todo o tem= / po** constar, que a deliberação que se tomou a respeito do proposto foi / com pleno conhecimento de cauza justa.
 - c) ...o Brigadeiro Jo-/ sé Correia Picanço, o Coronel João Antonio de Mello, o Major / Antonio Baptista Camacho, e o Capellão Manoel de Santa / Anna Macedo, **todos** pertencentes ao Corpo d'Armada, e Impe- / rial Brigada de Artilheria da Marinha...
 - d) Que he sem duvida que pelo Regulamento de 1808, / feito e publicado pela Regencia de Portugal / para as Milicias d'aquelle reino, entra este / privilegio no numero dos concedidos aos corpos / Milicianos, porem que aquelle Regulamento nunca foi mandado por em pratica no Brazil, neste sentido **em toda a parte** cauzará / transtornos //

Nota-se que: na década de 1820, usava-se alternar a presença de artigo em expressões e situações idênticas, como ilustram as sentença 28a e 28b; ‘a todo tempo’ e ‘em toda a parte’ são duas locuções adverbiais presentes no código; a sentença 28c é um exemplo de uso de ‘todos’ como aposto resumitivo. Quanto ao sentido, observaram-se: conjunto, inteireza e partição.

4 PADRÕES DE OCORRÊNCIA

Neste capítulo, são detalhados e comentados os aspectos mais importantes da análise, bem como emendam-se discussões já iniciadas, a fim de identificar os padrões que cercam a ocorrência de ‘tod*’, na amostra, principalmente, e também no *corpus*, para inferir os resultados no uso que se faz desses itens lexicais no universo da língua escrita, por meio da proposta de um modelo de funcionamento empírico, que reúne as informações no final do capítulo. Para refinar as reflexões, os temas discutidos são comparados com o que discorrem acerca deles alguns teóricos da língua.

4.1 SIGNIFICADO E FUNÇÕES SINTÁTICAS

Da análise feita, pode-se tirar um padrão de funcionamento de ‘tod*’, a começar pela discussão do uso do artigo e sua interferência no significado e então passar para as questões sintáticas.

4.1.1 A Questão dos Artigos Sucedendo ‘tod*’

Neste ponto, a discussão sobre o artigo preceder ‘toda’ e ‘todo’ nos casos em que poderiam ser sinônimos de ‘qualquer’ está superada, pois, apesar de não ser realizado nem um teste estatístico para confirmar a hipótese de que é indiferente o uso do artigo nessa situação específica, dada a dificuldade de seleção quantitativa dos dados, os exemplos são muitos e diversificados o suficiente – até mesmo os textos em que aparecem são de diversos tipos – para admitir-se que se trata de um uso. Esse fato é reafirmado nos textos do código da década de 1820, no qual duas locuções adverbiais

idênticas no sentido aparecem com e sem artigo, sem prejuízo da idéia expressa, e não se trata de erro de digitação. Historicamente, a alternância de artigo nessa situação data de antes do quinhentismo e Camões já a praticava (SAID ALI, 1964, p.122).

Para confirmação prática, alguns co-textos do *corpus* que apresentam esse fenômeno foram impressos e entregues a dez usuários da língua, de níveis de escolaridade diferentes, e pediu-se-lhes que explicassem os excertos. A partir do que reportaram (em palavras de usuário e não de pesquisador), observou-se que o artigo não interferia no sentido, entendido como ‘conjunto’, sem haver ameaça, portanto, de ambigüidade¹⁴. Dessa forma, para o usuário da língua, não há diferença entre sentenças como *Todo homem é mortal* e *Todo o homem é mortal*, sendo uma distinção de significado, do ponto de vista empírico, estipulada por estudiosos da língua e não por quem a utiliza para comunicar-se; mesmo porque essa diferenciação faz parte somente da escrita, já que se aglutinam as duas vogais ‘o’ na fala, desaparecendo o artigo. Vale destacar que o discernimento do usuário acerca da sua língua é relevante para qualquer estudo lingüístico, pois ele é que a modifica, praticando-a, e não os teóricos.

Caso similar ocorre com ‘todo mundo’ e ‘todo o mundo’, que, pode-se afirmar, são expressões indistintamente utilizadas para designar ‘coletividade humana’ – apesar de a segunda apresentar, em tese, potencial para ambigüidade por conta da noção de ‘inteireza’ –, embora a primeira seja preferida pelos usuários. Também nesse caso, não ocorre ambigüidade e não há relação direta entre a segunda expressão e um uso mais formal, ou clássico, da língua, tendo em vista que ambas aparecem em textos jornalísticos e literários.

Nesse sentido, também está confirmado, até estatisticamente, que é um uso, embora não privilegiado pelo usuário da língua, a omissão do artigo entre ‘todos’ ou ‘todas’ e um pronome possessivo. Almeida (1999, p.352) revela que a supressão do artigo posterior às flexões do singular e do plural – em qualquer caso, não somente com pronomes possessivos – era comum em autores antigos e até preferida por eles. Então, constata-se que, simplesmente, a preferência mudou, mas o uso continua.

¹⁴ Considera-se desnecessário explicitar uma metodologia para essa técnica prática, uma vez que somente se desejava reforçar uma aceção confirmada pela análise, ou seja, não se extraíram dela conceitos novos.

4.1.2 Valor Nominal e Valor Adverbial

Primeiramente, cabe destacar que Neves (2000) não registra uso de ‘todo’ e ‘toda’ com valor adverbial em nenhuma situação – para essa autora, ambas são sempre pronomes indefinidos –, e nem o menciona Said Ali (1964). O fato de não ocorrer o mesmo valor com as formas pluralizadas e de outros casos apresentarem classificação difícil mostram que essa postura tem fundamento e é mais simples. Contudo, há casos em que o valor adverbial desses itens lexicais é, argumenta-se, quase explícito, pois se encontram em posição que somente advérbios ocupam nas sentenças; e esses casos não são explorados ou mesmo mencionados pelos dois teóricos. Já Almeida (1999) sustenta a postura de que assumem valor adverbial. Sabe-se que a frequência com que esse caso ocorre é baixa, pois não ultrapassa 7,6% das ocorrências, sendo sensivelmente mais constante em se tratando de ‘toda’. A locução ‘de todo’ também faz essa mesma função, mas então não há motivo para polêmica.

Uma vez que há divergência de postura entre teóricos da língua renomados acerca do valor adverbial de ‘todo’ e ‘toda’, e já se puseram considerações sobre o uso do artigo após os itens lexicais analisados no sintagma, esse assunto merece ser debatido, mesmo porque a classificação adotada no capítulo 3, como esclarecido, representa uma possibilidade, viável, é claro. Realmente, trata-se de uma questão controvertida, pois, em alguns casos, mesmo parecendo esses dois itens lexicais ter valor adverbial, a decomposição da sentença em que estão presentes revela que as suas conexões com os outros termos da oração não é o que aparenta ser. Por exemplo, vejamos os excertos seguintes:

- 29
- a) Romário disse: “Ficamos **todos cansados** no final. Acho que todos sentiram o calor.”
 - b) O meu cavalo levou-me para o Engenho Velho e eu daí me encaminhei para a Tijuca, onde cheguei ao meio-dia, **todo molhado** e fatigado pelos maus caminhos.
 - c) Estavam **todos armados** da cabeça até os pés, além da grande espada de guerra que batia as ancas do animal.
 - d) Com exceção de uma breve cena externa, filmada no centro de São Paulo, “Sábado” foi **todo rodado** no velho estúdio da Vera Cruz, em São Bernardo do Campo.
 - e) Mas, na indústria, é **praticamente todo aproveitado**.
 - f) A tabela do IR precisaria ser **toda revista**.

- 29 g) Quando pegamos as cópias do filme, estavam **todas mutiladas**, para diminuir a duração.
- h) Era florada, lindo véu de branca renda
Se estendeu sobre a fazenda, qual um manto nupcial
E de mãos dadas, fomos juntos pela estrada
Toda branca e perfumada, pela flor do cafezal.
(*Cascatinha e Inhana – Flor do Cafezal*)

Por ser variável, ‘tod*’ sempre resgatará um termo da oração, dado que concorda com ele em gênero e número; mas as formas do singular também podem intensificar, daí a dificuldade de classificação. Para agravar esse estado, a organização de ‘tod*’ na sentença é complexa. Em 29a, ‘todos’, embora deslocada da posição normal de adjunção, é adjunto adnominal do núcleo do sujeito, ‘nós’ (*todos nós* ou *nós todos*), que não está expresso estruturalmente, e pode ser movido sem prejuízo da sintaxe e da semântica (*Todos ficamos*); em 29c, ‘todos’ é o sujeito da oração, embora esteja adjacente ao predicativo e, em 29g, ‘todas’ é sujeito de ‘estavam’. Mas, em 29b, não há como considerar ‘todo’ adjunto do núcleo do sujeito, também não expresso, ‘eu’, visto que, estando no predicativo e embora – alguém possa argumentar – concorde com o sujeito, é a ‘molhado’ que ele se refere e com o qual também concorda em gênero; nesse caso, não há como mover ‘todo’: *todo cheguei ao meio-dia, molhado e fatigado*, por isso, sua relação direta é com o adjetivo ‘molhado’ – aliás, pode-se dizer que ‘todo’ não se refere a ‘fatigado’ –; a mesma situação ocorre em 29h, em que a referência de ‘toda’ é ‘branca’, não havendo modo de mover ‘toda’ para junto de ‘estrada’, nem defesa para a estrutura *adjetivo+adjetivo*, isto é, para o juízo de que um adjetivo modifique outro, numa seqüência, sem que haja vírgula, conjunção ou se trate de palavra composta. Em 29d, em que se considera ‘todo’ adjunto adnominal do núcleo do sujeito ‘Sábado’, observa-se que a moção de ‘todo’ para perto do sujeito acarretaria estranheza pela falta de artigo: *Todo “Sábado” foi...*, mas, como explicado por Almeida (1999), a omissão do artigo após o item lexical de trabalho – quer no singular, quer no plural – desde longa data, é uso no português; então, continua havendo respaldo empírico para a defesa de que, também em casos como esse, as flexões do singular são adjuntos adnominais; contudo, cabe defender que, nesse caso, ‘todo’ modifica o verbo que o antecede, o que acarreta uma classificação oscilante. Em 29f, ‘toda’ – e também ‘todo’, em casos análogos, – parece ser adjunto adverbial, mas,

argumenta-se aqui, também pode ser adjunto adnominal deslocado do núcleo do sujeito: *Toda a tabela do IR*, pois, estando o verbo na voz passiva, cria-se um efeito de reforço à idéia expressa na sentença – como discutido no capítulo 3 – mediante o deslocamento do adjunto do sujeito para perto do verbo principal. Tal fato repete-se em 29e, em que, mesmo ‘todo’ estando entre um advérbio e o verbo no particípio, poderia ser considerado adjunto adnominal de um sujeito não expresso – percebe-se que esse recurso de deslocar o item lexical, adjunto adnominal, para perto do particípio (voz passiva analítica) possibilita combinações que não exigem a explicitação do sujeito passivo, valorizando a harmonia entre as palavras e a sonoridade da sentença, que seriam prejudicadas se a estrutura fosse: *Mas, na indústria, todo é praticamente aproveitado*. Cabe esclarecer que, na voz passiva sintética, os componentes do sujeito, normalmente, ficam reunidos após o verbo, como é a tendência, nesse caso (exemplo: ‘Nesse princípio se baseou toda a descrição’).

Dessa exposição, conclui-se que há rija defesa para considerar-se que, quando no singular e antecedendo ou, mais raramente, sucedendo um adjetivo, o item lexical estudado assume valor adverbial, sobretudo quando não há verbo de ligação que o anteceda; propriedade que não há nas flexões do plural, pois, nessa circunstância, o item não permite intensificação, mas sim reforço – que se acentua quando a flexão está no diminutivo – da idéia expressa na sentença. Vale ressaltar que, quando não há artigo antes do núcleo do sujeito, nem sempre a moção de ‘todo’ e ‘toda’ é possível, ou preferível, sob risco de desarmonia na sentença ou inaceitabilidade – quando assume valor distributivo, partitivo e de conjunto (exemplo: ‘Toda literatura é...’) – mas isso também não descaracteriza a sua função de adjunto adnominal, apenas, nesses casos, ou prefere-se a manutenção do adjunto com o restante do sintagma de que participa, ou o sentido exige-o. Ainda, observa-se que há casos em que a classificação oscila entre adjunto adverbial e adnominal, valendo a intenção do autor ao expressar a sentença.

Antes de finalizar esta seção, vale ilustrar um pouco mais a questão da imprecisão dos limites entre o advérbio e o adjetivo, em certos casos. Para tal, observe-se o seguinte grupo de excertos, no qual 30a foi extraído de uma página eletrônica sobre cinema e os demais da população:

- 30
- a) Nell Fenwick é a mulher que ama, mas ela anda **meia confusa** em relação aos homens.
 - b) A Williams, já **toda poderosa**, era só uma das escudérias interessadas no piloto.
 - c) É uma raquete **quase inteira preta**, muito mais discreta.
 - d) O novo CD é envolvido por uma capa **totalmente preta e branca**, adornada com uma escrita críptica espelhada.

Em 30a, o advérbio ‘meio’ sofreu flexão de gênero, mas não há como classificá-lo de outra forma que não seja adjunto adverbial modificador de um adjetivo, a menos que, como comentado anteriormente, fosse aceita a estrutura *adjetivo+adjetivo*, mas isso não parece ter cabimento. A princípio, ‘meio’ – do latim *mediu*, por via popular – era utilizado como substantivo e numeral (metade), no século XIII, mas seu uso como advérbio data do século XV (HOUAISS, VILLAR e FRANCO, 2001, p.1883; FERREIRA, 1999, p.1309). Assim, comparando-a a ‘todo’, tem-se um adjetivo que, em tese, passou a assumir valor adverbial e um substantivo que passou a assumir esse mesmo valor, embora o primeiro varie em gênero e o segundo, ordinariamente, não. A sentença 30b é comparável à 30a, e, nela, ‘toda’ assume valor adverbial. Em 30c, observa-se que se deve considerar ‘inteira’ um adjunto adverbial, especialmente quando se faz um cotejo com a sentença 30d.

Assim, corrobora-se a afirmação de que os tais limites mencionados são mesmo sutis em certos momentos e se robustece o conceito de que ‘todo’ e ‘toda’ podem assumir valor adverbial. Vale salientar que os teóricos tradicionais são unânimes em aceitar que o advérbio, ou adjunto adverbial, modifica um adjetivo e pelo menos um aceita que há palavras que variam em gênero, mas cuja função sintática é adjunto adverbial.

É importante lembrar que uma pesquisa, não muito meticulosa, revelou que o caso 30a não ocorre no *corpus* e é escasso na rede mundial, o que leva a crer que, na língua culta, constitui o que seria considerado erro gramatical; todavia, na fala, sabe-se, é fato que casos assim ocorrem com certa freqüência, inclusive em pronunciamentos de pessoas escolarizadas.

4.1.3 Funções e Significados

No aspecto sintático, mais especificamente, a primeira dedução é a de que, quando singular, os itens estudados são adjuntos, pois não constituem núcleos de sintagmas nominais; já, quando plural, dividem-se em adjuntos e núcleos de sintagmas nominais; como substantivo, ‘todo’ participa de qualquer sintagma em que essa classe seja aceita. Os sintagmas em que os itens estudados ocorrem e suas funções neles são os seguintes:

- Sujeito: estando no singular, o item lexical é sempre adjunto adnominal e, estando no plural, pode ser, além disso, núcleo (fórico) – pois identifica a pessoa gramatical, fazendo referência a ela, e então satisfaz a função de pronome pessoal – e, somente no caso de ‘todos’, núcleo afórico – quando designa coletividade de entes. Nenhuma forma diminutiva foi encontrada nessa posição.
- Predicativo: quando está após verbo de ligação, ‘tod*’ é adjunto adnominal deslocado do sujeito da sentença; porém, as formas do singular, quando não há verbo de ligação explícito, sobretudo, mas também, em alguns casos, na presença dele, são adjuntos adverbiais que intensificam o adjetivo, com a noção de ‘muito’, ‘bem’, ‘totalmente’, ‘inteiramente’, e então preservam sua propriedade de adjetivo e variam em gênero; nessas circunstâncias, as formas pluralizadas constituem ora adjuntos adnominais deslocados do sujeito, que o resgatam – referenciando-o, portanto, assumem comportamento pronominal –, e reforçam a idéia de conjunto que ele encerra, ora apostos. Isso vale para as formas diminutivas de cada flexão.
- Voz passiva: quando ‘tod*’ sucede verbo na voz passiva sintética ou precede particípio na analítica, pode ser adjunto adnominal ou adverbial, dependendo da interpretação dada à sentença, ou seja, do sentido atribuído às palavras.
- Complemento verbal: nessa condição, as formas do singular são sempre adjuntos adnominais do núcleo do objeto direto ou indireto; e as do plural, além disso, podem ser também sozinhas o complemento, dado o atributo fórico de ambas e o de designar coletividade de ‘todos’.
- Adjunto adverbial: em sintagmas com essa função, ocorre a maior participação de

‘todo’ como substantivo, em locuções adverbiais. Nesse caso, as formas do singular estão muitas vezes associadas a períodos de tempo, e são adjuntos adnominais – na grande maioria das vezes – e adverbiais; as do plural são adjuntos adnominais – na maioria dos casos –, apostos resumitivos, e ‘todos’ pode ser núcleo nominal.

- Nos outros sintagmas, não há fatos extraordinários.

Quanto ao sentido, é comum a todas as formas ‘conjunto’ e, argumenta-se aqui, ‘reforço’, cuja ocorrência é um tanto rara no plural. As formas do singular assumem também as noções de ‘inteireza’, ‘intensificação’, ‘distribuição’ e ‘partição’ (as duas últimas podem ser consideradas subdivisão de ‘conjunto’). O significado mais comum às formas do singular é ‘reforço’, que abrange cerca de 45% de cada item lexical estudado. Assim, apura-se que ‘tod*’ já não é mais usado, na maioria das vezes, com seu valor primitivo de ‘inteiro’. Não se constata relação direta entre sentido e função sintática. As noções de ‘conjunto’ e ‘inteireza’ podem ser agrupadas sob o rótulo ‘totalidade’; assim, a classificação fica mais simples e acertada, pois nem sempre é fácil distingui-las.

Desse modo, é importante frisar que os itens lexicais estudados diferem em características semânticas, ou seja, quanto à carga de significado que encerram, e a distinção dá-se entre plural e singular, como ilustra o quadro a seguir.

Quadro 6 – Comportamento dos itens lexicais estudados quanto ao significado

Itens lexicais	Acompanha substantivo	Reforça conjunto	Reforça não conjunto	Intensifica adjetivo ou verbo	Valor distr/part	Pronome subst.
Toda	x		x	x	x	
Todo	x		x	x	x	
Todas	x	x				x
Todos	x	x				x
Formas diminutivas	<i>Equivalem aos itens lexicais correspondentes, mas assumem valor de ‘reforço’ mais intenso.</i>					

Observa-se que ‘tod*’, mesmo não estando em posição de adjetivo, pode ser modificado por um advérbio, como ‘quase’, ‘praticamente’ e ‘nem’. Nessa condição,

ao ligar-se a ele, há negação do sentido de ‘totalidade’ que ‘tod*’ encerra. Eis que surge mais um ponto que necessita de amadurecimento nas gramáticas: o caso de advérbios modificarem substantivos e pronomes. As formas diminutivas constantes no *corpus* não são acompanhadas por advérbios.

4.2 RESPOSTAS

Agora já se podem construir respostas para as questões apresentadas no início da análise. Apesar de manterem suas características de adjunto, há distinção no uso dos itens lexicais analisados, como visto, mas a diferenciação está, sobretudo, entre as formas do singular e do plural; e é feita por meio da transposição da função de adjunto, no singular, para aposto, sujeito fórico e afórico, no plural – o último explica a maior incidência do masculino plural e mostra a diferença de função entre as formas plurais, uma vez que somente ‘todos’ é sujeito afórico.

Observou-se que ‘todos’, no sentido de ‘coletividade’, é um recurso vastamente utilizado em textos jornalísticos, o que suscita a hipótese de que, nesse tipo de texto, há muitas generalizações. Ainda, não se pode explicar a vasta ocorrência de ‘todos’ mediante a idéia simplista de que é masculino, pois esse gênero é aparente, haja vista que, a partir dos dados da amostra, percebe-se que o gênero em que é empregada, em grande parte das vezes, é o neutro, o qual, em português, coincide com o masculino, conforme expõe Câmara Jr. (1970, p.88): “o que mais podemos dizer, em referência ao gênero, do ponto de vista semântico, é que o masculino é uma forma geral, não-marcada, e o feminino indica uma especialização qualquer [...]”.

Já a expressiva freqüência de ‘todo’ é devida ao uso da frase feita ‘todo mundo’ – e, em casos mais raros, ‘todo o mundo’ – que é vastamente útil ao usuário da língua para designar coletividade de entes, na maioria das vezes, humanos. Portanto, mais um exemplo de generalização, em que é empregada no gênero neutro e não no masculino.

Finalmente, a escassez das formas diminutivas dos itens analisados deve-se ao seu uso ser coloquial e quase exclusivo da língua falada.

4.3 MODELO DE FUNCIONAMENTO

Agora, podem-se dispor essas informações, ou um resumo delas, em um quadro, para montar o modelo de funcionamento almejado.

Quadro 7 – Modelo de funcionamento de ‘todo’ e ‘toda’

TODO e TODA
CASO 1 – valor pronominal
Comportamento: agregam-se a um núcleo nominal
Ocorrência: sintagmas nominais e adverbiais
Sentido: REFORÇO (sintagma desmembra-se)
Artigo no sintagma: ausente, anteposto ao núcleo do sintagma, posposto à flexão
Exemplos: <ul style="list-style-type: none"> • Isso vai contra toda nossa história e toda nossa tradição. • Luiz Antônio planejou lição de casa politicamente correta para a turma toda. • Em 1967, com uma blusa de gola olímpica laranja e o famoso paletó xadrezinho, já reluzia para todo o País os acordes e versos do hino Alegria, Alegria – “sem lenço, sem documento, nada no bolso e nas mãos”.
Posição dos itens lexicais
Antepostos ou pospostos ao núcleo do sintagma nominal (inclusive na voz passiva sintética)
Exemplos: <ul style="list-style-type: none"> • Joca não perceberia a falta de instrumentos, pois todo ele, no seu corpo triunfal, na sua alegria rara, no impulso da sua alma, vivendo, espraiando-se na velha dançada raça, todo ele era movimento, era vibração, era música. • A grande barbada, lógico, é trabalhar na seção de turismo e viajar o mundo todo de graça. • Construiu-se a casa toda de plástico.
Pospostos a verbos de ligação
Exemplos: <ul style="list-style-type: none"> • O mérito é todo nosso. • A companhia já está toda informatizada.
Antepostos ao particípio da voz passiva analítica (posposição não aparece no <i>corpus</i>)
Exemplos: <ul style="list-style-type: none"> • O salário foi todo pago. • Essa série é toda construída sobre dois níveis.
Pospostos a verbo intransitivo
Exemplos: <ul style="list-style-type: none"> • Ela ajuntou-se todinha
Sentido: CONJUNTO (não se desmembra do sintagma)
Artigo no sintagma: ausente, anteposto ao núcleo do sintagma
Exemplos: <ul style="list-style-type: none"> • E todos, desde meu amigo mineiro Antonio Luís Paixão, sociólogo, até meu companheiro de mesa, versado em magias de toda espécie, o escritor Caio Fernando Abreu, bateram na mesma tecla. • Toda a gente é concorde em afirmar que o Brasil é um país não essencialmente agrícola, como também profundamente contrário à letra redonda. • Queríamos uma eleição com um só candidato e começamos a destruir todo mundo; no final, fomos destruídos também. • Atualmente, todo o mundo reconhece a região do Pacífico Asiático como o coração palpitante da economia.
Posição dos itens lexicais

Antepostos ao núcleo do sintagma

Exemplos:

- **Todo** dia será de sol.
 - Tem um amigo meu que está bêbado **toda** noite.
-

Sentido: **INTEIREZA** (desmembra-se do sintagma)

Artigo no sintagma: anteposto ao núcleo do sintagma, posposto à flexão

Exemplos:

- Fiquei **a** noite **toda** amarrada.
 - No correr de **toda a** infância, o livro que me causou maior impressão e exerceu maior influência foi História de Carlos Magno e os Doze Pares de França.
-

Posição dos itens lexicais

Antepostos ao artigo do núcleo nominal, pospostos ao núcleo nominal, pospostos a verbo de ligação

Exemplos:

- Durante **todo** o dia, as bandeiras azuis e brancas de Israel estiveram a meio-pau.
 - Na lista de clientes estão casais que trabalham o dia **todo** fora de casa
 - Não tenho de que me queixar; a culpa é **toda** minha.
-

Sentido: **QUALQUER** (não se desmembra do sintagma)

Artigo no sintagma: ausente ou anteposto ao núcleo nominal

Exemplos:

- Na verdade, algumas imagens são mais “calculáveis” do que outras, mas **toda** imagem pode ser a priori descrita por um algoritmo.
 - De **toda a** forma, o mínimo que se pode exigir é rapidez e total transparência nas investigações desses dois duplos homicídios.
-

Posição dos itens lexicais

Antepostos ao artigo do núcleo nominal e antepostos ao núcleo nominal

Exemplos:

- É um senhor de 78 anos, “com roupas e o vocabulário de antigamente”, segundo Leone, que está sempre contra **todo** o tipo de inovação.
 - Ela busca emoção, aventura, o que, no fundo é, sim, o sonho de **toda** mulher.
-

Sentido: **DISTRIBUIÇÃO** (não se desmembra do sintagma)

Artigo no sintagma: ausente

Exemplos:

- O problema é que **todo** dia surge um novo problema.
 - No ramo dos restaurantes, a história do engenheiro que virou suco se repete **toda** hora.
-

Posição dos itens lexicais

Antepostos ao núcleo nominal

Exemplos:

- **Todo** final de ano é a mesma coisa: é jacaré tropeçando, caindo, bonitinhas com a cabeça rodando...
-

CASO 2 – valor adverbial

Comportamento: intensificam um adjetivo ou um verbo

Ocorrência: orações com predicativo, sem verbo de ligação, sobretudo; orações com verbos intransitivos

Sentido: **INTENSIFICAÇÃO** (‘muito’, ‘inteiramente’, ‘completamente’, ‘bem’)

Posição dos itens lexicais

Antepostos a um adjetivo

Exemplos:

- Ao chegarem à varanda Ana Rosa, já em trajes de passeio, os esperava para sair **toda** debruçada no parapeito da janela e derramando sobre o Bacanga um olhar mole e cheio de incertezas.
 - Genelício dava o braço à noiva, encasacado numa casaca mal talhada, que punha bem à mostra a sua gibosidade, e caminhava **todo** atrapalhado nos apertados sapatos de verniz.
-

Após o núcleo nominal na voz passiva sintética

Exemplos:

- Construiu-se a casa toda de plástico.
-

Antepostos ao particípio da voz passiva analítica (posposição não aparece no *corpus*)

Exemplos:

- Clodovil disse que a casa foi **toda** revirada e que notou a falta de um aparelho de videocassete e também de algumas jóias.
 - Meu dinheiro é **todo** aplicado da melhor forma possível, visando a um melhor rendimento.
-

Pospostos a verbo intransitivo

Exemplos:

- Ela juntou-se **todinha**.
-

Observe-se que a distinção entre os valores adnominal e adverbial, por vezes, é determinada pelo termo ao qual se entende que 'todo' e 'toda' estão ligados: ao substantivo (adnominal); ao verbo ou ao adjetivo (adverbial).

EXPRESSÕES IDIOMÁTICAS E FRASES FEITAS QUE SE DESTACAM

- A toda, em/por toda (a) parte; toda vez (sempre); ao todo; de todo; de todo modo; em todo caso; a todo (o) vapor; a todo (o) custo; como um todo.
 - Toda (a) gente; todo (o) mundo.
-

Quadro 8 – Modelo de funcionamento de 'todos' e 'todas'

TODOS e TODAS

CASO 1 – valor pronominal

Comportamento: agregam-se a um núcleo nominal ou constituem um núcleo nominal fórico

Ocorrência: sintagmas nominais e adverbiais

Sentido: **REFORÇO** (desmembra-se do sintagma)

Artigo no sintagma: anteposto ao núcleo nominal

Posição dos itens lexicais

Pospostos a verbo, principalmente, de ligação; antepostos ao participio da voz passiva analítica

Exemplos:

- As partidas serão **todas** no ginásio.
- Os criminosos e os pobres eram **todos** arrebanhados nos mesmos locais de confinamento.

Sentido: **CONJUNTO** (não se desmembra do sintagma)

Artigo no sintagma: ausente (principalmente antes de pronome possessivo), anteposto ao núcleo do sintagma, posposto à flexão

Exemplos:

- Meditação, tai chi chuan, estudei todas as coisas... Gosto de pensar que **todas** nossas atividades são espirituais, que tudo que vemos está envolvido por algo que não vemos.
- **Os** jogadores **todos** se encaixam: o Mazinho completa o César Sampaio, o Zinho completa os dois e o Rincón completa os três.
- Se a união configurar-se estável, será como se Josimeire estivesse, para **todos os** efeitos práticos, legalmente casada.

Posição dos itens lexicais

Antepostos ao núcleo do sintagma, antepostos ao artigo e pospostos ao núcleo do sintagma

Exemplos:

- A nova divisão vai apurar **todos** crimes cometidos no estado.
- Ele tem 44% contra 38% da soma de **todos** os candidatos.
- Os outros **todos**, como se houvessem passado palavra, repetiam a mesma cantilena.

Constituindo núcleos nominais

Fórico: assumem função de sujeito ou aposto

Exemplos:

Sujeito:

- **Todos**, no entanto, disseram que a Seleção saberá superar os obstáculos, para conquistar a tão sonhada medalha olímpica.
- **Todas** vencem no final deste mês.

Aposto:

- Almoça na Associação Comercial de Porto Alegre (12h) e visita cidades de Novo Hamburgo, São Leopoldo, Esteio e Canoas, **todas** no RS.
-

-
- E foi assim que a vida das mulheres se povoou de maridos, amantes e noites sem manhãs, **todos** colocados no mesmo plano moral.
-

CASO 2 – valor substantivo (‘todos’)

Comportamento: constitui um núcleo nominal afórico

Ocorrência: sintagmas nominais

Sentido: **CONJUNTO**

Artigo no sintagma: ausente

Posição do item lexical

As mesmas do substantivo

-
- Quércia apenas reafirma o que **todos** já viram: no novo aquário brasileiro ele está mais para peixe fora d’água.
 - Blinda disse que jogar “com um forte calor”, como aconteceu ontem, “é difícil para **todos**, mas creio que nos adaptamos melhor”.
-

CASO 3 – valor adverbial (‘todas’)

Comportamento: intensifica um verbo com sentido intransitivo

Ocorrência: orações com verbos intransitivos

Sentido: **INTENSIFICAÇÃO** (‘muito’, ‘em demasia’)

Posição do item lexical

Posposto a um verbo

Exemplo:

-
- O que ela queria era **aprontar todas** e mais um pouco com a pantera cor-de-rosa.

EXPRESSÕES IDIOMÁTICAS E FRASES FEITAS QUE SE DESTACAM

-
- De uma vez por todas; (estar) em todas; (aprontar, beber) todas; de todos os tempos.
 - Todas as pessoas; todos (indicando coletividade de seres).
-

CONSIDERAÇÕES FINAIS

Ao chegar ao fim desta pesquisa, a maior sensação é a de que estudar o uso é, talvez, o único modo de efetivamente conhecer uma língua, já que, não há como negar, o uso é o que a mantém viva; e, não obstante as críticas que se fazem ao estudo da língua a partir de *corpus*, fica a certeza de que essa abordagem permite vislumbrar uma maneira apropriada e eficaz de coletar, tratar e estudar os usos lingüísticos.

Por meio de uma metodologia que envolveu a lingüística de *corpus*, a lingüística computacional e a estatística e contou com o apoio, principalmente, da gramática normativa para organizar os dados em categorias, pretendia-se formular os padrões de ocorrência do item lexical ‘todo’ e suas flexões em textos escritos, pertencentes à chamada norma culta da língua, e então propor um modelo de funcionamento empírico de tais itens. Na busca desse intuito, as maiores lições que se tiraram do desenvolvimento desta pesquisa são: quando se decide trabalhar com a dinamicidade da língua, deve-se estar pronto para enfrentar fatos extraordinários; muito antes do final, podem-se colher outros achados pelo caminho; e o trabalho com base em um *corpus* volumoso e diversificado pode trazer contribuições valorosas ao estudo da língua portuguesa, sobretudo à escrita.

O modelo criado – uma proposta que tem por fim aprimorar a teoria empírica sobre a língua – cobre boa parte dos fatos lingüísticos que envolvem o item lexical com que se trabalhou, pois se constatou que os itens de trabalho ainda reservam atributos para um estudo maior e mais detalhado. Todavia, foi possível estabelecer padrões de uso mais aprofundados, que complementam as explanações teóricas existentes sobre tal tema lingüístico, dos quais fazem parte casos ainda não tratados, ou equivocadamente tratados, pelos teóricos da língua.

Em certos momentos, os resultados suscitaram reflexões e agitações internas na pesquisadora, sobre temas ligados ao objeto estudado, que não puderam ser

desenvolvidas propriamente, ou mesmo mencionadas, dado o tempo de duração desta pesquisa. Disso, fica a garantia de que o conhecimento sobre milhares de palavras da língua portuguesa seria enriquecido e ampliado a partir de análises que reunissem os usos que as circundam, que as relevasse quando proferidas pelos usuários, respeitando a finalidade natural da língua: a comunicação. Este estudo mostrou que, nessa condição, o pesquisador aproxima-se mais da palavra e pode fazer descobertas surpreendentes, que normalmente vão além de concepções prévias, despedaçando expectativas, fugindo a padrões preestabelecidos e considerados completos.

Como exemplo disso, evidenciou-se que a coletividade que utiliza a língua diversifica o seu uso, de forma que muito do que seria considerado inaceitável por alguns pode ser, além de usual para outros, formas de expressão cristalizadas na língua desde longa data – como ocorreu na investigação do artigo sucedendo os itens lexicais de trabalho, tanto no plural como no singular. Então, uma análise deve considerar o maior número possível de usos, sob risco de chegar a conclusões que não se sustentam perante usuários de todos os níveis de letramento – nem mesmo perante níveis selecionados para estudo, como o que foi escolhido para ser objeto deste trabalho: a norma padrão escrita – ou de discriminar usos.

Classificar os dados, em algumas etapas, revelou-se uma tarefa árdua e que implica julgamentos que poderiam aventar polêmicas. Todavia, entende-se que isso faz parte do trabalho empírico e não há teoria lingüística isenta de controvérsias. Nesse ponto, referendou-se ser essencial adotar uma teoria para servir de apoio na classificação dos dados que seja dominada pelo pesquisador, pois isso facilita o trabalho, especialmente a tomada de decisão, e robustece as discussões. Em consequência, as próprias teorias adotadas são debatidas e assim podem ser aprimoradas.

A estatística – aqui configurada na forma que se empregou para selecionar os dados de trabalho; na aplicação, por duas vezes, de um teste de hipóteses; na criação de um modelo probabilístico; e em várias ocasiões em que se utilizaram percentuais, que, apesar de serem cálculos simples, mostraram-se uma ótima ferramenta de pesquisa – constitui um exemplo real de como a matemática pode ser aplicada no estudo empírico da língua, mediante a focalização de dados, comparações,

comprovações e a oferta de um diferente ângulo de visão, uma outra perspectiva. Infelizmente, faltam publicações que auxiliem os pesquisadores a utilizar essa ferramenta especificamente no estudo da língua, sobretudo portuguesa, e permitam-lhes enriquecer seus conhecimentos nesse ramo – o conhecimento lingüístico seria o maior beneficiado com o desenvolvimento dessa área. Certamente, é um investimento que merece ser feito. Neste trabalho, não se fez jus ao seu potencial, uma vez que sua participação foi pequena, porém não pouco importante.

As outras duas partes da metodologia adotada, a lingüística computacional e a lingüística de *corpus* mostraram-se, como esperado, eficientes e dinâmicas na pesquisa da língua. Por meio delas, a estatística pôde dar sua contribuição e questões lingüísticas suscitaram, constituindo tema para investigações, que somente puderam ser realizadas devido à agilidade do computador em processar informações. Integrada ao meio eletrônico, a rede mundial mostrou-se um *corpus*, pois se encaixa, com poucas ressalvas, nas principais recomendações para que uma massa lingüística seja adequada para pesquisa, como: os textos não devem ter sido criados para fins de pesquisa e devem tratar de temas variados. Dinamizar e sistematizar a rede mundial como fonte de pesquisa lingüística é uma sugestão da pesquisadora para trabalhos futuros. Por meio do *corpus* diacrônico, verificou-se que fatos aparentemente novos podem ser mais antigos do que se imagina e demonstrou-se que a ordem de ocorrência dos itens lexicais estudados, no que se refere a proporção, mantém-se constante há muito tempo, no mínimo, desde a época do império no Brasil.

Em conclusão, considera-se que o estudo foi frutífero em vários aspectos, seja nas discussões apresentadas com o objetivo de incentivar reflexões, seja na metodologia adotada – isto é, o tratamento dado aos itens lexicais investigados –, que pode servir de base para estudos vindouros, e no próprio modelo construído, que pode ser aprimorado a partir de outras pesquisas. Além disso, e sobretudo, porque demonstrou que, no uso real, as palavras escapam às definições estabelecidas e isso fornece margem para discrepâncias entre os teóricos, o que indica que argumentar e convidar à reflexão é necessário, para que, dentre outros motivos, no ensino de língua portuguesa, possa valorizar-se mais a intuição e o conhecimento lingüístico do aluno, sem esperar que ele repita os dizeres da teoria adotada pelo professor, mas lhe

permitindo deduções – como, por exemplo, que tem cabimento em dizer que ‘todo’ e ‘toda’ podem conectar-se a um verbo ou a um adjetivo, até mesmo a um substantivo, em certos casos, assumindo valor adverbial, apesar de variarem em gênero – e, amparado pelo respeito à suas concepções lingüísticas, compreender as relações inter-
-termos da versão escrita e padrão, neste caso, da sua língua, a fim de manuseá-la propriamente para integrar-se à sociedade.

REFERÊNCIAS

- ALMEIDA, N. M. de. **Gramática metódica da língua portuguesa**. 44. ed. São Paulo: Saraiva, 1999.
- BARBETTA, P. A. **Estatística aplicada às ciências sociais**. 2. ed. Florianópolis: Ed. da UFSC, 1998.
- BECHARA, E. **Lições de português – pela análise sintática**. 3. ed. Rio de Janeiro: Fundo de Cultura, 1964.
- BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus Linguistics: investigating language structure and use**. Cambridge University Press, Cambridge, 1998.
- CAMARA JR. J. M. **Estrutura da língua portuguesa**. 16. ed. Petrópolis: Vozes, 1970.
- CASCATINHA e INHANA (intérpretes). **Flor do Cafezal**. Disponível em: <www.kazaa.com>. Acesso em: 18 jun. 2003.
- DUARTE, V. L. V. **Ambigüidade lexical e sua interferência na interpretação do discurso: estudo com o verbo arrasar**. Em: 5º ENCONTRO DO CELSUL, 2002, Curitiba: UFPR, CD-ROM. No prelo.
- FERREIRA, A. B. de H. **Novo Aurélio Século XXI: o dicionário da língua portuguesa**. 3. ed. Rio de Janeiro: Nova Fronteira, 1999.
- GRÉGGIO, S.; MERKLE, C.; DUARTE, V. L. **Transcrição de ‘Ofícios do Conselho Supremo Militar ao Governo da Capitania/Presidente da Província – 1821 a 1827’**. Florianópolis, 1999. Não publicado. O original pertence ao acervo do Arquivo Público de Santa Catarina.
- HOUAISS, A.; VILLAR, M. de S.; FRANCO, F. M de M. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2001.
- LEECH, G. Corpora and theories of linguistics performance. In: Jan Svartvik (Org.). **Directions in corpus linguistics**, Mouton de Gruyter, Berlim, 1992.
- MANNING, C. D. e SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. The MIT Press, Cambridge, 2000.
- McENERY T. e WILSON A. **Corpus linguistics**. Edinburg University Press, Edinburg, 1997.
- NASCENTES, A. **Dicionário etimológico resumido**. Rio de Janeiro: Instituto Nacional do Livro/MEC, 1966.
- NEVES, M. H. de. **Gramática de usos do português**. São Paulo: Unesp, 2000.
- NEVES, M. H. de. **Guia de uso do português: confrontando regras e usos**. São Paulo: Unesp, 2003.
- NILC – Núcleo Interinstitucional de Lingüística Computacional (Org.). **Corpus lingüístico**. Disponível em: <<http://www.nilc.icmc.usp.br/nilc/tools/corpora.htm>>. Acesso em: 9 jan. 2002.

- NILC – Núcleo Interinstitucional de Lingüística Computacional. **Mini Gramática: sintagma**. Disponível em: <<http://www.nilc.icmc.usp.br/minigramatica/mini/sintagma.htm>>. Acesso em: 24 mar. 2003.
- PEAD (Português – Ensino a Distância). **O sintagma nominal**. Disponível em: <<http://acd.ufrj.br/~pead/tema15/por-tm15.html#ponto7>>. Acesso em: 24 mar. 2003.
- ROCHA, M. Relações anafóricas no português falado: uma abordagem baseada em *corpus*. **Revista D.E.L.T.A.**, São Paulo, v. 16, n. 2, 2000. p. 229-261.
- ROCHA, M. O corpus computadorizado em lexicografia: guia do consumidor. **A construção de dicionários e léxicos a partir de corpus** (Mesa redonda). Marco Rocha e Maria Tereza Biderman. Coordenação: Clotilde Murakawa. In: SEMINÁRIO DE PESQUISA, PERFILANDO POLÍTICAS E PROJETOS. Universidade Estadual Paulista, Faculdade de Ciências e Letras, Araraquara, 21 nov. 2001.
- SAID ALI, M. **Gramática secundária e histórica da língua portuguesa**. 3. ed. Brasília: Editora UnB, 1964.
- SARDINHA, T. B. Lingüística de corpus: histórico e problemática. **Revista D.E.L.T.A.**, São Paulo, v. 16, n. 2, 2000. p. 323-367.
- SINCLAIR, J. **Corpus, concordance, collocation**. Oxford University Press, Oxford, 1991.
- WALSH, A. **Statistics for the social sciences: with computer applications**. Boise State University, Harper & Row, publishers, New York, 1990.

APÊNDICE – Fórmulas e Tabela do Teste Qui-Quadrado

FÓRMULAS E TABELA DO TESTE QUI-QUADRADO

a) Freqüências esperadas (E):

$$E = \frac{(\text{total da linha}) \times (\text{total da coluna})}{(\text{total geral})}$$

(Fórmula aplicada a cada célula da tabela.)

b) Estatística (ou contribuições) do qui-quadrado

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Na qual se tem:

O = Freqüência observada

E = Freqüência esperada

c) Fragmento da Tabela da Distribuição Qui-quadrado

<i>gl</i>	Área da cauda superior						
	0,250	0,100	0,050	0,025	0,010	0,005	0,001
1	1,32	2,71	3,84	5,02	6,63	7,88	10,8
2	2,77	4,61	5,99	7,38	9,21	10,6	13,8
...