

FRANCIELLE ALVES VARGAS

**ENSAIO DE LEXICOLOGIA SÓCIO-HISTÓRICA
COM RECURSOS DA LINGUÍSTICA DE CORPUS
E DA LINGUÍSTICA COMPUTACIONAL**

Trabalho de conclusão de curso apresentado
ao programa de graduação da Faculdade de
Letras da Universidade Federal de Minas
Gerais.

Áreas de concentração:
Filologia Românica

Orientador:
Prof. Dr. César Nardelli Cambraia

Belo Horizonte
2014

E o que sinto em mim, quando diante do computador busco a essência do homem, a essência profunda, do animal e da pedra, que me permitirá escrevê-los, o que sinto, intensamente, é que eu a procuro dentro de mim, através de mim, através da minha própria, mais profunda, essência. E que essa é, antes de mais nada, uma essência de mulher.

Marina Colasanti

AGRADECIMENTOS

Gostaria de agradecer, primeiramente a Deus por estar sempre comigo. Em seguida, agradecer meu orientador e professor César Nardelli por seu um profissional e mestre brilhante e por ter me apresentado ao universo da linguagem. Aos meus pais e familiares, que mesmo longe, torceram por mim. Aos meus amig@s pela cumplicidade e, aos/as meus/minhas coleg@s de trabalho.

RESUMO

Neste trabalho analisou-se o termo *vadia* para identificar mudanças semânticas ao longo dos séculos XIX a XXI. Testou-se a hipótese de que houve mudança semântica no termo *vadia* motivada por mudanças sociais. Para testar a hipótese, foi construído um *corpus* que reuniu todos os registros do termo no acervo digital do *Jornal do Brasil*, que cobriu o período de 1891 a 2010. A análise dos dados empíricos, em consonância com a lexicologia sócio-histórica, foi feita através do uso de recursos da linguística de *corpus* e da linguística computacional. A hipótese foi confirmada com base nos dados, que revelaram que a cada ciclo de lutas sociais das mulheres o termo em estudo ganhou novos sentidos, sendo associado a palavras de diferentes campos semânticos. Foi ainda proposto neste trabalho, um modelo teórico para a construção de lematizadores automáticos.

Palavras-chave: Lexicologia; Linguística de *Corpus*; Linguística Computacional.

ABSTRACT

In this work we analyzed the word *vadia* to identify semantic changes throughout the 19 to 21 centuries. We tested the hypothesis that semantic change in the word *vadia* was motivated by social changes. To test the hypothesis, we built a *corpus* with all the records of the word in the digital archive of *Jornal do Brasil*, which covered the period from 1891 to 2010. The analysis of empirical data, in line with the socio-historical lexicology, was taken through the use of resources of *corpus* linguistics and computational linguistics. The hypothesis was confirmed based on the data, which revealed that in each cycle of social struggles of women the word *vadia* gained new meanings, being associated with words from different semantic fields. It was also proposed in this paper a theoretical model for building automatic stemmers.

Keywords: Lexicology; Corpus Linguistics; Computational Linguistics.

LISTA DE FIGURAS

Figura 1. Representação de uma hierarquia do LM, agrupadas por tipos (DI FELIPPO; DIAS-DA-SILVA, 2006, p. 5)	27
Figura 2. Algoritmo que descreve o pseudocódigo da extração de termos com o uso de regras (DUQUE, 2012, p. 75)	27
Figura 3. Algoritmo que descreve o pseudocódigo da extração de termos com o uso de regras e dicionário (DUQUE, 2012, p.76).....	27
Figura 4. Ocorrências do termo vadia no acervo do JB na HDB	32
Figura 5. Textos com ocorrência do termo vadia enquadrados versus não-enquadrados no critério de seleção.....	33
Figura 6. Exemplo de um texto do jornal usado para transcrição	34
Figura 7. Paradigma derivacional nominal do PB.....	38
Figura 8. Paradigma flexional verbal do PB	39
Figura 9. Fases de desenvolvimento da aplicação proposta	40
Figura 10. Arquitetura básica para aplicações de PLN (DIAS-DA-SILVA, 1996, p. 234)	41
Figura 11. Representação gráfica das tabelas para a formação da BDL.....	42
Figura 12. Total de ocorrências de vadia em intervalos de sessenta anos.....	46
Figura 13. Total de ocorrências de vadia em intervalos de dez anos	46
Figura 14. Total de ocorrências de vadia em intervalos de cinco anos.....	47
Figura 15. Cinquenta lexemas mais frequentes no corpus	51
Figura 16. Intercessão dos grupos A e B dos 10 lexemas mais frequentes	54
Figura 17. Contextos de usa do lexema OLHO (1891 – 1950)	55
Figura 18. Contextos de usa do lexema MÃE (1891 – 1950)	56
Figura 19. Contextos de usa do lexema RUA (1891 – 1950)	57
Figura 20. Contextos de usa do lexema NOITE (1891 – 1950)	58
Figura 21. Contextos de usa do lexema POBRE (1891 – 1950).....	59
Figura 22. Contextos de usa do lexema TEMPO (1891 – 1950)	60
Figura 23. Contextos de usa do lexema FILME (1951 – 2010).....	61
Figura 24. Contextos de usa do lexema HISTÓRIA (1951 – 2010)	62
Figura 25. Contextos de usa do lexema VIDA (1951 – 2010).....	63
Figura 26. Contextos de usa do lexema HOMEM (1951 – 2010)	64
Figura 27. Contextos de usa do lexema MUNDO (1951 – 2010).....	65
Figura 28. Contextos de usa do lexema COISA (1951 – 2010)	66
Figura 29. Contextos de uso de vadia no primeiro período (1891–1950)	68
Figura 30. de uso de vadia no segundo período (1951 – 2010)	71
Figura 31. Estrutura básica de um sistema especialista.....	74

LISTA DE TABELAS

Tabela 1. Total de textos com ocorrência do termo <i>vadia</i>	33
Tabela 2. Textos com mais de uma ocorrência do termo <i>vadia</i>	34
Tabela 3. Parâmetros para o algoritmo de processamento	43
Tabela 4. Número de <i>tokens</i> e <i>types</i> no <i>corpus</i>	45
Tabela 5. Campos semânticos dos cinquenta lexemas mais frequentes no <i>corpus</i>	51
Tabela 6. Lexemas por recorte temporal de sessenta anos	53

LISTA DE SIGLAS

BDL: Base de Dados Lexicais

GRM: Gramaticais

HDB: HEMEROTECA DIGITAL BRASILEIRA

IA: Inteligência Artificial

JB: Jornal do Brasil

LCM: Lexias complexas

LM: Léxico Mental

NON: Não-Verbais

PB: Português Brasileiro

PLN: Processamento de linguagem natural

VIR: Verbos Irregulares

VRE: Verbos Regulares

LISTA DE ANEXOS

Anexo I : Lista das 1400 lexias mais frequentes no *corpus*.....80

Anexo II: Lista de lexemas a partir das 1400 lexias mais frequentes no *corpus*....96

SUMÁRIO

Agradecimentos.....	3
Resumo	4
Abstract.....	5
Lista de figuras	6
Lista de tabelas.....	7
Lista de siglas	8
Lista de anexos.....	9
1.0 Introdução	12
2.0 Fundamentação teórica	13
2.1. Lexicologia	15
2.1.1 Conceitos fundamentais.....	15
2.1.2. Lexicologia social	17
2.1.3. Lexicologia sócio-histórica.....	20
2.2 Linguística de <i>corpus</i>	22
2.3 Linguística computacional.....	23
2.3.1 Processamento de línguas naturais (PLN).....	23
2.3.2 Algoritmos	25
2.4. Designações de gênero	29
3.0 Hipótese de trabalho e objetivos.....	31
3.1. Hipótese de trabalho	31
3.2. Objetivos	31
3.2.1 Geral.....	31
3.2.2. Específicos	31
4.0 Metodologia	32
4.1 O <i>corpus</i>	32
4.2 Lematização do <i>corpus</i>	36
5.0 Lematização automática.....	36
5.1 Modelo ideal para lematização automática	40
5.1.1 Fase I: Base de dados lexical (BDL).....	40
5.1.2 Fase II: Algoritmo para tratar cadeia de caracter.....	42
6.0 Análise dos dados.....	45
6.1 Frequência	46

6.2 Co-ocorrências	50
6.2.1 Geral.....	50
6.2.2 Por recorte histórico de sessenta anos	53
7.0 Considerações finais.....	73
Referências Bibliográficas	75

1.0 INTRODUÇÃO

As palavras são, de certa forma, uma maneira de medir a concepção de mundo dos falantes a partir das representações que criam desde a infância. Essas representações são uma poderosa ferramenta para construção do conhecimento e possibilita ver o mundo de perspectivas diversas.

As representações expressas pelas palavras são construídas ao longo da história de vida dos falantes e sofrem modificações. Essas representações que constituem os conceitos veiculados pelas palavras são passíveis de mudança. Elas são construídas e desconstruídas na mente dos falantes na medida em que se apropriam de novos conhecimentos. Mudam as percepções sobre o objeto e, consequentemente, muda o conceito que o representa.

As estruturas sociais que são representadas no léxico das línguas não são estáticas. Elas mudam e se reconfiguram no tempo e no espaço. Grupos sociais em constantes mudanças, reformas políticas, sociais, econômicas, apropriação e negação de direitos, todos estes itens desencadeiam mudanças sociais e, consequentemente, mudanças lexicais.

Grupos sociais que vem alcançando significativas mudanças sociais, ao longo da história, são os grupos feministas representados por mulheres que lutam contra a herança sócio-histórica de negação de direitos.

A partir do conceito de sociedade patriarcal, criou-se uma representação da mulher enquanto coadjuvante na vida política e social. Para esta, devido a sua possibilidade biológica de reprodução, foram-lhe atribuídos apenas papéis sociais distantes da vida pública, ou seja, dentro dos lares, cuidando de suas famílias e filhos. Criou-se então uma expectativa de que a mulher ocuparia apenas papéis sociais no âmbito privado, sem que essa pudesse atuar ativamente na vida política e social de sua comunidade, papel este reservado ao outro gênero. Talvez seja o conceito de sociedade patriarcal o dado inicial do hipermonitoramento da sexualidade da mulher, dado que o homem, para se certificar que seus herdeiros não fossem bastardos, usava-se da presença do hímen para comprovação da virgindade da mulher e, assim, da legitimidade dos seus herdeiros, ou seja, a castidade era uma moeda de grande valor. Associados a castidade, fatores como pouca visibilidade pública e passividade eram características grandemente valorizadas.

Na medida em que as mulheres romperam com as expectativas sociais e reivindicaram seus direitos historicamente negados, gradativamente elas avançaram nos espaços políticos, apropriando-se de papéis sociais antes ocupados unicamente por homens. A apropriação de direitos políticos, econômicos, morais e sociais parece ter ressignificado os papéis sociais, alterando os eixos referenciais dos gêneros e modificando a forma que a sociedade via essa mulher. Na medida em que a mulher se apropriou de direitos, até então, a ela negados, novos comportamentos surgiram, comportamentos estes que romperam com paradigmas morais. Um exemplo é a mudança na forma de que as mulheres lidavam com sua sexualidade, que foi oprimida historicamente por vários fatores. As mulheres romperam com interdições e construções arcaicas no que diz respeito a sua sexualidade.

Um bom exemplo para representar os conflitos dessa ruptura são os termos associados a mulher. O termo *vadia*, por exemplo, atribuído às mulheres, na contemporaneidade diz respeito àquela mulher cujo(s) comportamento(s) fere(m) alguma norma de natureza moral ou social. Uma questão que se coloca é se a história desse termo refletiria as mudanças sociais que ocorreram nas décadas passadas. Para responder a esta questão, apresenta-se aqui um estudo do termo *vadia* sob a perspectiva da lexicologia sócio-histórica.

2.0 FUNDAMENTAÇÃO TEÓRICA

O léxico de uma língua representa o tesouro simbólico de uma geração, seu patrimônio cultural. Para Biderman (1996, p.44), a herança cultural é passada às novas gerações através da linguagem. A língua é o veículo por excelência da transmissão da cultura. É o léxico da língua que se constitui de um tesouro de signos linguísticos que, em forma de código semiótico, permite esse milagre. O léxico funciona como uma espécie de representação social da realidade. Segundo Matoré (1973, p. 39) o léxico é instrumento de compreensão social. Através do léxico, é possível inferir sobre representações sociais.

No desenvolvimento dos estudos do léxico, algumas metodologias e ferramentas específicas têm sido utilizadas. Uma metodologia empírista de grande valia neste campo de estudos é a linguística de *corpus*. Segundo Léon (2007), a oferta de grandes corpora e a possibilidade de tratamento de enormes volumes de dados linguísticos foram a origem, nos anos 1990, de uma renovação do interesse pelo

tratamento estatístico e probabilístico. Esse interesse adquiriu uma importância crescente e tornou-se atualmente um campo de pesquisa dominante nas ciências da linguagem (Léon, 2007, p.52).

A metodologia empírica consolidou-se com o advento das novas tecnologias, que possibilitaram a criação de grandes bases digitais textuais, além de resultados muito mais eficazes no tratamento desses dados. As possibilidades tecnológicas tem sido a chave-mestra para a otimização do tratamento desses dados. Várias aplicações estão sendo criadas, a fim de garantir ao pesquisador maior precisão nas análises, além da ampliação de possibilidades de pesquisa.

Atualmente, é possível encontrar algumas ferramentas de manipulação de corpora que geram dados importantes para as análises linguísticas; todavia, essas ferramentas lidam apenas com o número de lexias (formas concretas) presentes no *corpus*¹. Para estudos no campo da lexicologia, é imprescindível que, além do número de lexias, se obtenha também o número de lexemas (paradigma a que se vinculam as formas concretas). Neste trabalho, ensaiaremos o processo automático de quantificação de lexemas. Proporemos um modelo que, a partir de uma lista de lexias, se possa gerar uma lista de lexemas contabilizando sua frequência. Essa proposta só foi possível, a partir da evolução dos estudos da linguística computacional.

A linguística computacional tem como objetivo estabelecer um diálogo entre as ciências da linguagem e as ciências computacionais. É a partir desse diálogo que se tem estabelecido modelos computacionais para o processamento de línguas naturais (PLN) e, consequentemente, para o tratamento do léxico. Segundo Ranchhod (2001, p. 13-47), nos últimos anos, tornou-se evidente que os elementos linguísticos e, em particular os elementos lexicais, são a pedra de toque para qualquer sistema de processamento de linguagem natural. O léxico é, sem sombra de dúvidas, um componente importante para a PLN. Todavia, a complexidade morfossintática e a polivalência das unidades lexicais tornam-se dificultadores quando se trata de processamento de línguas naturais.

A propósito do diálogo entre a linguística computacional e o léxico, Biderman (2001, p. 90-91) diz:

¹O número de lexias equivale também ao número de *types* do *corpus*. Segundo Bybee (2001), *token* (*frequency*) é a ocorrência de cada item lexical em determinado *corpus* e *type* (*frequency*) é a frequência de determinado padrão na língua.

[n]os primeiros tempos o léxico foi tido como um gargalo no desenvolvimento da linguística computacional por causa da magnitude do patrimônio vocabular que constitui o acervo herdado de uma língua moderna de civilização, além do obstáculo adicional criado pelo volume incessante de neologismos que se incorporam continuamente no léxico. Entretanto, e isso é essencial nas teorias sobre a linguagem, o léxico constitui um repositório de informações sobre o mundo, uma vez que o conhecimento está condensado em palavras.

Vejamos, a seguir, questões relacionadas a cada um desses três campos do conhecimento necessário à execução da pesquisa que ora se apresenta: lexicologia, linguística de *corpus* e linguística computacional.

2.1 LEXICOLOGIA

2.1.1 CONCEITOS FUNDAMENTAIS

A lexicologia é uma ciência relativamente jovem. Os estudos do léxico durante muito tempo foram deixados de lado. Por algum tempo, muito pouco se fazia para analisar as palavras de uma língua. Elas eram organizadas alfabeticamente e coletadas a partir de obras literárias. A partir do século XIX, com o florescimento da onomasiologia e o *VII Congresso Internacional de Linguística* em 1952 em Londres, os conceitos linguísticos foram elaborados sobre a base fenomenológica para a elaboração de um sistema extralingüístico de referências (ABBADE, 2011).

O objeto da lexicologia é o léxico. Segundo Di Felippo& Dias-da-Silva (2006), o léxico é concebido, segundo os moldes gramaticais modernos, como uma rede de correspondência em vários níveis; e o item lexical, como uma trilha de propriedades fonológicas, morfológicas, sintáticas e semânticas. No léxico de uma língua é importante fazer distinções como lexias e lexemas.

Segundo Biderman (1978, p.130 – 4) a designação lexema trata-se de a “unidade léxica abstrata em uma língua”. Como exemplo ela cita o lexema MENINO, grafado com maiúsculo para demonstrar um processo abstrato que pode variar passando para o plural e adquirir a forma “meninos”, ou permanecer no singular, “menino”. Essas formas foram conceituadas por Biderman com o nome de lexias.

Para os estudos do léxico é necessária a classificação de lexias e de lexemas. Essa classificação envolve o processo de lematização, processo em que se agrupam

distintas lexias em seu lexema correspondente. Segundo Cúrsio (2006, p.5), a lematização é um processo que atua como uma espécie de filtro, deixando o número total de formas de uma maneira que evite a sua repetição, como, por exemplo, as diferentes conjugações de um mesmo verbo ou as flexões que indicam número e gênero em substantivos e adjetivos. A lematização é uma maneira de garimpar o texto, de modo que as formas sejam contabilizadas mais estritamente.

Saussure, ao introduzir os conceitos de signo linguístico e de sistema, lançou, naquele momento, as primeiras noções para um modelo de léxico mental (LM). Segundo Aitchison (1987), o LM é comparável ao mapa de metrô de Londres, com elementos conectados ponto a ponto:

Estamos tentando produzir um diagrama das conexões no léxico mental, que é, em alguns aspectos, comparável ao mapa do metrô de Londres. No entanto, existe um modo em que esse mapa mental é bastante diferente. Nós podemos ir para dentro do subterrâneo e mapear as conexões entre as estações. Mas não podemos ver as conexões diretamente na mente (AITCHISON, 1987, p.30)²

Segundo Bierwisch e Schreuder (1992), o LM diz respeito à parte do conhecimento lexical do falante determinado pela estrutura da língua. Ainda segundo os autores, o LM é elemento fundamental no processamento cognitivo da linguagem, o qual envolve três tipos de processos: (i) conceitualização (especificação de conceitos); (ii) formulação (seleção de itens) e (iii) construção de representações sintáticas.

Segundo Di Felippo & Dias-da-Silva (2006, p. 13), conforme a abordagem de cada campo de conhecimento, no léxico se constroem várias conceituações:

Ressalta-se que a principal diferença dos olhares dessas três áreas reside na concepção ou definição do objeto “léxico”. Sob o olhar da Psicolinguística, trata-se do léxico mental, ou seja, a parte do conhecimento lexical do indivíduo delimitada por sua língua, sob o ponto de vista da linguística, trata-se de um dos componentes da gramática e, sob o ponto de vista do PLN, trata-se de um “megarquivo”.

² No original: “We are trying, then, to produce a diagram of the connections in the mental lexicon which is in some respects comparable to a plan of the London Underground. However, there is one way in which this mental map is quite different. We can go down into the Underground and map the connections between stations. But we cannot view the connections in the mind directly.”

A utilização de hierarquias de tipos imprime ao léxico, organização global e tem por objetivo reduzir parte da informação redundante no léxico (FLICKINGER, POLLARD E WASOW, 1985; KOENIG, 1999).

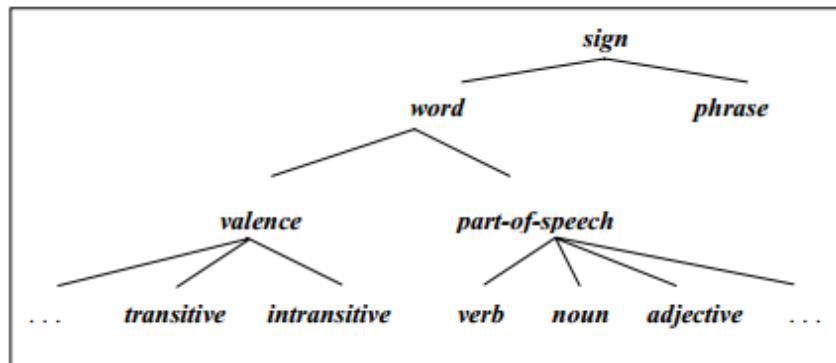


Figura 1. Representação de uma hierarquia do LM, agrupadas por tipos
(DI FELIPPO; DIAS-DA-SILVA, 2006, p. 5)

O LM de um falante contém um número significativo de elementos resultando uma complexa “teia” combinatória. Para Mel’čuk, (1988), dada essa complexidade, pressupõe-se que os itens lexicais estejam organizados de maneira funcional, para que o falante possa recuperar rapidamente não só o significado de um item, mas também todas as suas características gramaticais e usos.

Estudos mais aprofundados sobre o LM podem contribuir grandemente para o desenvolvimento da linguística computacional na tentativa de se criar um modelo computacional-lexical que represente a forma que processamos o léxico em nossa mente. Segundo Allen (1987), teoricamente as arquiteturas propostas para sistemas de PLN acabam por espelhar a arquitetura proposta para o sistema linguístico.

2.1.2. LEXICOLOGIA SOCIAL

Uma das abordagens de estudo do léxico é a lexicologia social. Não há registros claros sobre o inicio da utilização deste termo. Segundo Cambraia (2013, p. 159),

[a] falta de informação mais segura sobre a primeira ocorrência da expressão lexicologia social no campo dos estudos linguísticos impede que se afirme que Georges Matoré (1908-1998) tenha sido seu criador, mas pode-se dizer seguramente que foi ele quem a colocou em evidência desde seu artigo seminal “La lexicologie sociale” publicado em periódico *L’information Littéraire* em 1949.

Estudos de lexicologia social são de fundamental importância para os estudos linguísticos pela sua interdisciplinaridade, dialogando com as outras ciências. Segundo Nunes (2006, p.149-150), a lexicologia social é disciplina imprescindível na linguística, por suas afinidades com a gramática das formas (morfologia) e com a semântica e por ser um ponto de encontro nas ciências humanas.

É o diálogo entre a lexicologia social e outras áreas de conhecimento que possibilita categorizar o léxico, suas unidades semânticas e as relações com o contexto social. Foi a partir do pressuposto de que cada item lexical envolve uma rede de traços semânticos e de que, por isso, integra um determinado campo de significação, que, segundo Lyons (1977) *apud* Gritti (2010, p.138), alguns linguistas das décadas de 1920 e 1930, dentre eles Ipsen (1934), Porzig (1934) e Trier (1931), criaram a teoria dos campos lexicais. Essas teorias tem afinidade com o postulado de Saussure de língua enquanto sistema, princípio básico da teoria dos campos lexicais.

Segundo Trier (1931) *apud* Lyons (1977, p. 253), “os campos semânticos são realidades vivas intermediárias entre palavras (com status isolados) e totalidade de vocabulário”. Os campos semânticos são, portanto, estruturas subjacentes de representação ideológica pelas palavras. São estes campos que, imersos numa teia de significação, revelam um comportamento político. Segundo Robin (1977, p. 47), não são as palavras, mas a repartição das palavras no texto que implica um modelo ideológico.

O léxico de uma língua pode dizer muito sobre aspectos sociais e ideológicos. Segundo Matoré (1973, p.40), a criação de uma palavra equivale à criação de um conceito “no dia que a palavra árvore aparece, nasce o conceito”. Essa representação conceitual do mundo e a concretização desses saberes sobre das coisas convergem em uma construção da realidade. Quando nomeamos as coisas através de um sistema simbólico estamos representando nosso mundo social. Essa construção da realidade é chamada por Bourdieu (1989, p. 9) de *poder simbólico* - um poder de construção da realidade que tende a estabelecer uma ordem gnoseológica: o sentido imediato do mundo (e, em particular do mundo social).

As palavras atuariam, portanto, como uma espécie de termômetro social. A partir do léxico utilizado por cada grupo social, é possível identificar diferenças entre esses grupos. Essa identidade entre léxico e comportamento de um grupo

sociopolítico, subjacente na obra de Matoré, é o que Robin (1977, p. 44) chama de *isomorfismo*:

Para tanto [Matoré] formula uma nova hipótese: a de um isomorfismo entre os grupos sociopolíticos e o léxico. Cada grupo político seria assinalado pelo uso específico que faria de certas palavras. Colocadas e conhecidas as atitudes políticas, espera-se, deseja-se que lhes correspondam clivagens linguísticas – mais exatamente, léxicas – pertinentes.

O vocabulário seria assim uma espécie de etiqueta fixada sobre cada grupo político, na qual se revela uma visão do mundo, uma linguagem específica. Portanto, não seria necessário conhecer a posição política deste ou daquele grupo: seu vocabulário bastaria para defini-lo, para designar-lhe um lugar específico no tabuleiro político (ROBIN, 1977).

Robin (1977), no entanto, adverte que a hipótese do isomorfismo entre léxico e grupo sociopolítico não pode ser aceita sem reservas, já que um falante pode “mascarar” algumas das características lexicais do seu grupo, a fim de “parecer” integrante de um outro grupo social:

Convém concluir provisoriamente que, no discurso, o locutor se dá tal como quer aparecer, ou tal como se vê em função de seus sistemas de representações, e não diretamente tal como sua prática política e como a significação, a função social de sua prática política o define. (ROBIN, 1977, p.47)

A lexicologia social de Matoré baseia-se fundamentalmente no estudo de neologismos. Como assinala Alves (1984, p. 119), “neologismo constitui uma unidade lexical de criação recente, uma acepção nova que se atribui a uma palavra já existente, ou, então um termo recentemente emprestado a outro código linguístico.”. Os neologismos podem ser de três tipos: o formal, o semântico e por empréstimo. Segundo a autora, neologismos formais são criados por meio de derivação, composição, siglas, redução de palavras ou pela articulação de uma ou diversas sílabas que possuem um valor significante inédito; o semântico, pela atribuição de novo significado a um mesmo segmento fonológico; e, por fim, o por empréstimo, pela adoção de um lexema estrangeiro.

2.1.3. LEXICOLOGIA SÓCIO-HISTÓRICA

Para Guariglia (2006, p. 1696), “a historicidade faz existir novas práticas discursivas: um determinado momento histórico provoca a organização de enunciados referentes às formações discursivas que coexistem em uma sociedade”.

Da mesma forma que uma sociedade se transforma, o léxico de uma língua também se modifica, agregando valores novos e reconfigurando-se. Para Bakhtin (1988, p. 17), o léxico de uma língua veicula, de maneira privilegiada, a ideologia; e a ideologia, para ele, é uma superestrutura em que as transformações sociais de base se refletem – logo se refletem também na língua que as veicula. Para esse autor, a palavra serve como “indicador” das mudanças. Conforme os falantes se interagem e se modificam em torno dos processos sociais, as mudanças que acontecem em decorrência deste processo refletem no léxico da língua dos falantes. Quando esse processo se materializa na língua, temos o que chamamos *neologismo*. Guariglia (2006, p. 1697) usa o seguinte exemplo para representar neologismo motivado por fatores socio-históricos:

o empresariado deu preferência à palavra *funcionário*, referindo-se ao trabalhador assalariado, em substituição a *empregado*; mais recente, há tendência para utilização de *colaborador*. À medida que se admitiu a importância funcional do trabalhador para o sucesso de um empreendimento empresarial, passou-se a entendê-lo como participante ativo do processo. Aliás, as próprias empresas, outrora nomeadas *firmas*, hoje são preferencialmente chamadas de *organizações*, e já há quem utilize *gestoras*. Esses dois termos mais recentes destacam a importância da interação entre os diversos departamentos, como um organismo; sob uma gestão que faça compor um todo a partir das partes, para o sucesso do empreendimento. Não que as firmas de antigamente não tivessem de ser assim; mas, é inegável que o contexto da globalização tornou a nova gestão empresarial absolutamente imprescindível para a sobrevivência delas.

A proposta de Matoré (1973) sofreu duras críticas pela insuficiência dos dados apresentados e por não ter demonstrado sua(s) fonte(s), dentre outros aspectos. Porém, nenhum outro modelo foi proposto que pudesse abranger um modelo que desse conta da mudança do léxico e suas motivações sociais. Segundo Cambraia (2013, p.167), o próprio Matoré (1973, p. 91), no entanto, já havia deixado claro em sua obra que o estado incipiente dos estudos lexicológicos de então não lhe permitiam apresentar mais do que hipótese.

Cambraia (2013), em seu trabalho intitulado *Da lexicologia social a uma lexicologia sócio-histórica: caminhos possíveis*, apresenta um estudo de caso com o objetivo de propor um modelo empírista, quantitativo e de orientação laboviana que pudesse articular fatores intralingüísticos e extralingüísticos numa proposta de estruturação do léxico:

pretende-se mostrar, por meio de um estudo caso, que um modelo de organização e funcionamento do léxico (incluindo-se aí a questão das mudanças lexicais) não pode prescindir da articulação de fatores intralingüísticos e extralingüísticos. Essa articulação, já bem assentada nos estudos de mudança linguística em uma orientação laboviana, tem sido lamentavelmente negligenciada em estudos do léxico. (CAMBRAIA, 2013, p. 167-168)

No estudo de caso realizado pelo autor, buscou-se, inicialmente investigar o percurso histórico das expressões *esquadrão da morte*(EM) e *grupo de extermínio*(GE) e o que esse percurso revelaria em relação ao processo de criação e incorporação de neologismos ao léxico de uma língua.

A metodologia utilizada nesse trabalho ora dialoga com o método de Matoré (1973) ora rompe com ele: “[n]o que diz respeito ao recorte temporal, rejeita-se aqui a proposta de Matoré (tão criticada) de faixas de 30 a 36 anos” (CAMBRAIA, 2013, p.168). O autor adota o recorte de 5 anos que, segundo ele, seria uma extensão “não tão curta a ponto de dificultar a percepção de tendências gerais, nem tão extensa a ponto de impedir a percepção de qualquer tendência” (CAMBRAIA, 2013, p.169).

O autor demonstra a partir dos dados analisados do *corpus* que

seja por meio dos números absolutos, seja por meio dos números relativos, fica claro que EM predomina, sobretudo ao longo do período da ditadura no Brasil (1964-1985), enquanto GE tem sua prevalência no período posterior a ela. (CAMBRAIA, 2013, p. 173).

Ele argumenta que é esse paralelismo que sugere que o léxico estaria marcando uma mudança social, retomando Matoré (1973).

A conclusão do estudo de caso realizado por Cambraia (2013) discute o modelo proposto por Matoré. Esse modelo, segundo Cambraia (2013), é uma estratégia interessante desde que seja aplicado parcialmente, e não como havia sido concebido. Cambraia (2013, p.183) afirma que:

é possível sim articular critérios (internos) e externos (sociais) para representar a estrutura do léxico de uma época: assim, por exemplo, no que diz respeito ao período da ditadura no Brasil, é possível elaborar um microcampo como o de violência.

E, para finalizar, o autor assinala que os estudos de lexicologia sócio-histórica devem se encaixar em três eixos principais (temporal, social e espacial) e sugere que

um tal modelo de lexicologia sócio-histórica supõe uma prática mais baseada em estudos aprofundados de casos específicos com o objetivo, a longo prazo, de associação entre eles em dados sócio históricos do que análises descontextualizadas de grandes listas de neologismos. (CAMBRAIA, 2013, p.184)

2.2 LÍNGUISTICA DE CORPUS

A partir da criação de um *corpus* é possível realizar análises linguísticas baseadas em dados empíricos. Cria-se uma base dados textuais que possibilita, das buscas estruturadas neste banco de dados, análise de forma, conteúdo, conceitos, contextos de usos, entre outras. Há para isso, diversas ferramentas de *corpora* que podem auxiliar esse processo. O conjunto de práticas desta área de conhecimento é chamado linguística de *corpus*. A propósito do termo *corpus*, Léon (2007, p. 54) diz:

Podemos dizer que “corpus” é um termo ambíguo que se refere, ao mesmo tempo, a um conjunto de dados e a um conjunto de métodos. No primeiro sentido (conjunto de dados), podemos dizer todo linguista é potencialmente um utilizador de *corpus* pois a linguística permanece de modo indiscutível, uma disciplina empírica. No segundo sentido (um conjunto de métodos), as pesquisas sobre *corpus* implicam métodos indutivos mais que hipotético-dedutivo o que quer dizer que as análises conduzidas pelos dados (*data-driven*) são preferidas àquelas conduzidas por regras (*rule-driven*). Além disso, esses métodos são frequentemente estatísticos e/ou probabilísticos, embora se possam encontrar também métodos sobre *corpus* fundados a partir de simples concordância.

Segundo Berber Sardinha (2000, p. 325),

[a] Linguística de *corpus* ocupa-se da coleta e exploração de *corpora* ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.

No ano de 1999 comemorou-se o aniversário de 35 anos da criação do primeiro *corpus* linguístico eletrônico, o *corpus Brown*. Lançado em 1964, o *Brown University*

Standard Corpus of Present-Day American English continha uma grande quantidade de dados para a época: um milhão de palavras (Berber Sardinha, 2002, p. 323-324).

2.3 LINGUÍSTICA COMPUTACIONAL

Segundo Othero (2006, p.342), “a área responsável pela investigação do tratamento computacional da linguagem e das línguas naturais é conhecida como linguística computacional”.

As investigações recentes deram evidências positivas da possibilidade do diálogo entre áreas do conhecimento distintas: de um lado, as ciências exatas e, de outro, as ciências humanas. A síntese dessa oposição sempre será rica, porque sempre será o resultado de um trabalho interdisciplinar. Acreditamos que a ferramenta informática já é parte constitutiva das pesquisas que envolvem o tratamento de corpora linguísticos, e a matemática, pelo viés da estatística paramétrica, a fornecedora dos elementos analíticos necessários à compreensão do discurso. (Calvalcanti, 2005)

A palavra segundo o critério gráfico é o canônico constante para o desenvolvimento da linguística computacional. Segundo Basílio (2004, p. 13), “graficamente, podemos definir palavra como a sequência de caracteres que aparece entre espaços e/ou pontuação e que corresponde a uma sequência de sons que forma uma palavra”. É a partir do diálogo e a colaboração mútua entre a ciência da computação e a ciência da linguagem que se materializa a linguística computacional. (Alencar, 2006, p. 12-13).

2.3.1 PROCESSAMENTO DE LÍNGUAS NATURAIS (PLN)

As primeiras investigações institucionalizadas sobre o PLN começaram a ser desenvolvida no início da década de 50, depois da distribuição de 200 cópias de uma carta, conhecida como *Weaver Memorandum*, escrita por Warren Weaver, então vice-presidente da Fundação Rockefeller e exímio convededor dos trabalhos sobre criptografia computacional. Nessa carta, divulgada em 1949, Weaver convidava universidades e empresas, interessados potenciais, para desenvolver projetos sobre um novo campo de pesquisa que ficou conhecido como “tradução automática”, “tradução mecanizada” ou simplesmente MT (abreviação do inglês “Machine

Translation") (Silva, Montilha, Specia, Nunes, Junior, Martins, Pardo, 2007, p. 6.),, 2007, p. 6).

Para o desenvolvimento em PLN, devido a sua amplitude, faz-se necessário um recorte para que se saiba o tipo de metodologia a se utilizar. O PLN abrange várias e complexas áreas do conhecimento e, por isso, exige uma certa perspectiva a fim de traçar uma visão da área (Silva, Montilha, Specia, Nunes, Junior, Martins, Pardo, 2007, p. 6),, 2007, p.4)

O recorte utilizado neste trabalho será o processamento de línguas escritas e, especificamente, o do português do Brasil (PB).

Os estudos em PLN têm motivado diálogos frutíferos sobre a interface computador-máquina. Segundo Othero (2005, p. 26),

Atualmente, a busca por interfaces mais amigáveis e por softwares voltados para o trabalho com linguagem natural tem motivado muitas pesquisas na área, e diversos são os frutos resultantes de estudos na área de inteligência artificial e da linguística computacional.

Um exemplo deste diálogo é o software ELIZA. Criado pelo pesquisador Joseph Weizenbaum, esse aplicativo simula diálogos intermitentes com o usuário, exercendo a função de um psicanalista. Segundo Othero (2005), o primeiro *chatterbot*³ desenvolvido foi ELIZA. Ele foi programa para ser uma espécie de psicólogo, ou conselheiro sentimental. Na verdade, o objetivo principal de seu criador não era criar um “psicólogo virtual”, mas apenas um programa de conversação, utilizando um sistema que se baseava em “moldes” (*templates*, em inglês) para construir suas frases.

A linguística computacional é categorizada pela ciência da computação no campo de pesquisa chamado *inteligência artificial* (IA). Inteligência artificial é a área de conhecimento que se preocupa com a criação de protótipos computacionais que simulem faculdades cognitivas humanas.

Imprescindíveis para o avanço da IA foram os trabalhos dos matemáticos dos séculos XVII a XIX. No séc. XIX, surge a figura de Alan Turing⁴, mas só em 1956 é que a IA começa a ser reconhecida como ciência.

³ O nome *chatterbot* vem da junção de duas palavras inglesas: *to chat* (“conversar, bater papo”) e *bot* (abreviação de *robot*). Os *chatterbots* são programas desenvolvidos para interagir com usuários humanos através de diálogos em linguagem natural, na modalidade escrita.

⁴ Considerado um dos pais da computação, o cientista britânico foi um dos primeiros a vislumbrar a possibilidade de que as máquinas poderiam se tornar inteligentes e criou um modelo matemático teórico para o computador universal, antes que os primeiros equipamentos desse tipo de fato existissem. (ALAN, 2012)

Uma das perguntas que levantamos a partir do diálogo entre PLN e IA é: como fazer com que um computador comprehenda uma língua natural? E, a partir dessa pergunta, há outros questionamentos: É possível mapear processos cognitivos do reconhecimento, processamento e armazenamento do LM? É possível criar um protótipo que simule o funcionamento do LM?

2.3.2 ALGORITMOS

Algoritmos são descrições organizadas de um conjunto de comandos que, efetuados, resultam numa sucessão finita de ações. Podemos defini-lo também como uma lista de instruções (comandos) ordenadas que tenham por finalidade resolver um determinado problema.

Knuth (1973) apresenta uma versão para a origem da palavra *algoritmo*. Ela seria derivada do nome de um famoso matemático chamado Abu Ja'far Maomé ibn Mûsâ al-Khowârîsm (825 d.C.), que, traduzido literalmente, quer dizer “Pai de Ja'far, Maomé, filho de Moisés, de Khowârîzm”. Khowârîzm é hoje a cidade de Khiva, na ex-União Soviética. Esse autor escreveu um livro chamado *Kitab al jabrw'al-muqabala* (“Regras de restauração e redução”). O título do livro deu origem também a palavra *álgebra*. O significado da palavra *algoritmo* é muito similar ao de “receita”, “procedimento”, “técnica”, “rotina”.

Um algoritmo opera sobre um conjunto de entradas (no caso de um bolo, a farinha, os ovos, o fermento, etc.) de modo a gerar uma saída que seja útil (no caso de um bolo, agradável) para o usuário (no exemplo, o bolo pronto). Um algoritmo tem cinco características importantes:

- (i) Finitude: um algoritmo deve sempre terminar após um número finito de passos.
- (ii) Definição: cada passo de um algoritmo deve ser precisamente definido. As ações devem ser definidas rigorosamente e sem ambiguidades.
- (iii) Entradas: um algoritmo deve ter zero ou mais entradas, isto é, quantidades que são lhe fornecidas antes do algoritmo iniciar.
- (iv) Saídas: um algoritmo deve ter uma ou mais saídas, isto é, quantidades que tem uma relação específica com as entradas.

(v) Efetividade: um algoritmo deve ser efetivo. Isto significa que todas as operações devem ser suficientemente básicas de modo que possam ser em princípio executadas com precisão em um tempo finito por um humano usando papel e lápis.

Para representarmos os algoritmos, as formas mais comuns são as seguintes:

(i) Linguagem natural: os algoritmos são expressos diretamente em linguagem natural, como nos exemplos anteriores.

(ii) Fluxograma convencional: esta é uma representação gráfica que emprega formas geométricas padronizadas para indicar as diversas ações e decisões que devem ser executadas para resolver o problema.

(iii) Pseudo-linguagem: prega uma linguagem intermediária entre a linguagem natural e uma linguagem de programação para descrever os algoritmos.

Não existe consenso entre os especialistas sobre qual seria a melhor maneira de representar um algoritmo. Atualmente a maneira mais comum de se representarem algoritmos é através de uma pseudo-linguagem ou pseudo-código. Essa forma de representação tem a vantagem de fazer com que o algoritmo seja escrito de uma forma próxima de uma linguagem de programação de computadores.

Um dos objetivos do algoritmo é a resolução de um problema. Por exemplo, há diferentes maneiras de chegar a um dado lugar. A escolha do melhor caminho será baseada no melhor critério adotado em função da necessidade.

Segundo Duque (2012, p. 69),

a abordagem baseada em regras é utilizada para extrair automaticamente termos relevantes, mediante padrões encontrados nas sentenças de interesse (i.e., sentenças de tratamento). A técnica de processamento de língua natural Part-Of-Speech (POS) é utilizada para etiquetar as palavras em suas respectivas classes gramaticais (i.e., classificar nas classes substantivos, adjetivo, verbo, dentre outras).

A seguir um exemplo de algoritmo utilizado com a finalidade de extração de informações baseado em regras:

Algoritmo 1 - Extração de termos com o uso de regras.

```

1   for (i = 0; i < quantidade das sentenças;i++)
2       for (j = 0; j<= quantidade de regras; j++)
3           MatrizResultado[i][j] = 0;
4   regra[] <- getRegras(); //termo representativo+POS e somente POS
5   sentença[] <- getSentençasClassificadas();
6   for (i = 0; i <= quantidade das sentenças; i++)
7       for (j = 0; j<= quantidade de regras; j++)
8           if (regra[j] casou com sentença[i])
9               print ("Termo extraido");
10              MatrizResultado[i][j] = 1;
11      end
12  end

```

Figura 2. Algoritmo que descreve o pseudocódigo da extração de termos com o uso de regras
(DUQUE, 2012, p. 75)

Outro exemplo dado pela autora é o algoritmo para extração de termos com uso de regras e dicionário (Figura 3). Esse algoritmo descreve o pseudocódigo da extração de termos com o uso de regras e de dicionário.

Algoritmo 2 - Extração de termos com o uso de regras e dicionário.

```

1   for (i = 0; i < quantidade das sentenças;i++)
2       for (j = 0; j<= quantidade de termos no dicionário; j++)
3           MatrizDicionario[i][j] = 0;
4   for (i = 0; i < quantidade das sentenças;i++)
5       for (j = 0; j<= quantidade de regras; j++)
6           MatrizRegras[i][j] = 0;
7   dicionário[] <- getTermos();
8   regra[] <- getRegras(); //termo representativo+POS e somente POS
9   sentença[] <- getSentençasClassificadas();
10  for (i = 0; i <= quantidade das sentenças; i++)
11      for (k = 0; k<= quantidade de termos no dicionário; k++)
12          if (existe termo validado em dicionário[k] em
13              sentença[i])
14              print ("Termo existente na sentença");
15              MatrizDicionario[i][j] = 1;
16      end
17      for (j = 0; j<= quantidade de regras; j++)
18          if (regra[j] casou com sentença[i])
19              Termo <- termoIdentificado(regra[j]);
20              if (não existe termo em dicionário)
21
22                  insere (termo, dicionário, "Não Validado");
23                  print ("Termo novo extraído");
24                  MatrizRegras[i][j] = 1;
25              end
26              else print ("Termo extraído já existe no
27                  dicionário");
28      end
29  end
30 end

```

Figura 3. Algoritmo que descreve o pseudocódigo da extração de termos com o uso de regras e dicionário (DUQUE, 2012, p.76)

Segundo a autora, a abordagem com dicionário seria necessária para dar conta de fatos como flexão de plural e flexão de gênero. Porém, no que diz respeito à proposta do presente estudo, a solução proposta será outra: trabalharemos apenas com um algoritmo por regras. Iremos discutir esse assunto melhor mais adiante.

São utilizados também, a serviço da PLN, métodos estatísticos computacionais. O método matemático-estatístico-computacional de tratamento e análise de textos de André Camlong está descrito nas obras *Méthode d'analyse lexicale textuelle et discursive* (Camlong, 1996) e *Do léxico ao discurso pela informática* (Zapparoli; Camlong, 2002). Para o desenvolvimento de aplicações matemática-computacionais, Ebecken, Lopes e Costa (2003) citam algumas técnicas de aprendizagem de máquina:

- Hidden Makov Model
- Naive Bayes, Suport Vector Machines
- Árvores de decisão

Além dos cálculos probabilísticos, há alguns outros recursos computacionais oriundos das linguagens de programação de computadores, como funções que possibilitam a otimização dos processos que envolvem processamento de cadeia de caractere. Listamos alguns:

(a) Busca aproximada por *strings*: Dada uma coleção de *strings*, a busca aproximada por *strings* tem o objetivo de identificar, de forma eficiente, *strings* similares àquelas informadas em uma dada consulta.

(b) Flexibilização de consultas (*query relaxation*): Muitas vezes, ao serem executadas consultas, os termos de busca não coincidem com os dados existentes no banco de dados. Estas divergências podem ter diversas causas, dentre elas erros nos termos utilizados na consulta ou conhecimento limitado do usuário sobre o assunto pesquisado, inconsistências nos dados da base de dados, entre outros motivos. A partir da *query relaxation*, é possível retornar dados potencialmente interessantes ao usuário, com base na similaridade destes com os especificados na consulta.

(c) Limpeza de dados (*data cleaning*): Informações provenientes de diversas fontes podem acarretar em dados inconsistentes, ou seja, um mesmo objeto do mundo real pode ter diversas representações. Estas inconsistências podem ser geradas por inúmeros motivos, como erros no processo de coleta de dados ou falhas humanas. A limpeza de dados tem o objetivo o de detectar, e, então, corrigir ou remover os registros incorretos ou imprecisos.

2.4 DESIGNAÇÕES DE GÊNERO

O termo *gênero* é complexo por atravessar conceitos não muito bem definidos entre natureza e cultura ou biológico e social. Segundo Levi Strauss (1982, p. 43), o homem é um ser biológico ao mesmo tempo em que é um indivíduo social.

A designação *gênero* é utilizada para nos referirmos a características arraigadas no social, ou seja, em práticas culturais. Para nos referirmos a aspectos que abordem a biologia ou aspectos do “natural”, utilizamos o termo *sexo*.

A palavra *gênero* representa as múltiplas possibilidades dentro da cultura de nos reconhecermos femininos e masculinos, mulheres e homens. As relações de gênero vão além do biológico: são construções sociais, políticas e culturais de percepções sobre o que é ser um dado gênero.

Numa sociedade em que as condutas estabelecidas como “femininas” ou “masculinas” são reproduzidas na cultura, as identidades são delimitadas por espaços e comportamentos sob um rígido controle e vigilância das relações entre gênero e sexualidade. A sexualidade pode ser entendida como a vivência dos prazeres e desejos, constituída por uma complexa combinação de sentidos, representações e valores, passíveis de alterações (visto que não se apresentam de forma estática), pelas mudanças que histórica e socialmente foram modificando os conceitos e as formas de se ver e viver as práticas sexuais.

Um caso interessante de designação relacionada a gênero tornou-se evidente a partir de 2011, com um episódio acontecido no Canadá, mas que teve repercussões inclusive no Brasil, com a chamada *Marcha das Vadias*.

O movimento batizado em português brasileiro *Marcha das Vadias* (originalmente em inglês *Slutwalk*) teve inicio em Toronto, no Canadá, no de 2011, após diversos casos de violência sexual sofrida por mulheres na Universidade de Toronto. Devido a esse fato, um agente de polícia local foi chamado para se pronunciar e, durante sua participação, afirmou que as mulheres deveriam evitar se vestir como vadias (em inglês, *sluts*) para não sofrerem o estupro (RUSH, 2011). O termo *vadia* foi adotado e batizou esse movimento feminista. Em um espaço curto de tempo, o movimento ganhou milhares de adeptos nos EUA e na America Latina (MENEZES, 2014).

O carro chefe desse movimento é a luta contra todo tipo de violência simbólica e sexual sofrida historicamente pelas mulheres. Sobre os estudos sobre violência contra mulheres no Brasil, Santos e Izumino (2005, p. 2) dizem:

Os primeiros estudos sobre o tema têm por objeto as denúncias de violência contra as mulheres nos distritos policiais e as práticas feministas não-governamentais de atendimento às mulheres em situação de violência. Em meados dos anos 80, com o surgimento das delegacias da mulher, passam a privilegiar as ações do Estado nas esferas da segurança pública e da Justiça.

Santos e Izumino (2005, p. 2) esclarecem também que:

Entre os trabalhos que vieram a se constituir como referências a esses estudos, identificamos três correntes teóricas: a primeira, que denominamos de dominação masculina, define violência contra as mulheres como expressão de dominação da mulher pelo homem, resultando na anulação da autonomia da mulher, concebida tanto como “vítima” quanto “cúmplice” da dominação masculina; a segunda corrente, que chamamos de dominação patriarcal, é influenciada pela perspectiva feminista e marxista, compreendendo violência como expressão do patriarcado, em que a mulher é vista como sujeito social autônomo, porém historicamente vitimada pelo controle social masculino; a terceira corrente, que nomeamos de relacional, relativiza as noções de dominação masculina e vitimização feminina, concebendo violência como uma forma de comunicação e um jogo do qual a mulher não é “vítima” senão “cúmplice”.

A violência contra as mulheres resulta da socialização machista. Dada sua formação de macho, o homem julga-se no direito de espancar sua mulher. Esta, educada que foi para submeter-se aos desejos masculinos, toma este “destino” como natural. (SAFFIOTI, 2004, p. 79-80)

O fator agravante deste processo de dominação é a culpabilização da mulher pela violência praticada pelo outro. A partir do momento que estabelecemos um conjunto de práticas sociais “permitidas” a um gênero e não a outro, cria-se um imaginário social de que pessoas cuja prática social não se adéqua neste jogo não são vítimas e sim propulsoras de violência. Muitas frases levantadas durante a Marcha da Vadia representavam esse fenômeno. “Meu corpo, minhas regras”, “Se ser livre é ser vadia. Eis-me aqui!” (MARCIEIRA, 2011). Quando se estabelece que o outro “motivou” a violência, usando o exemplo das mulheres, pela a exibição de partes do corpo ou a adoção por estas de outras práticas marginalizadas moralmente a este gênero, transfere-se a posição da mulher do lugar de vítima para agressora, eximindo o real agressor da violência. Daí a sentença de culpada é atribuída à mulher.

Como se vê, o termo *vadia* parece ter passado por um processo de ressignificação que merece ser estudado, pois seria um exemplo interessante de como mudanças sociais acarretam transformações na semântica das palavras e, consequentemente, no léxico de uma língua.

3.0 HIPÓTESE DE TRABALHO E OBJETIVOS

3.1. HIPÓTESE DE TRABALHO

Toma-se aqui como hipótese de trabalho que o termo *vadia* terá passado por mudanças semânticas entre os sécs. XX e XXI em função de mudanças sociais como a emergência da *Marcha das Vadias*.

3.2. OBJETIVOS

3.2.1 GERAL

Analizar o uso do termo *vadia* a fim de identificar mudanças semânticas ao longo dos séculos XIX a XXI.

3.2.2. ESPECÍFICOS

- a) Formação de *corpus* composto de artigos de jornal brasileiro dos sécs. XIX e XXI.
- b) Coleta de ocorrências do termo *vadia* no *corpus*.
- c) Análise quantitativa e qualitativa dos dados.
- d) Teste da hipótese de que houve mudança semântica no termo *vadia* motivada por mudanças sociais.

4.0 METODOLOGIA

4.1. O CORPUS

Uma análise sócio-histórica do léxico nos permite acompanhar mudanças semânticas de uma língua ao longo do tempo levando em conta seu contexto social.

Atualmente é possível encontrar acervos online que possibilita a pesquisa em documentos jornalísticos num recorte histórico significativo seja ele diacrônico ou sincrônico. Um deles é a HEMEROTECA DIGITAL BRASILEIRA.

A HEMEROTECA DIGITAL BRASILEIRA⁵ é responsável por armazenar periódicos nacionais digitalizados proporcionando ampla consulta em acervos como jornais, revistas, anuários, boletins etc. Um dos objetivos da hemeroteca é possibilitar aos pesquisadores do mundo inteiro acesso livre e sem custo algum a documentos como *Diário Carioca*, *Correio da Manhã* e outros documentos que não circulam mais na forma impressa.

Para realizar o estudo proposto neste trabalho, realizamos uma busca no acervo da HEMEROTECA DIGITAL BRASILEIRA relativo ao *Jornal do Brasil* (JB) utilizando como critério de busca a palavra-chave *vadia*. A busca retornou os seguintes dados:

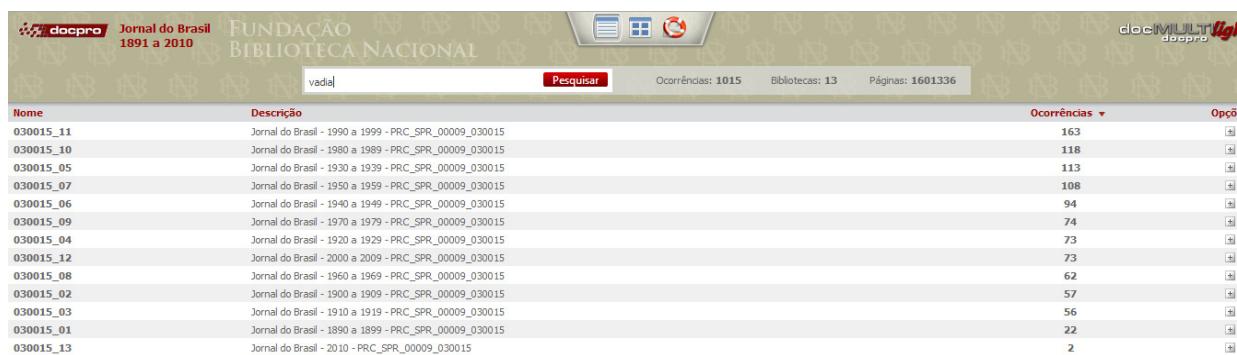


Figura 4. Ocorrências do termo *vadia* no acervo do JB na HDB.

Foram analisados todos os 1.015 textos de jornal retornados na busca do JB. Esta busca abarcou o período de 1890 a 2010. Em seguida, classificamos os registros que se referiam, especificamente, ao campo semântico de mulher. Dos 1.015

⁵ Disponível em <http://hemerotecadigital.bn.br>.

registros, 178 atenderam essa condição; porém, 21 destes registros, mesmo atendendo essa condição, não foram enquadrados por se tratarem de partes de palavras cruzadas. Adotamos a decisão metodológica não enquadrar esses 21 textos na análise, por constarem em seção do jornal com pouca riqueza semântica para análise. Portanto, os textos enquadrados no critério de análise estabelecido neste trabalho foram 157. Os textos válidos equivaleram a 15,5% da busca inicial.

Total bruto de textos	Total de textos com referência ao campo semântico de mulher	Total de textos com referência ao campo semântico de mulher (sem dados de palavras cruzadas)
1015 (100%)	178/1015 (17,5%)	157/1015 (15,5%)

Tabela 1. Total de textos com ocorrência do termo *vadia*

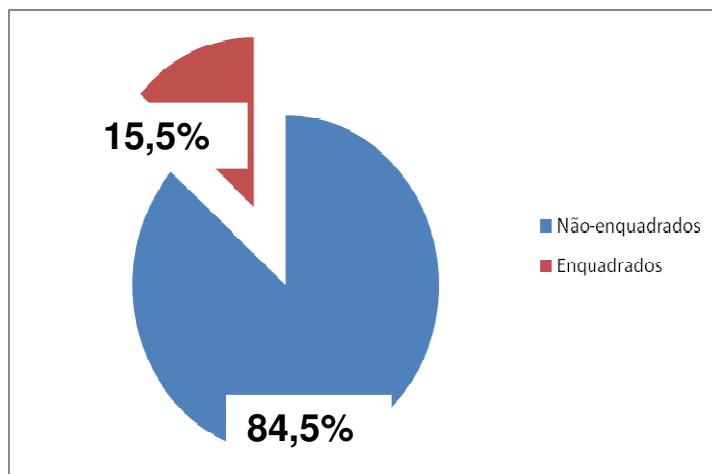


Figura 5. Textos com ocorrência do termo *vadia* enquadrados versus não-enquadrados no critério de seleção

Após a coleta dos textos (em formato de imagem), realizamos sua conversão em textos digital (em formato txt). Optamos por digitalizar apenas o texto do jornal que atendia o critério de presença do termo *vadia* (como na figura 8):

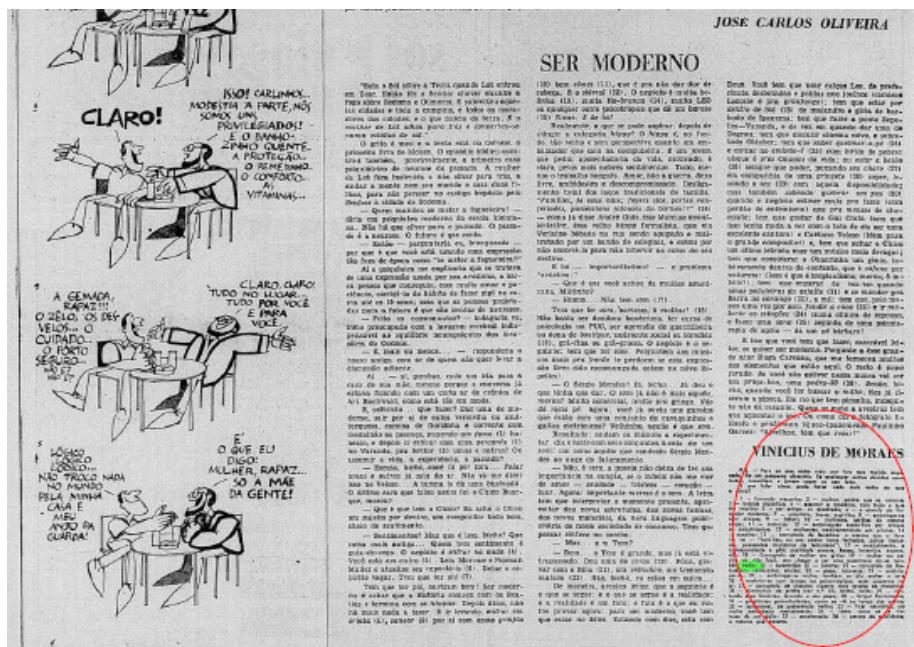


Figura 6. Exemplo de um texto do jornal usado para transcrição.

Houve apenas 8 casos em que o termo *vadia* foi encontrado mais de uma vez num mesmo texto. Em geral, a relação foi de 1 para 1. Apenas os textos listados abaixo tiveram uma relação diferente.

	<i>Data</i>	<i>Número de ocorrências de vadia</i>	<i>Localização no acervo do JB na HEMEROTECA DIGITAL BRASILEIRA</i>
1	10/11/1932	2	http://memoria.bn.br/DocReader/DocReader.aspx?bib=030015_05&PagFis=28035
2	24/04/1938	3	http://memoria.bn.br/DocReader/DocReader.aspx?bib=030015_05&PagFis=84016
3	08/05/1938	2	http://memoria.bn.br/DocReader/DocReader.aspx?bib=030015_05&PagFis=84318
4	09/07/1980	3	http://memoria.bn.br/DocReader/DocReader.aspx?bib=030015_10&PagFis=9655
5	28/07/1987	2	http://memoria.bn.br/DocReader/DocReader.aspx?bib=030015_10&PagFis=144008
6	16/10/1988	2	http://memoria.bn.br/DocReader/DocReader.aspx?bib=030015_10&PagFis=175251
7	23/03/1990	2	http://memoria.bn.br/DocReader/DocReader.aspx?bib=030015_11&PagFis=5085
8	01/06/1993	10	http://memoria.bn.br/DocReader/DocReader.aspx?bib=030015_11&PagFis=5085

Tabela 2. Textos com mais de uma ocorrência do termo *vadia*

Ao realizar um trabalho científico, faz-se necessário ressaltar aspectos metodológicos seguidos durante a pesquisa. Trataremos a seguir de alguns pontos que consideramos importantes discutir.

Os arquivos coletados do acervo do JB estavam em formato PDF. Optamos pela utilização de tecnologia OCR (*OpticalCharacterRecognition*)⁶ com o objetivo de realizar a primeira etapa de transcrição dos documentos. Feito isso, realizamos a segunda etapa de transcrição: a revisão, corrigindo os erros gerados pelo OCR. Uma das dificuldades da pesquisa é justamente esse processo de transcrição, que ainda é, infelizmente, realizado em grande parte, de forma manual devido a limitações da ferramenta OCR. Mesmo a tecnologia OCR mostrando uma significativa evolução para os processos de transcrição, ela ainda é limitada. Observamos durante o desenvolvimento deste trabalho as seguintes limitações desta ferramenta: (i) o OCR tem dificuldade de reconhecer alguns caracteres como *x*, *y* e *w*; (ii) em relação a trechos de línguas estrangeiras e trechos em itálico, a ferramenta também não se comporta bem; e (iii) palavras acentuadas, também não são corretamente reconhecidas.

Outro aspecto relevante observado durante a transcrição dos arquivos foi a diferença de alguns trechos do jornal à ortografia atual da variante padrão culta. Listamos alguns: *processal-a*, *comtudo*, *assumpto*, *emquanto*, *surprehende*, *sahe*, *collegas*, *illictio*, *fel-o*, *solemnamente*, *alumnas* (JB, 17.04.1924).

Além dos fatores discutidos até agora, um grande obstáculo no processo de manejo e processamento dos arquivos diz respeito à pouca legibilidade em que alguns arquivos se apresentam no acervo do JB, o que tornou a revisão da transcrição ainda mais penosa.

No processo de lematização deste *corpus*, lexias com variações gráficas foram agrupadas ao lexema correspondente. Por exemplo, lexias como *atenção*, *atenções*, *attenção* e *atencao* foram contabilizadas no lexema ATENÇÃO; as lexias *pae* e *paes* foram compiladas no lexema PAI; e lexias como *hontem* e *ontem* foram compiladas no lexema ONTEM.

Optamos por não excluir nomes próprios da lista de frequência do *corpus* analisado.

Outra opção metodológica foi a distinção de gênero na lematização. Embora tradicionalmente se subordinem as flexões de masculino e feminino a um só lexema, pareceu-nos relevante que neste trabalho fossem tratadas separadamente, já que o objeto de estudo é o termo *vadia* referindo-se a mulher. Portanto, todas as lexias do

⁶ O reconhecimento óptico de caracteres (OCR) é uma tecnologia que converte imagens de texto como, por exemplo, documentos digitalizados, em caracteres de texto atuais. (MICROSOFT, 2014).

corpus que designavam gênero não foram lematizadas. Por exemplo, as lexias como *garota* e *garoto* não foram agrupadas em ao lexema GAROTO, mas sim respectivamente em GAROTO e GAROTA. Listamos algumas que foram subordinadas à forma de feminino: *menina/meninas*, *filha/filhas*, *aluna/alunas*, *loba/lobas*, *pequena/pequenas*.

4.2. LEMATIZAÇÃO DO CORPUS

Em função da limitação de tempo imposta ao presente trabalho, não foi possível concretizar o modelo ideal de lematização automática discutida na seção 2.2.4.

Para que o estudo do termo *vadia* pudesse ser levado a cabo, adotamos um processo de lematização semi-automatizada, utilizando em especial o programa AntConc⁷ (versão 3.2.4w). O processo se deu da seguinte forma:

- 1) Os 157 textos do JB foram processados pelo AntConc para gerar uma lista de lexias: essa lista ficou composta de 13.978 lexias.
- 2) Elegemos como recorte para lematização as 1.400 primeiras lexias (10% das lexias do *corpus*). Dada a dificuldade de realizar manualmente a lematização, impôs a necessidade de realizar o tal recorte.
- 3) Essas 1.400 lexias foram lematizadas, gerando uma lista de 725 lexemas.

5.0 LEMATIZAÇÃO AUTOMÁTICA

Atualmente, um dos grandes desafios em estudos em lexicologia está na impossibilidade de se gerar listas de lexemas automaticamente. O que é possível encontrar hoje, enquanto processo lexical que já é possível ser realizada automaticamente é o chamado *stemming*. O processo de *stemming* consiste em reduzir a palavra à sua raiz. Segundo Porter *apud* Nunes (2002,p.13), “stemmer ou

⁷ Disponível em <<http://www.antlab.sci.waseda.ac.jp/software/antconc3.2.4w.exe>>.

‘Porter stemmer’ é o processo de remoção do comum morfológico e terminações flexionais de palavras em inglês”⁸.

Embora a lematização e o *stemming* sejam processos distintos, ambos são “ferramentas” que possibilitam o estudo e a compreensão do léxico de uma língua natural. Ainda segundo Nunes (2002, p. 14),

Lematização difere fundamentalmente de *stemming*. Enquanto lematização existe puramente no contexto lexicográfico, *stemming* não. Lematização é, pois, a representação da palavra através de seu masculino singular, adjetivos e substantivos e infinitivo (verbos), apenas no contexto da lexicologia. *Stemming* é a retirada de sufixos do radical, enquanto *stem* é o radical. Assim, as estruturas são distintas, embora eventualmente possam ser graficamente semelhantes.

Segundo Biderman (1936, p.84), a lematização é o processo de agrupar todas as formas relacionadas ou variantes morfológicas de uma palavra sob uma única entrada (lema) e a tarefa correlata de separar os homônimos. É no processo de lematização que vamos nos focar neste trabalho. Um das propostas deste trabalho é propor um modelo de sistema de processamento que possibilite a criação de listas de lexema a partir de listas de lexias.

Um dos grandes desafios no desenvolvimento de sistemas para o processamento de línguas naturais reside na ambiguidade das palavras, em que, cada uma delas pertence a uma classe linguística distinta, sendo necessária uma análise da sentença ou do contexto. Além disso, outra dificuldade reside na formação morfologia do PB: muitas são as formas verbais ou nominais que partilham de mesmas características mórficas, sendo difícil distinguir a classe de algumas dessas palavras fora do contexto. Além disso, o PB possui um inventário flexional, complexo trazendo mais um desafio para os estudos de PLN.

Laroca (2003, p. 37-49) apresenta os paradigmas derivacionais nominais e flexionais verbais do PB da seguinte maneira:

⁸ No original: “stemmer or ‘Porter stemmer’ is a process for removing the commoner morphological and in flexional endings from words in English”

Sufixos formadores de substantivos femininos	Substantivos primitivos masculinos	Substantivos derivados femininos
-ada	Boi, moleque, feijão, limão	Boiada, molecada, feijoada, limonada
-aria	Livro, papel, peixe	Livraria, papelaria, peixaria
-esa/-essa	Duque, barão, conde	Duquesa, baronesa, condessa.
-ia	Reitor, delegado, burguês	Reitoria, delegacia, burguesa.
-iria	Maestro	Maestrina
-inha	Galo	Galinha

Sufixos formadores de substantivos femininos	Verbos	Substantivos derivados femininos
-ção	Arrumar, trair	Arrumação, traição
-triz	Imperar	Imperatriz
-ura	Assinar, pintar	Assinatura, pintura

Substantivos formadores de substantivos femininos	Adjetivos	Substantivos derivados femininos
-eza	Triste, pobre	Tristeza, pobreza
-ia	Alegre, valente	Alegria, valentia
-(i)dade	Útil, igual	Utilidade, igualdade
-(i)dão	Escuro, manso	Escuridão, mansidão
-ice	Tolo, velho	Tolice, velhice
-itude	Magno, alto	Magnitude, altitude
-ura	Formoso, branco	Formosura, brancura

Sufixos formadores de substantivos masculinos	Substantivos primitivos femininos	Substantivos derivados masculinos
-al	Laranja, banana, areia, pomba	Laranjal, bananal, areial, pombal
-ão	Mulher, faca, porta	Mulherão, portão
-ario	Rã, bicicleta, erva	Ranário, bicicletário, herbário.
-eiro	Palha, formiga, tinta, agulha	Palheiro, formigueiro, tinteiro, agulheiro
-io	Mulher, casa	Mulherio, casario

Sufixos formadores de substantivos masculinos	Verbos	Substantivos derivados masculinos
-dor	Imperar	Imperador
-douro	Beber, suar	Bebedouro, suadouro
-mento	Casar, ferir	Casamento, ferimento
-tório	Lavar	Lavatório

Sufixos formadores de substantivos masculinos	Adjetivos	Substantivos derivados masculinos
-or	Fresco, amargo	Frescor, amargor
-ume	Negro	Negrume

Figura 7. Paradigma derivacional nominal do PB

VERBO	PRON.	INDICATIVO						SUBJUNTIVO			IMPERATIVO
		PRESENTE	PRETÉRITO PERFEITO	PRETÉRITO IMPERFEITO	PRETÉRITO MAIS-QUE-PERFEITO	FUTURO DO PRESENTE	FUTURO DO PRETERITO	PRESENTE	IMPERFEITO	FUTURO	
TRABALHAR	Eu	trabalho	trabalhei	trabalhava	trabalhara	trabalharei	trabalharia	trabalhe	trabalhasse	trabalhar	-
	Você	trabalha	trabalhou	trabalhava	trabalhara	trabalhará	trabalharia	trabalhe	trabalhasse	trabalhar	trabalhe
	Ele/Ela	trabalha	trabalhou	trabalhava	trabalhara	trabalhará	trabalharia	trabalhe	trabalhasse	trabalhar	trabalhe
	Nós	trabalhamos	trabalhamos	trabalhávamos	trabalháramos	trabalharemos	trabalhariam	trabalhemos	trabalhássemos	trabalharmos	trabalhemos
	Vocês	trabalham	trabalham	trabalhavam	trabalharam	trabalharão	trabalhariam	trabalhem	trabalhassem	trabalharem	trabalhem
	Eles/Elas	trabalham	trabalham	trabalhavam	trabalharam	trabalharão	trabalhariam	trabalhem	trabalhassem	trabalharem	trabalhem
ESCREVER	Eu	escrevo	escrevi	escrevia	escrevera	escreverei	escreveria	escreva	escrevesse	escrever	-
	Você	escreve	escreveu	escrevia	escrevera	escreverá	escreveria	escreva	escrevesse	escrever	escreva
	Ele/Ela	escreve	escreveu	escrevia	escrevera	escreverá	escreveria	escreva	escrevesse	escrever	escreva
	Nós	escrevemos	escrevemos	escrevíamos	escrevérâmos	escreveremos	escreveríam	escrevemos	escrevéssemos	escreveremos	escrevemos
	Vocês	escrevem	escreveram	escreviam	escreveram	escreverão	escreveriam	escrevam	escrevessam	escreverem	escrevam
	Eles/Elas	escrevem	escreveram	escreviam	escreveram	escreverão	escreveriam	escrevam	escrevessem	escreverem	escrevam
ASSISTIR	Eu	assisto	assissti	assistia	assistira	assistirei	assistiria	assista	assistisse	assistir	-
	Você	assiste	assissti	assistia	assistira	assistirá	assistiria	assista	assistisse	assistir	assista
	Ele/Ela	assiste	assissti	assistia	assistira	assistirá	assistiria	assista	assistisse	assistir	assista
	Nós	assistimos	assistimos	assistíamos	assistíramos	assistiremos	assistíramos	assistimos	assistíssemos	assistirmos	assistimos
	Vocês	assistem	assisstiram	assistiam	assistiram	assistirão	assistiriam	assistam	assistissem	assistirem	assistam
	Eles/Elas	assistem	assisstiram	assistiam	assistiram	assistirão	assistiriam	assistam	assistissem	assistirem	assistam

Outros Verbos Regulares: Falar, Voltar, Cantar, Mandar, Amar, Andar, Continuar, Tentar, Usar / Escolher, Correr, Dever, Bater, Viver / Partir, Dividir, Imprimir, Existir, Possuir, Desistir, Insistir, Resistir.

Alterações Ortográficas Regulares - Muitos verbos regulares apresentam alteração na última letra do radical, em determinadas formas. Isso acontece porque esta letra entra em contato direto com as várias flexões do verbo: sua junção com a vogal inicial da desínfência ocasiona as alterações, tipicamente devido a regras de acentuação e ortografia. Alterações em que a letra final do radical é vogal:

- verbos terminados em *oer* (*doer, moer, roer...*): formas rizotônicas do presente do indicativo recebem acento (*ó, óis, ói, óem*).
 - verbos terminados em *or* (*doar, perdoar, voar, magoar...*): acento em (*eu*) perdido por causa do duplo "o";
 - verbos terminados em *ear* (*bazar, golpear, passear...*): formas rizotônicas dos dois presentes ganham um "l" na junção do radical à desinência (*baseio, baseias,.. baseiam, baseie,.. baseiem*).
 - *estrear* (outros): além do "l" nas formas rizotônicas (como o caso acima), recebe acento no "e" adjacente, por causa do ditongo aberto (*estréio, estreias,.. estréiam, estréie,.. estréiem*).
 - verbos terminados em *air* (*cair, sair, trair, esivar...*): vogal temática "i" aparece inadvertidamente em (*eu*) *caio* e no presente do subjuntivo (*caia, caías, caíamos, caíais, caíam*).
 - verbos terminados em *guarquir* (*avenguir, adequar, obliquar...*): recebem acento no "U" das formas rizotônicas dos dois presentes (*avengúo, avengúas,.. avengúem, avengúe.. avengúem*).
 - verbos terminados em *guirquir* (*arguir, delinquir, outros...*): formas rizotônicas acentuadas, exatamente como o caso anterior (*argúo, argúas,.. argúem, argúa,.. argúam*);
- Alterações quando o final do radical é consoante:
- verbos terminados em *guir* (*rediguir, distinguir...*): "gu" vira "g" quando adjacente às vogais *a/o* (*distinguo, distinguia,.. distinguimos, distinguis...*).
 - verbos terminados em *guer* (*erguer, reerguer, erguer...*): mesmo caso, "gu" vira "g" quando adjacente às vogais *a/o* (*ergo, erga,.. ergamus, ergais...*).
 - verbos terminados em *gar* (*carregar, julgar, negar, rogar...*): caso oposto aos anteriores, "g" vira "gu" quando junto a "e" (*julguei, julgue,.. jüguemos, jügueis...*).
 - verbos terminados em *ger* (*proteger, reger...*): "g" vira "j" quando unido às vogais *a/o* (*protejo, proteja, protejais, protejamos...*).
 - verbos terminados em *gir* (*agir, reagir, dirigir, fingir...*): mesmo caso, "g" vira "j" quando adjacente às vogais *a/o* (*ajo, aja, ajas,.. ajam*).
 - verbos terminados em *cer* (*ficar, tocar, aplicar, explicar, implicar...*): "c" vira "qu" quando junto a "e" (*tiquei, tique,.. tiquemos, tiqueis...*).
 - verbos terminados em *gar* (*forgar, comegar, alcançar, capar...*): "g" vira "d" quando junto a "e" (*forcei, force,.. forcemos, forceis...*).
 - verbos terminados em *cer* (*crescer, parecer, aquecer, esquecer...*): caso oposto ao de cima, "c" vira "ç" quando adjacente às vogais *a/o* (*esqueço, esqueça,.. esqueçemos, esqueça/s...*).
 - verbos terminados em *zer*, *zir* (*jazer, luzir, induzir, conduzir, produzir, traduzir...*): 3a. pessoa singular do presente perde a desinência "e" normal dessa flexão (*conduz, em vez de conduze*).

Figura 8. Paradigma flexional verbal do PB

Portanto, lematizar as unidades lexicais do português brasileiro não é uma tarefa fácil. São necessários estudos e diálogos contínuos e aprofundados entre as áreas da linguística e da ciência da computação. Como o presente estudo é um trabalho de conclusão de curso de graduação, proporemos um modelo ideal para a resolução do nosso problema inicial: gerar uma lista de lexemas a partir da lista de lexias. A proposta se resume em apresentar um modelo teórico que abarque a lematização das lexias a partir da classificação em classes gramaticais e a utilização de uma base de dados lexicais. Para a realização do estudo do termo *vadia*, adotaremos um modelo simplificado, tratando de forma manual os casos que o modelo simplificado não der conta.

5.1. MODELO IDEAL PARA LEMATIZAÇÃO AUTOMÁTICA

Nesta seção discutiremos a arquitetura de um modelo ideal de lematização automática. Ilustraremos algumas das dificuldades de um tal modelo utilizando dados do *corpus* da presente pesquisa (que será detalhadamente apresentado na seção de metodologia).

Um modelo ideal deve ter a capacidade de gerar lista de frequência de lexemas a partir da lista de lexias e divide-se em três fases de desenvolvimento:

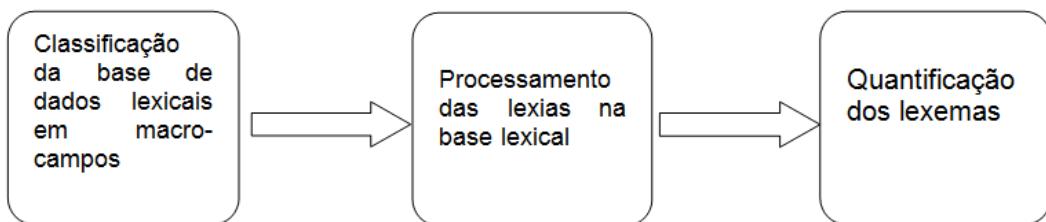


Figura 9. Fases de desenvolvimento da aplicação proposta

A fase I consiste em montar a base de dados que dará suporte ao algoritmo da fase II. A fase II testará as entradas $[x(n)]$ conforme os parâmetros estabelecidos na figura acima. A fase III consiste em agrupar os lexemas e quantificar as frequências, gerando assim uma lista de frequência de lexemas a partir da lista de lexias proposta.

5.1.1 FASE I: BASE DE DADOS LEXICAL (BDL)

Há um variado arsenal de arquiteturas voltadas para o processamento natural da linguagem, pois os sistemas podem variar de acordo com as especificidades da aplicação. No entanto, segundo Dias-da-Silva (1996, p. 234), dois grupos de componentes são imprescindíveis para a implementação de qualquer sistema desse tipo: as *bases de conhecimento* e os *módulos de processamento* que atuam sobre essas bases:

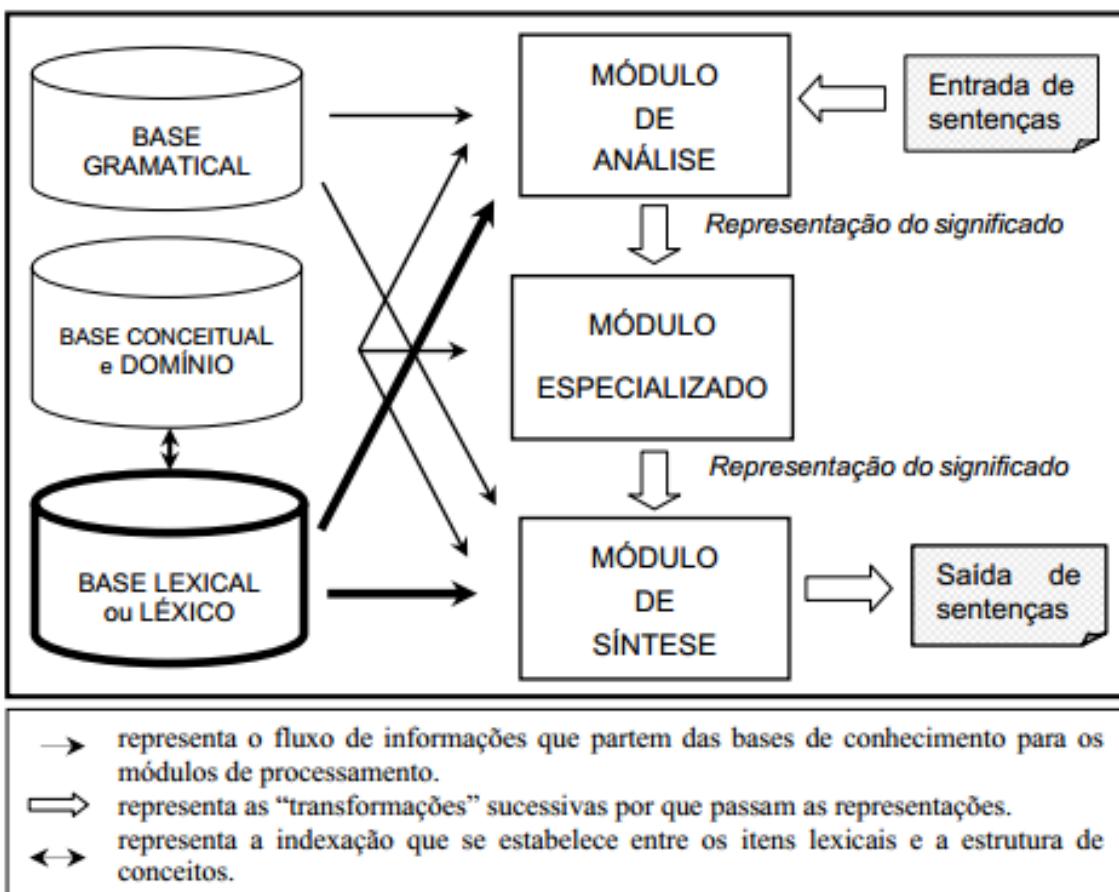


Figura 10. Arquitetura básica para aplicações de PLN (DIAS-DA-SILVA, 1996, p. 234)

Segundo Basílio (2004, p. 13), “palavras da língua são aquelas que aparecem listadas nos dicionários”. Portanto, deve-se utilizar nesta aplicação uma base de dados composta pelas unidades lexicais encontradas no dicionário de língua do PB. Este tipo de base é a base de dados lexicais (BDL). Para Biderman (2001, p. 90) “as bases de dados lexicais (BDL), estruturais e anotadas, são bases de conhecimentos lexicais”.

Segundo Correia (2009, p. 34), os dicionários estão também a serviço de sistemas de PLN:

Até às ultimas décadas do século XX, os dicionários eram apenas publicados em forma de livro, em suporte de papel. O desenvolvimento da informática e dos computadores pessoais levou também à produção de dicionários em suporte digital, ou dicionários eletrônicos, como são também conhecidos. Existem dois tipos de dicionários em suporte digital: aqueles que são concebidos para serem usados por máquinas, isto é, para servirem de base a sistemas diversos de processamento de língua natural (PLN) e os que são organizados para serem usados por pessoas.

Para o lematizador automático de um modelo ideal são necessários dois algoritmos. O primeiro é de classificação e o segundo é de processamento. O primeiro

vai classificar as unidades lexicais e o segundo deve processar as cadeias de lexias, testar os critérios na base de dados e retornar os lexemas correspondentes.

Para que o segundo algoritmo seja executado, é preciso que o primeiro algoritmo classifique a base lexical em verbos regulares, verbos irregulares, elementos não-verbais (substantivos, adjetivo e advérbio) e itens gramaticais,. Após a classificação, esses dados devem ser agrupados respectivamente em quatro tabelas distintas: VRE, VRI, NON e GRM.

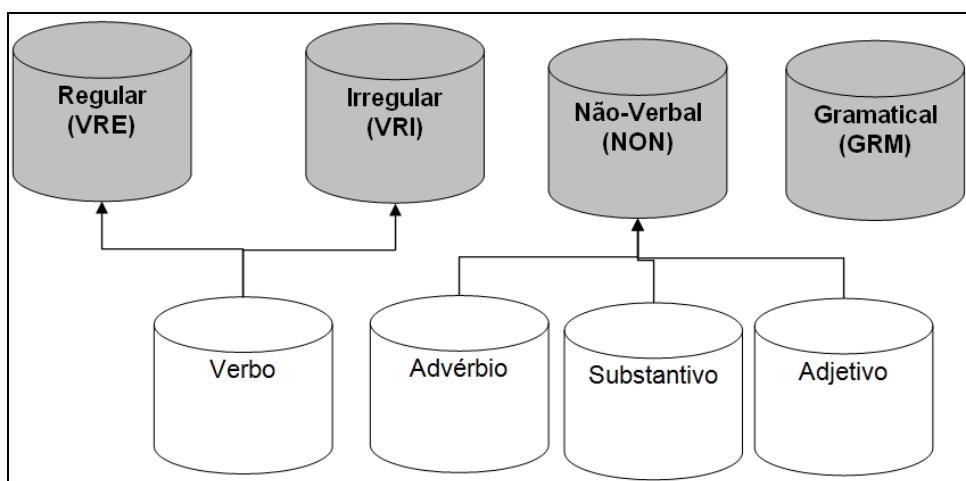


Figura 11. Representação gráfica das tabelas para a formação da BDL

5.1.2 FASE II: ALGORITMO PARA TRATAR DE CADEIA DE CARACTER

Apesar do grande número de variações flexionais do PB, é possível estabelecer alguns critérios comuns entre eles, utilizando a lógica de cadeias sufixais de análise e elegendo as cadeias de denominadores comuns, ou seja, as cadeias cujo comportamento compartilha de características comuns. Um exemplo seriam as flexões dos verbos regulares que não sejam afetadas por alomorfia morfologicamente condicionada. Este critério serve de grande valia a projeção de arquiteturas de PLN.

Uma proposta de algoritmo para tratamento de cadeias de caracteres é formada por um conjunto de condições com o objetivo de direcionar a busca na base lexical das entradas x , onde ($x = \text{lexia}$), sendo uma lexia uma cadeia de caracteres de número n . Por exemplo, no PB, há um percentual pequeno de cadeias de caracteres terminadas em ar, er e ir que não pertencem à classe dos verbos. Incluindo uma regra segundo a qual cadeias terminadas com esses três finais devem ser testadas na base de verbo primeiramente, diminui-se acentuadamente o número de processamentos

em nossa base de dados. Evitamos, portanto, que a entrada x seja processada em uma tabela à qual não é correspondente (p. ex., a cadeia amar será testada primeiramente na tabela de verbos antes de ser testada em outras tabelas) .

Baseando-nos na representação do paradigma flexional e derivacional proposto por Laroca (2003) (ver tabela 1), temos os seguintes parâmetros:

Tipo	Parâmetro	Descrição
Grupo dos verbos regulares (VRE)	Flexão verbal igual a 2 caracteres	<i>am, em, ei, ou, eu, iu, ia, ar, er, ir.</i>
	Flexão verbal igual a 3 caracteres	<i>ava, ara, era, ira, ará, era, irá.</i>
	Flexão verbal igual a 4 caracteres	<i>amos, aram, arão, arei, arem, ária, asse, avam, emos, eram, erão, erei, eria, esse, iam, imos, iram, irão, irei, irem, iria, iriam, ando, endo, indo.</i>
	Flexão verbal igual a 5 caracteres	<i>ariam, eriam, iriam, assim, essem, esse, issem, armos, ermos, irmos.</i>
	Flexão verbal igual a 6 caracteres	<i>ávamos, íamos, aramos, éramos, iramos, aremos, eremos, iremos.</i>
	Flexão verbal igual a 7 caracteres	<i>assemos, éssemos, aríamos, eríamos, iríamos, íssemos.</i>
Grupo dos elementos não-verbais (NON)	Derivação nominal igual a 2 caracteres	<i>la, al, ao, io, or</i>
	Derivação nominal igual a 3 caracteres	<i>ada, esa, iria, cão, ura, eza, dão, ice, ura, dor, ume</i>
	Derivação nominal igual a 4 caracteres	<i>aria, essa, inha, triz, dade, ario, eiro,</i>
	Derivação nominal igual a 5 caracteres	<i>idade, itude, douro, mento, tório</i>
Grupo dos verbos irregulares (VIR)	Sem parâmetro	<i>Lista</i>
Grupo dos itens gramaticais (GRM)	Sem parâmetro	<i>Lista</i>

Tabela 3. Parâmetros para o algoritmo de processamento

O comando de regras do algoritmo de processamento deve interpretar as lexias enquanto cadeias de caracteres: x = cadeia de caracteres (entrada); n = n.^º de

caracteres da cadeia; $z = n.$ ^o de caracteres do parâmetro. A fórmula que monta a cadeia de caracteres que será executada na BDL é a seguinte:

$$[x(n) - x(z)]$$

Por exemplo, para a cadeia de caracteres *comprar* (com 7 caracteres), no parâmetro flexão verbal igual a 2 caracteres, que compara os últimos 2 caracteres da cadeia, o item que o algoritmo irá buscar na base de dados é *compr*.

Para o modelo proposto, algumas questões são passíveis de discussão:

- Questão 1: O resultado da busca com base no parâmetro acima, ao identificar a lexia *comprei*, o algoritmo vai procurar na base lexical, a cadeia *compr*. Neste caso, a consulta retorna dois lexemas: COMPRAR e COMPREENDER. Para resolver este problema, ao escrever o algoritmo na sintaxe da linguagem de programação escolhida, deve-se explicitar a seguinte regra: a cadeia de caracteres procurada na base lexical deve ter tamanho igual a cadeia de caracteres da lexia procurada mais dois. Outro exemplo seria a lexia *alegria*. Antes do tratamento deste problema, a consulta encontra na base lexical os seguintes lexemas: ALEGRIA e ALEGREMENTE. Com a solução proposta neste item, esse problema é solucionado.

- Questão 2: Outra questão diz respeito às alomorfias morfologicamente condicionadas do PB. Um exemplo seria o verbo *abraçar*. Em alguns tempos verbais e na primeira pessoa do indicativo o *c* toma lugar do *ç* (cf. *abraçar* x *bracei*). Neste caso, é preciso rastrear estes casos e inseri-los numa cláusula de exceção.

- Questão 3: Acepções lexicais de nomes próprios e lexias de idiomas diferentes ao PB devem ser retomadas a sua forma original, ou seja, a lexia terá o mesmo valor do lexema nestes dois casos.

- Questão 4: O problema das lexias que tiverem com problemas na forma e não forem atribuídas ao lexema corresponde na forma gráfica correta deve ser corrigido pelo usuário na lista de lexias. Por exemplo, formas como *malabarismo* e *hontem* não serão associadas aos lexemas MALABARISMO e ONTEM, respectivamente. Por terem variação gráfica e não corresponderem à ortografia atualda variante padrão culta, estes casos são tratadas como nomes próprios ou de outro idioma.

- Questão 5: Erros gráficos, o modelo proposto não consegue tratar. Estes casos serão lematizados enquanto lexema distinto, ou seja, novo lexema. Exemplo: *xinema, circunstnacia, colidiano, collega, colleglo, cometetrem, commetera*, etc.
- Questão 6: Inadequações de acentuação, o algoritmo poderá tratar a partir da criação de funções específica para eliminar caracteres especiais de strings. Por exemplo, utilizando a linguagem PL SQL é possível criar uma empregando o mecanismo chamado “*Collate*” que trata especificamente caracteres especiais, tais como cedilha, acentos, e outros símbolos. Exemplo: *circunstância* x *circunstancia*.
- Questão 7: Casos de sinonímia, o modelo proposto não consegue tratar. Exemplo: *colar* (substantivo x verbo), *completas* (adjetivo x verbo), *banca* (substantivo x verbo), etc.
- Questão 8: Apenas a lematização dos itens lexicais no plural não enquadram na regra proposta pela equação acima. Neste caso, isso deve entrar no item de exceção. A equação proposta neste trabalho que trabalha com cadeias referenciais baseada nas flexões derivacionais do PB não abrange a complexidade do plural desta língua. Para este caso, é necessário que, neste item, esteja fora dos argumentos baseados na equação: $[x(n) - x(z)]$. Para isso, basta mapear os casos de uso do plural e criar um argumento específico para ela.

Processar lexias do PB não é uma tarefa simples. É necessário, além de um amplo conhecimento em linguística e em formação das palavras do português brasileiro, conhecimento em linguagens de programação, desenvolvimento de software, banco de dados e outras áreas dentro da ciência da computação. Daí o grande desafio desta tarefa.

6.0 ANÁLISE DOS DADOS

A partir da transcrição dos 157 textos, criamos o *corpus* deste trabalho. O conjunto dos textos apresentam as seguintes características:

<i>Tokens</i>	<i>Types</i>
60.693	13.978

Tabela 4. Número de *tokens* e *types* no *corpus*

Foram lematizadas as 1.400 primeiras lexias (ou types), gerando-se uma lista de 725 lexemas, que serviu de base para as análises que se seguem.

6.1 FREQUÊNCIA

Realizamos três tipos de recorte temporal da primeira ocorrência do termo *vadia* (em 09.11.1891) até a última (em 12.09.2009): intervalos de sessenta, década e quinquênio anos. No total houve 185 ocorrências (distribuídas em 157 textos): foram 184 ocorrências no singular e 1 ocorrência no plural. Houve ainda 8 ocorrências de *vadio* e 3 de *vadios*, mas, como se esclareceu na metodologia, as formas femininas foram consideradas separadamente em relação às masculinas.

O cenário de registros de *vadia* nos textos do JB foi o seguinte:

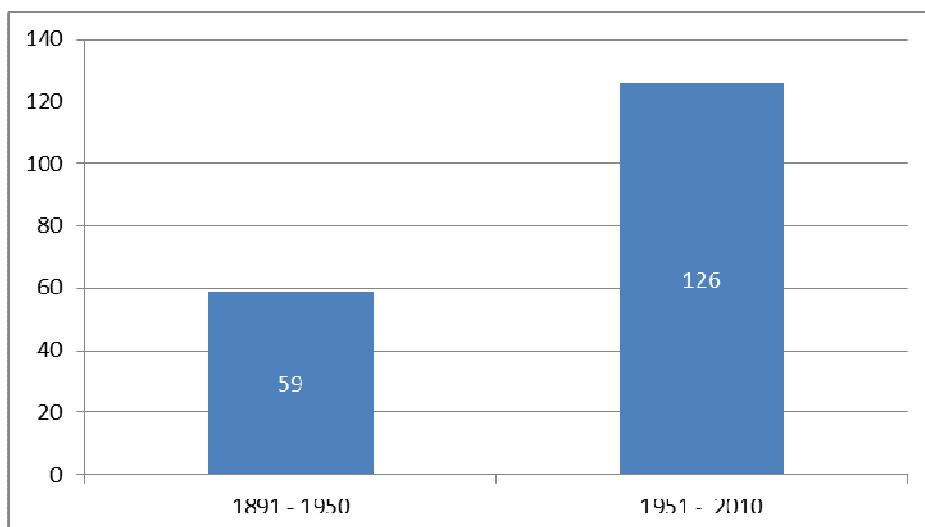


Figura 12. Total de ocorrências de *vadia* em intervalos de sessenta anos

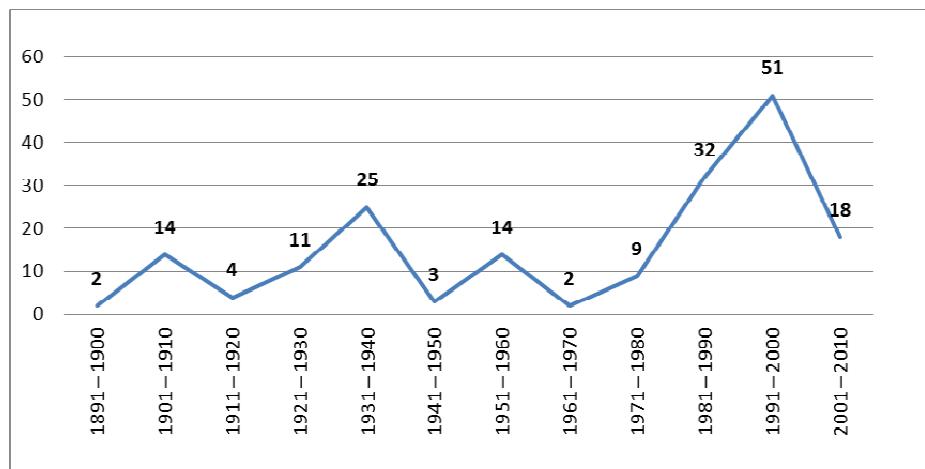


Figura 13. Total de ocorrências de *vadia* em intervalos de dez anos

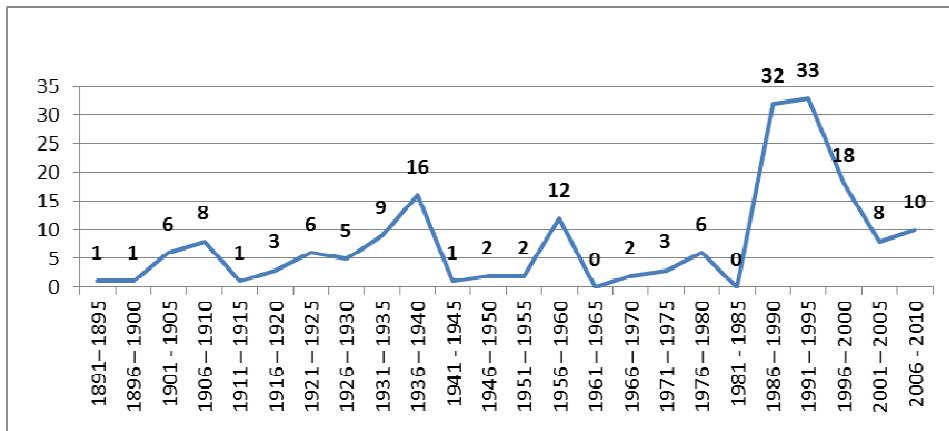


Figura 14. Total de ocorrências de *vadia* em intervalos de cinco anos

Durante o período de recorte deste trabalho (1891 a 2010), percebemos alguns picos de frequência do emprego do termo *vadia*. Na figura 14 (por sessenta anos), percebe-se uma maior frequência na faixa 1951-2010. Na figura 15 (por década), os picos estão em 1901-1910, 1931-1940, 1951-1960 e 1991-2000. Olhando com maior detalhe, como permite a figura 16 (por quinquênio), vemos que os picos são mais especificamente em 1906-1910, 1936-1940, 1956-1960 e 1986-1990 em conjunto com 1991-1995.

Para entender esses cinco picos, tomaremos como referência o cenário político do século XX e ações realizadas e organizadas por grupos de mulheres.

É curioso observar que os picos de ocorrência de *vadia* não estão desassociados de marcos históricos relacionados ao processo de gradativa emancipação política, social e moral das mulheres.

Segundo Brasil (2013, p. 1),

A história do movimento feminista possui três grandes momentos. O primeiro foi motivado pelas reivindicações por direitos democráticos como o direito ao voto, divórcio, educação e trabalho no fim do século 19. O segundo, no fim da década de 1960, foi marcado pela liberação sexual (impulsionada pelo aumento dos contraceptivos). Já o terceiro começou a ser construído no fim dos anos 70, com a luta de caráter sindical. No Brasil, o movimento tomou forma entre o fim do século 18 e início do 19, quando as mulheres brasileiras começaram a se organizar e conquistar espaço na área da educação e do trabalho. Nísia Floresta (criadora da primeira escola para mulheres), Bertha Lutz e Jerônima Mesquita (ambas ativistas do voto feminino) são as expoentes do período.

1906-1910: Segundo Timoteo (2013, p.93-94), “[n]a primeira década do século XX, existiam organizações feministas socialistas, anarquistas e liberais em vários países da América Latina”.

1936-1940: Neste período, o país era governado pelo presidente Getúlio Vargas e no ano de 1940 surgiu um movimento popular liderado por mulheres que almejavam o direito de atuar e serem inseridas na vida política do País, direito este negado até então. Neste período elas conquistaram um grande feito: o direito ao voto. (BRASIL, 2013) Ainda segundo Brasil (2013, p.1),

Mesmo assim, um ano após de conquistado o direito ao voto, em 1934, Carlota Pereira Queiróz⁹ torna-se a primeira deputada brasileira. Naquele mesmo ano, a Assembleia Constituinte assegurava o princípio de igualdade entre os sexos, o direito ao voto, a regulamentação do trabalho feminino e a equiparação salarial entre os gêneros.

Como sabemos o período da ditadura foi um momento de grande repressão aos movimentos sociais. E com o movimento feminista não foi diferente, porém mesmo atuando longe do palco social, o movimento de mulheres continuou com suas lutas. Segundo Otto (2004, p.239),

Com o golpe de 1937 ocorre um longo período de refluxo do movimento feminista que se estende até as primeiras manifestações nos anos 1970. No entanto, isso não significa que durante esse longo período as mulheres não tiveram nenhum papel no mundo público; muito pelo contrário, houve momentos importantes de participação da mulher, como o movimento no início da década de 1950 contra a alta do custo de vida, por exemplo.

1956-1960: Neste período o País vivia o chamado “milagre econômico”, que gerou um processo de industrialização paralelamente marcado por evolução tecnológica e científica. Os grupos de frente, que reivindicavam melhorias de condições sociais das mulheres, polemizaram assuntos como contraceptivos. No Brasil, as discussões sobre os métodos contraceptivos modernos entre eles os das pílulas anticoncepcionais diz respeito a políticas internacionais voltadas a redução de natalidade. Segundo Pedro (2003, p.241),

No Brasil — assim como nos países do terceiro mundo —, a divulgação dos métodos contraceptivos modernos, entre estes o das pílulas anticoncepcionais, fez parte de políticas internacionais voltadas para a redução da população. Isto foi muito diferente do que ocorreu com mulheres

⁹ Formou-se médica pela Faculdade de Medicina de São Paulo e foi a primeira mulher eleita deputada federal.

de países europeus, cujas políticas natalistas tinham adquirido muita força após as guerras mundiais. Assim, enquanto em lugares como a França a pílula somente foi liberada para consumo em 1967, no Brasil a pílula anticoncepcional e o DIU foram comercializados sem entraves desde o início da década de 60.

Note-se que, no período que abarcou a ditadura militar no Brasil, houve um decréscimo na recorrência do termo *vadia*, decréscimo este significativo. Dado que nos intervalos temporais que cobriam o período da ditadura (recortes de 1961 a 1985) houve apenas 11 registros do termo. Segundo Brasil (2013, p. 1),

Com a ditadura do Estado Novo, em 1937, o movimento feminista perde força. Só no fim da década seguinte volta a ganhar intensidade com a criação da Federação das Mulheres do Brasil e a consolidação da presença feminina nos movimentos políticos. Mas logo vem outro período ditatorial, a partir de 1964, e as ações do movimento arrefecem, só retornando na década de 70.

Um fato curioso ocorreu durante este período no Brasil em 1975 e nos anos 80, que Brasil (2013, p. 1) descreve da seguinte maneira:

Um dos fatos mais emblemáticos daquela década foi a criação, em 1975 (Ano Internacional da Mulher), do Movimento Feminino pela Anistia. No mesmo ano a ONU, com apoio da Associação Brasileira de Imprensa (ABI), realiza uma semana de debates sobre a condição feminina. Ainda nos anos 70 é aprovada a lei do divórcio, uma antiga reivindicação do movimento. Nos anos 80, as feministas embarcam na luta contra a violência às mulheres e pelo princípio de que os gêneros são diferentes, mas não desiguais. Em 1985 é criado o Conselho Nacional dos Direitos da Mulher (CNDM), subordinada ao Ministério da Justiça, com objetivo de eliminar a discriminação e aumentar a participação feminina nas atividades políticas, econômicas e culturais.

1986-1990 e 1991-1995: A maior ocorrência do termo *vadia* foi nesta década, que é período pós-ditadura militar no Brasil. Segundo Otto (2004, p.239) é fruto da ditadura militar o fenômeno que a autora chama de “novo feminismo”.

A emergência do feminismo em pleno governo Médici determinou que ele surgisse dentro e fora do país e em boa parte no exílio. Na Europa e nos Estados Unidos havia cenários de grande efervescência política, de revolução dos costumes, de radical renovação cultural, enquanto no Brasil o clima era de ditadura militar, repressão e morte. As características que o movimento feminista teve nos dois hemisférios estão intimamente ligadas a esses cenários, sendo que os primeiros grupos feministas em 1972, em São Paulo e no Rio de Janeiro, foram inspirados no feminismo do Hemisfério Norte.

Além de tratar do “novo feminismo” que emerge da ditadura, Otto (2004) discorre sobre o período pós-ditadura que ela chama de “feminismo na redemocratização”. Nesse período, surgem grupos temáticos de discussão dos direitos das mulheres, delegacias voltadas à mulher e o fenômeno que ela chama de “feminismo acadêmico”.

Surgiram grupos feministas temáticos, como também houve espaço para o surgimento e o desenvolvimento do que se poderia chamar de feminismo acadêmico, ancorado no Departamento de Pesquisa da Fundação Carlos Chagas, em São Paulo, e em pesquisas de ciências humanas e educação realizadas nas grandes universidades do país, em algumas das quais surgiram Núcleos de Pesquisa em Estudos da Mulher. A partir de 1985, foram criadas as delegacias especializadas. O feminismo, as feministas e as delegacias da mulher não resolveram a questão da violência, mas a criação das delegacias foi um avanço na medida em que a mulher passou a ser reconhecida como vítima de violência. O segundo tema que se tornou central no movimento feminista a partir da década de 1980 foi a implantação do Programa de Atenção Integral à Saúde da Mulher (PAISM), pelo Ministério da Saúde, que envolvia três temas: planejamento familiar, sexualidade e aborto (OTTO, 2004, p. 240)

É justamente no período pós-ditadura que há a maior ocorrência de registros do termo *vadia*. Esse fenômeno de aumento de frequência que vêm se arrastando por todo o século parece ter refletir mudanças na percepção da mulher pela sociedade. Além disso, os papéis sociais atribuídos a este gênero parecem ter sido ressignificados. Segundo esclarece Otto (2003, p.240), “pode-se inferir que, ao evidenciar a existência de múltiplas identidades e de um feminismo difuso na sociedade, (...) vislumbra que o sujeito na história não é mais o de uma figura universal, tendo o homem como o protagonista”.

6.2 CO-OCORRÊNCIAS

6.2.1. GERAL

Uma das formas de apreender o significado de uma palavra é verificar com quais outras ela co-ocorre em um mesmo texto. Para analisar isso, elaboramos uma lista de frequência.

Primeiramente, apresentamos uma lista tomando como referência os 50 lexemas mais frequentes:

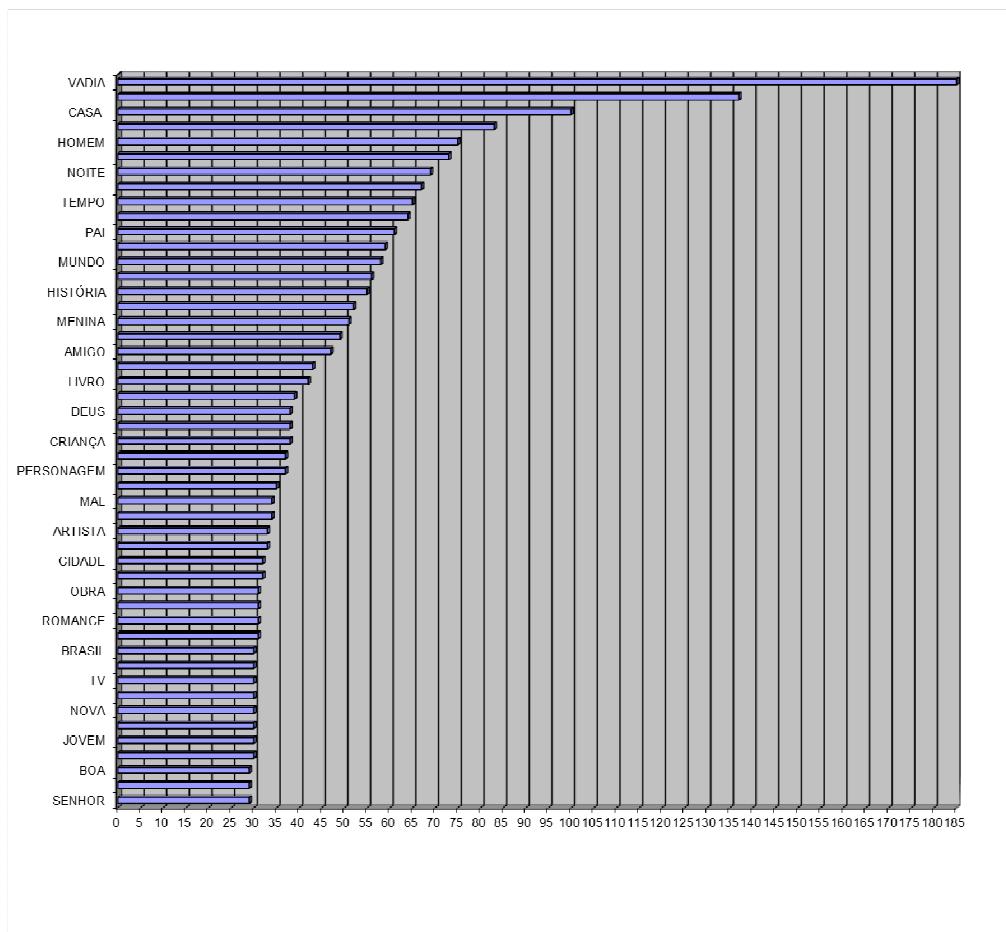


Figura 15. Cinquenta lexemas mais frequentes no *corpus*

Esses 50 primeiros lexemas podem ser agrupados em certos campos semânticos¹⁰:

Campo Semântico	Lexemas	Quantidade
<i>Cultura (ficção)</i>	FILME, HISTÓRIA, LIVRO, PERSONAGEM, ARTISTA, OBRA, ROMANCE, TV, MÚSICA	9
<i>Qualificadores</i>	MAL, VADIA, PEQUENA, CERTO, NOVA, BOA, NOVO, BOM	8
<i>Homem</i>	HOMEM, PAI, AMIGO, FILHO, MARIDO, SENHOR	6
<i>Lugar</i>	CASA, MUNDO, RUA, RIO, CIDADE, BRASIL	6
<i>Tempo</i>	DIA, NOITE, TEMPO, VIDA	4
<i>Mulher</i>	MULHER, MÃE, MENINA, FILHA	4
<i>Econômico</i>	MISERÁVEL, POBRE, CONTA	3
<i>Parte do corpo</i>	OLHO, MÃO	2
<i>Pessoa</i>	CRIANÇA, JOVEM	2

¹⁰ No foram colocados neste quadro os lexemas COISA, CASO e PORTA por se considerar que não formam um campo específico.

<i>Religião</i>	DEUS, FAMÍLIA	2
<i>Sentimento</i>	AMOR	1
<i>Lei</i>	POLÍCIA	1

Tabela 5. Campos semânticos dos cinquenta lexemas mais frequentes no *corpus*

É interessante verificar, baseando-nos nos dados da tabela 4 que o campo semântico de *cultura (ficção)* está no topo da lista. É possível deduzir, a partir desses dados, que o termo *vadia* é majoritariamente usado, com base no *corpus* analisado, em contextos de ficção como filmes, teatro, novelas, livros, entre outros.

Outra observação interessante diz respeito ao segundo campo semântico mais frequente, o dos *qualificadores*: parece-nos que o uso de qualificadores associados ao termo analisado traduz a perspectiva dos julgamentos envolvidos quando utilizado este termo. Inserido num contexto moral, o termo *vadia* é por essência, um julgamento moral, traduzindo a perspectiva social ou individual de alguém em relação ao outro alguém, julgamento este baseado em práticas sociais que transgridem práticas morais para aquele contexto histórico.

O que mais chama a atenção, porém, é o terceiro campo mais frequente do gráfico de campos semânticos: o de *homem*. O campo semântico de *homem* é recorrente em contextos em que termo *vadia* é utilizado. O termo *vadia*, que remete especificamente à mulher, associado ao campo semântico *homem*, remete, no imaginário social, à promiscuidade.

Além disso, outros casos interessantes são os campos de *lugar* e de *tempo*. Os dois campos tiveram alto índice e grande ocorrência de lexemas. Porém, são lexemas encontrados em vários gêneros textuais e recorrentemente usados em contextos múltiplos, sendo indispensáveis para a construção textual, o que levanta a discussão sobre quais campos semânticos se devem evidenciar em trabalhos como este e se tais campos semânticos trariam alguma evidência empírica e semântica significativa para análise.

Aspectos econômicos foram temas recorrentes no *corpus* analisado. Segundo estudos de gênero, a transição da atuação da mulher do ambiente privado para o público trouxe múltiplas identidades para essa “nova” mulher, que saía dos afazeres domésticos e passava a trabalhar profissionalmente no ambiente público, ambiente este povoadão majoritariamente por homens. No contexto econômico, a mulher passou

da situação dependente para independente, ou seja, ela mesma conquistava seu próprio dinheiro.

Quanto ao campo de *religião*, o que nos parece é existir a perspectiva da mulher enquanto reflexo de Maria, que gradativamente foi sendo alterada pela inserção desta mulher no contexto público, o que a possibilitou uma nova forma de lidar com seu corpo e sua sexualidade.

Já o contexto policial, representado pelo campo semântico de *lei*, parece revelar que há resquícios do termo *vadia* relacionado ao crime de vadiagem, sentido este primitivamente atribuído ao termo.

6.2.2. POR RECORTE HISTÓRICO DE SESSENTA ANOS

Tomando como referência o recorte histórico de sessenta anos (que divide a faixa temporal total de 1891 a 2010 em dois grandes blocos), criamos 2 *subcorpora*: o primeiro *subcorpus* é o Grupo A, que compreende os arquivos referentes ao período de 1891 a 1950, e o segundo *subcorpus* é o Grupo B, que compreende os arquivos de 1951 a 2010.

Selecionamos para comparação os 10 lexemas mais frequentes em cada um desses *subcorpora*:

Período	Posição	Lista de lexemas	Frequência
Grupo A Período: 1890 - 1950 Nº de arquivos: 46 Tokens: 23.232 Types: 6.674	1º 2º 3º 4º 5º 6º 7º 8º 9º 10º	DIA VADIA CASA OLHO MÃE RUA MULHER NOITE POBRE TEMPO	61 59 54 41 40 36 34 33 29 29
Grupo B Período: 1951 - 2010 Nº de arquivos: 111 Tokens: 37.463 Types: 9.672	1º 2º 3º 4º 5º 6º 7º 8º 9º 10º	VADIA MULHER DIA FILME VIDA HOMEM MUNDO CASA COISA HISTÓRIA	126 103 83 65 59 52 47 46 46 43

Tabela 6. Lexemas por recorte temporal de sessenta anos

A partir dos dados apresentados é possível dizer que, em função da diferença de frequência significativa de lexemas em cada grupo, o termo *vadia* se insere em campos semânticos distintos.

É interessante notar que o número de textos de cada grupo é diferente. O grupo B possuiu quase o dobro de textos em relação ao grupo A (grupo A com 46 textos e grupo B com 111 textos), o que implica que houve um aumento significativo do emprego do termo *vadia* neste período.

Alguns lexemas, durante ambos os períodos, foram compartilhados. São eles, CASA, DIA, MULHER, e VADIA. Porém, o que mais chama atenção são os lexemas distintos entre esses dois meio séculos. Por exemplo, o lexema MÃE ocorre apenas no grupo A e o lexema HOMEM somente no GRUPO B. Para o período analisado tivemos 40% dos lexemas que apareceram em ambos recortes históricos (grupo C), porém tivemos 60% de lexemas distintos, ou seja, lexemas que aparecem enquanto frequência significativa apenas em um período e não em outro (grupo A e B):

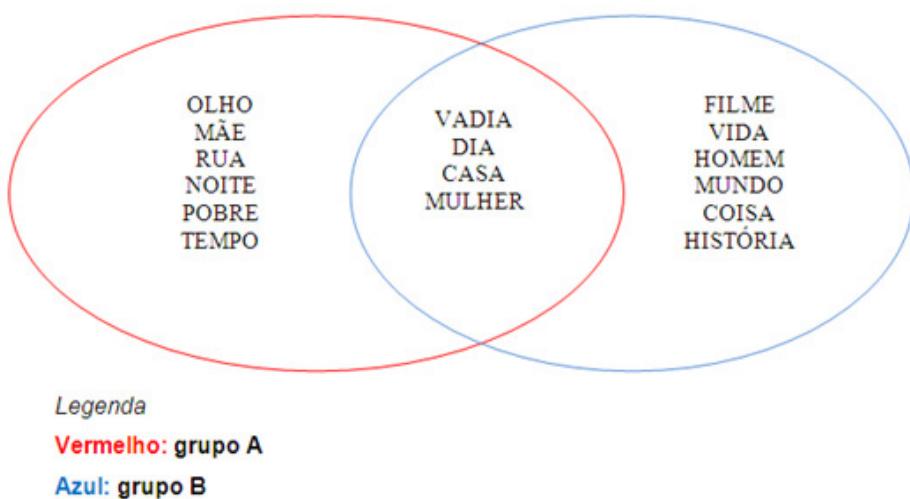


Figura 16. Intercessão dos grupos A e B dos 10 lexemas mais frequentes

Abaixo, analisaremos cada um dos lexemas que aparecem nesses grupos:

Grupo A:

OLHO: Os contextos de uso deste lexema, na maioria dos casos, remetiam aos olhos da mulher.

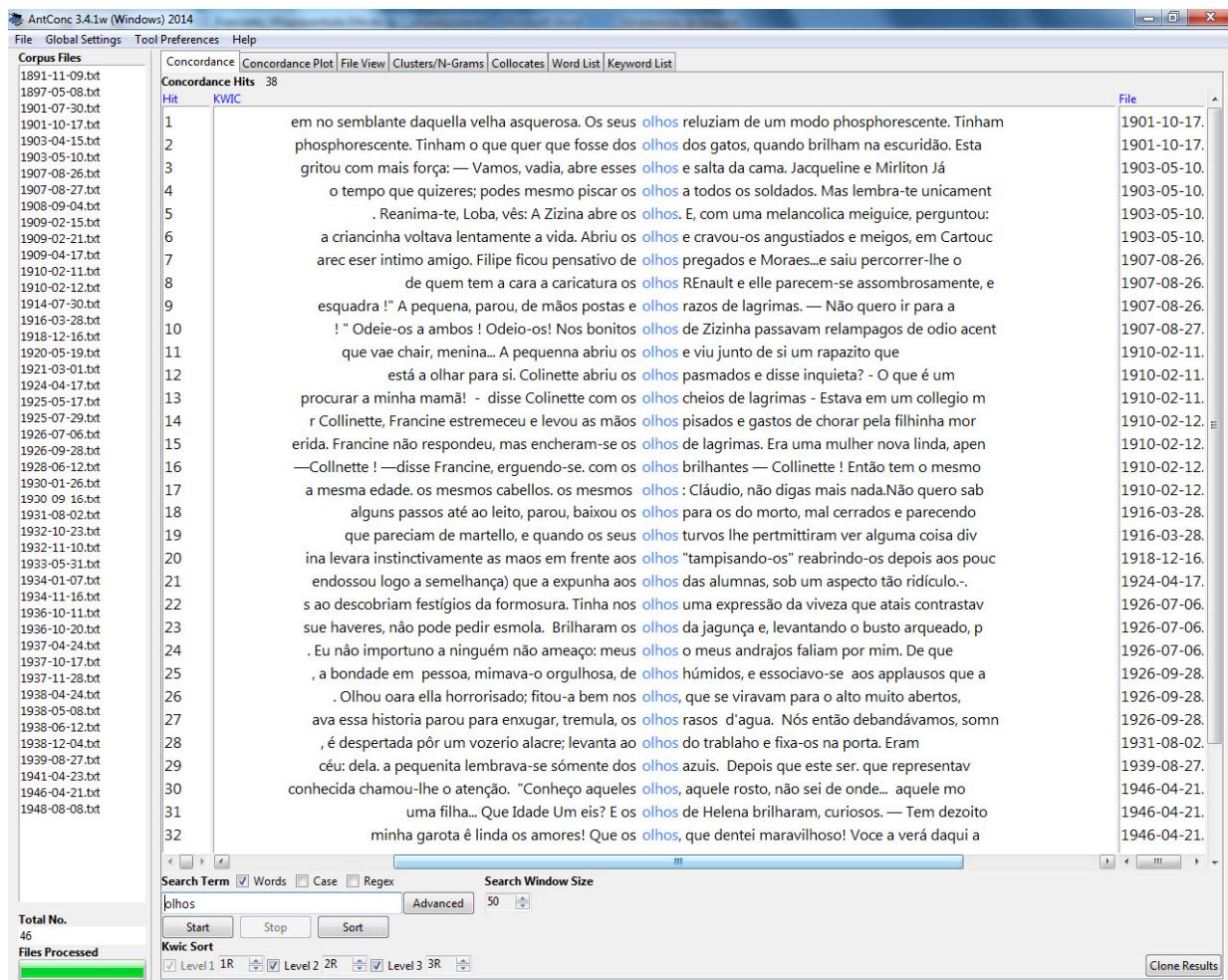


Figura 17. Contextos de uso do lexema OLHO (1891 – 1950)

MÃE: Este lexema compartilha dois campos semânticos: o de mulher e o de religião (mãe de Deus).

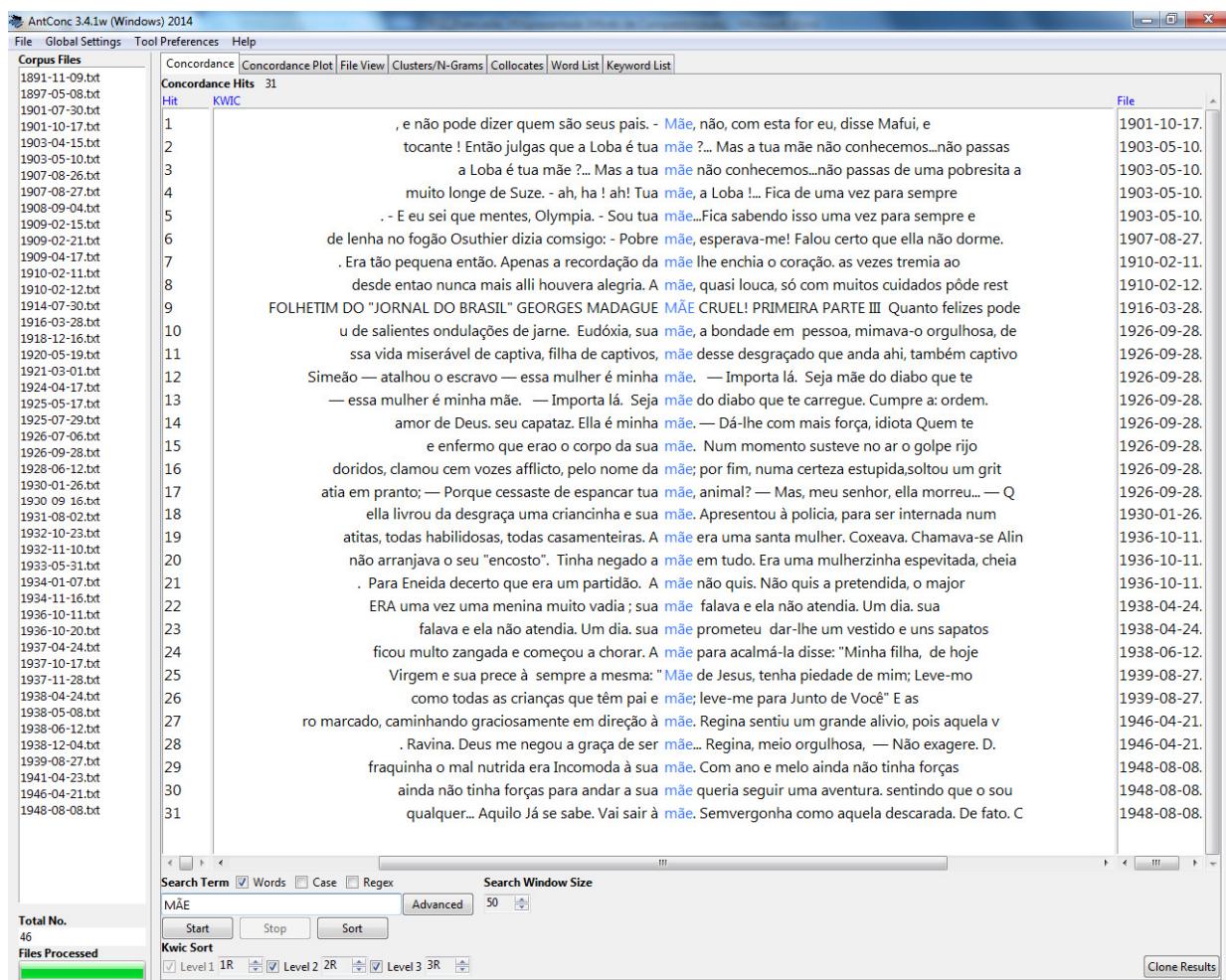


Figura 18. Contextos de uso do lexema MÃE (1891 – 1950)

RUA, NOITE e POBRE: Estes lexemas remetem ao campo semântico de vadiagem. A ausência de atividade laboral ou o não atendimento às atividades escolares era considerado vadiagem, por isso se usa o termo *vadia*.

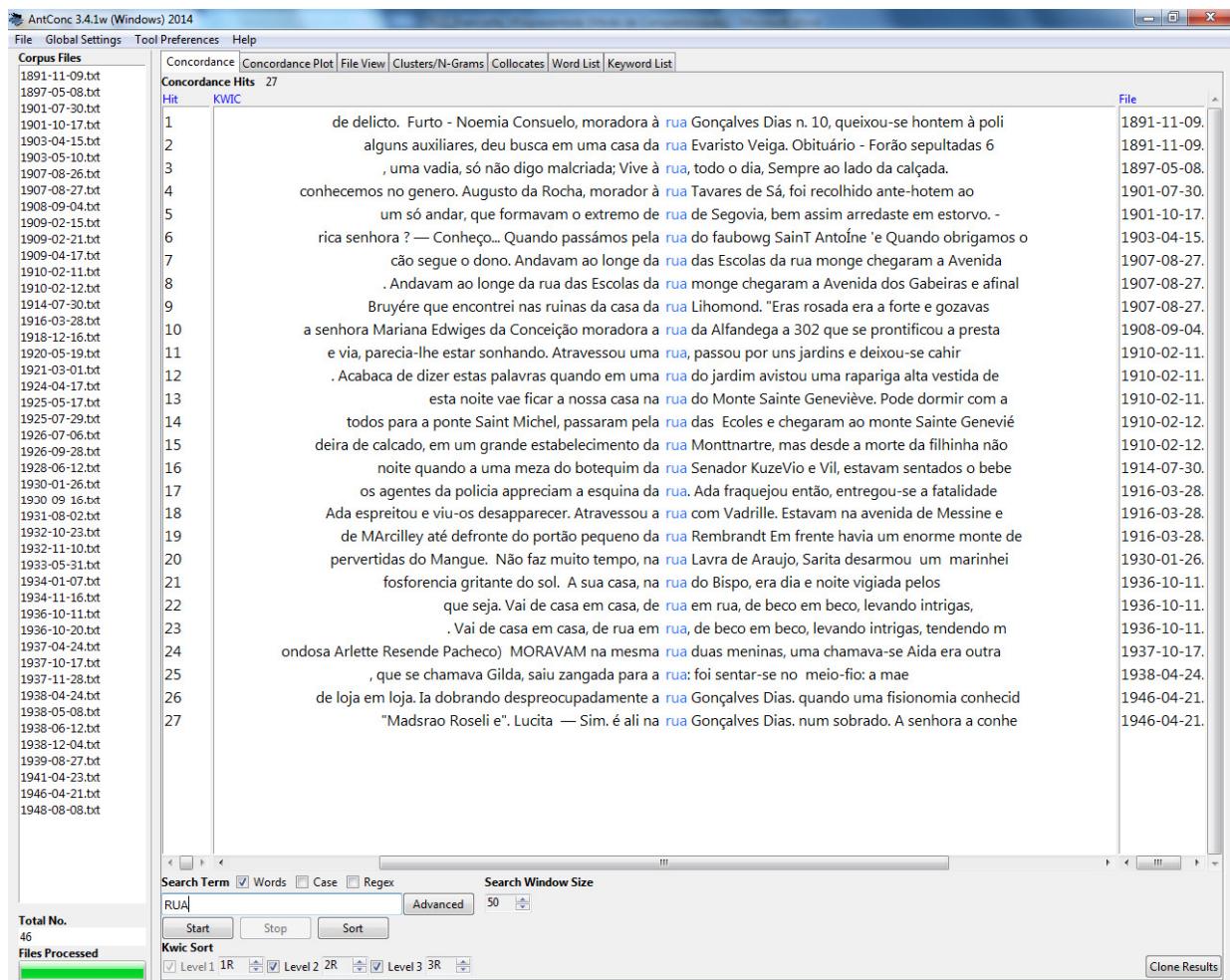


Figura 19. Contextos de uso do lexema RUA (1891 – 1950)

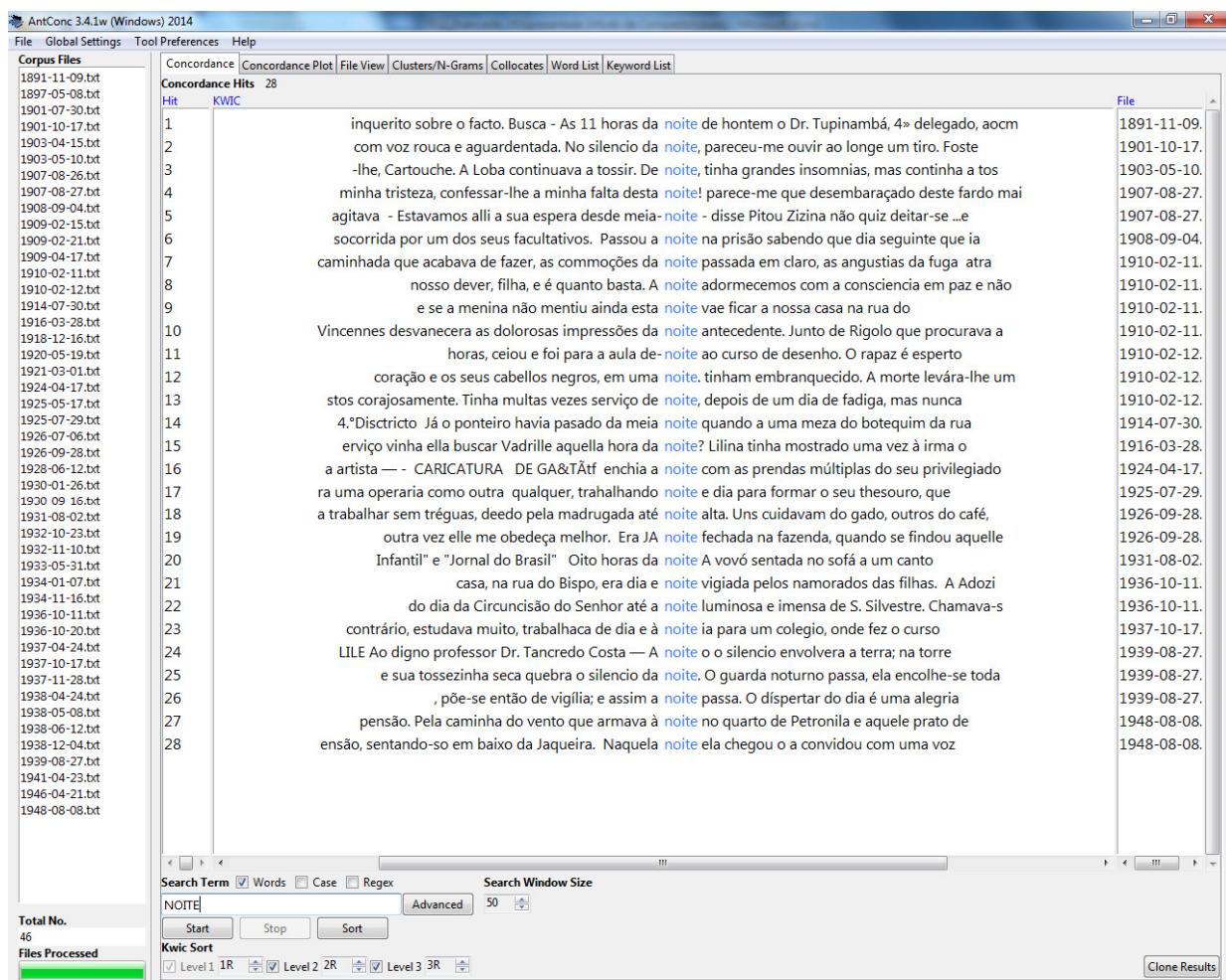


Figura 20. Contextos de uso do lexema NOITE (1891 – 1950)

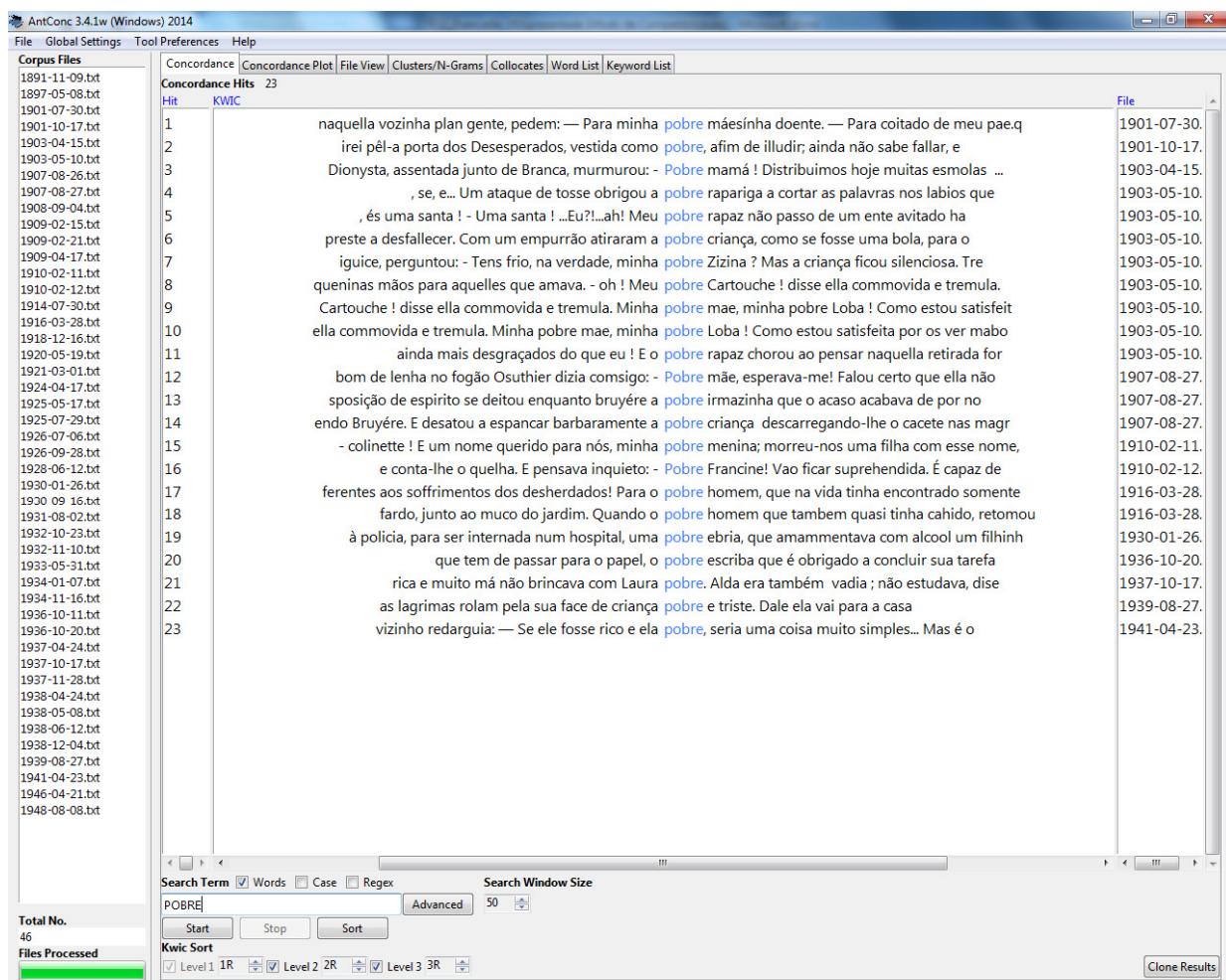


Figura 21. Contextos de uso do lexema POBRE (1891 – 1950)

TEMPO: Lexemas de marcas temporais podem ser interpretados, neste contexto, no sentido de “gastar o tempo” ou “perder o tempo”. No *corpus* analisado, remete também ao campo semântico de vadiagem.

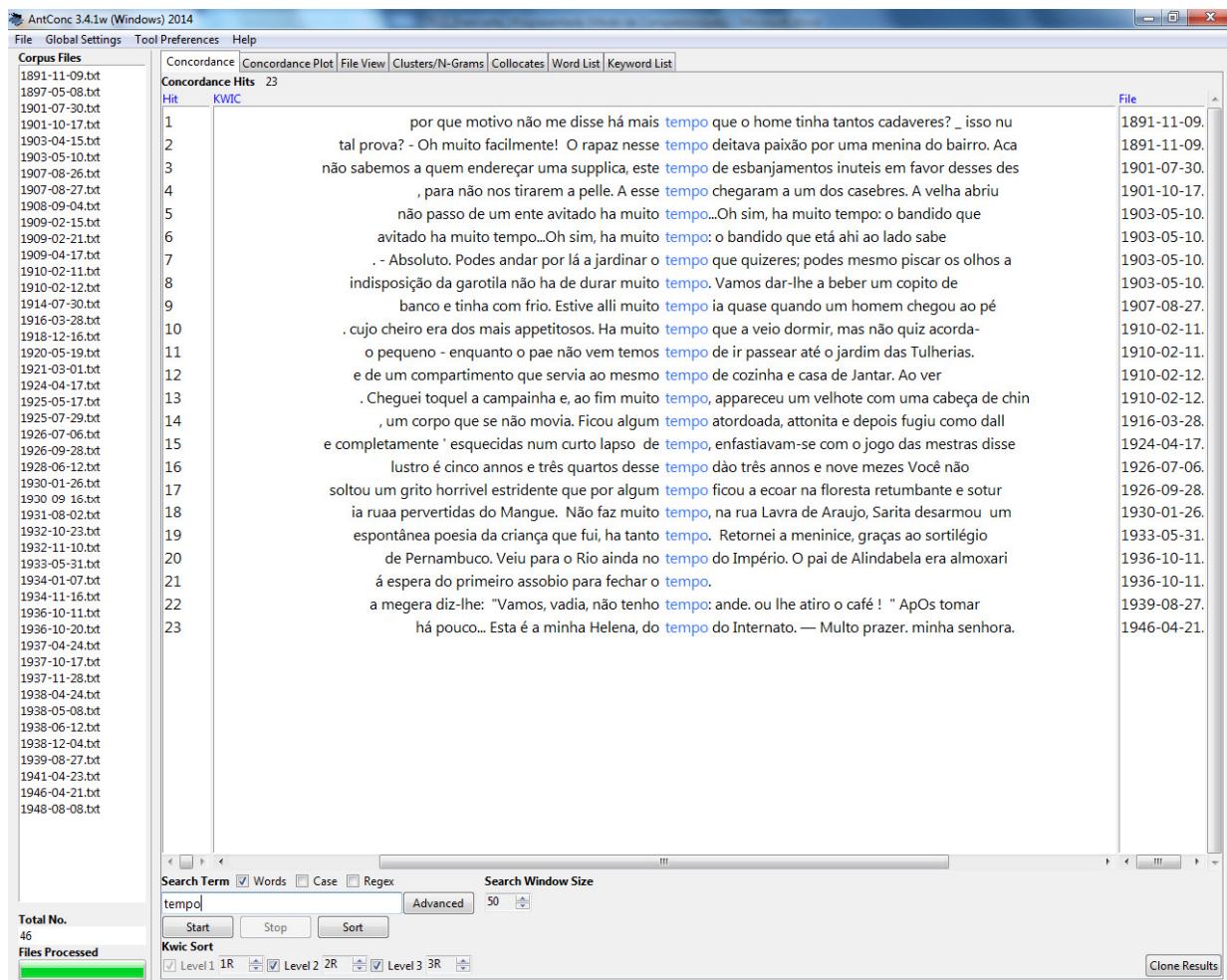


Figura 22. Contextos de uso do lexema TEMPO (1891 – 1950)

Grupo B:

FILME e HISTÓRIA: Estes lexemas remetem à ficção. O aparecimento dos lexemas HISTÓRIA E FILME no grupo B nos leva a levantar a seguinte hipótese: *vadia* como neologismo de sentido estando associado ao campo da moral teria ocorrido inicialmente no campo da ficção? Tal hipótese pode ser analisada em trabalhos posteriores, com mais cuidado.

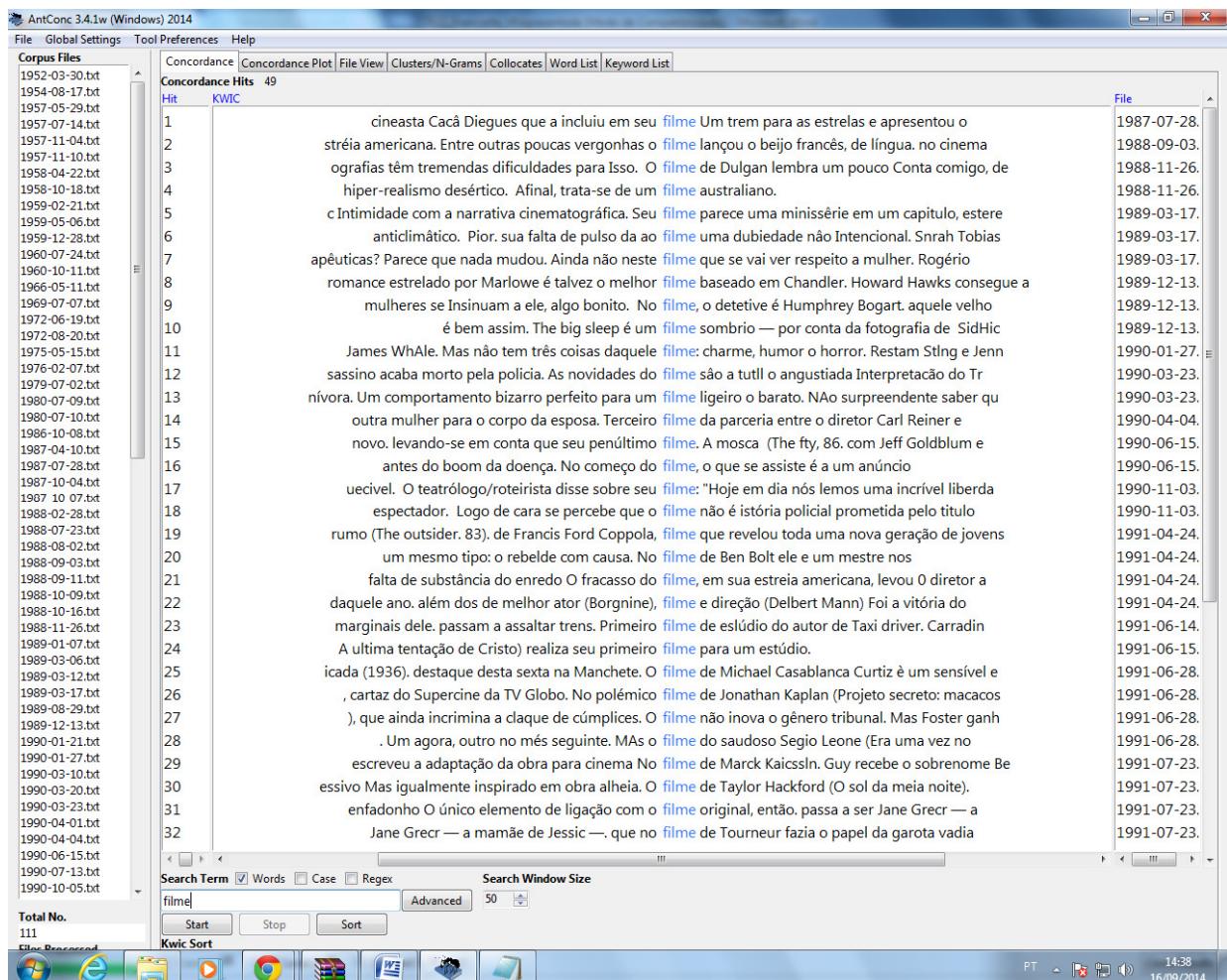


Figura 23. Contextos de uso do lexema FILME (1951 – 2010)

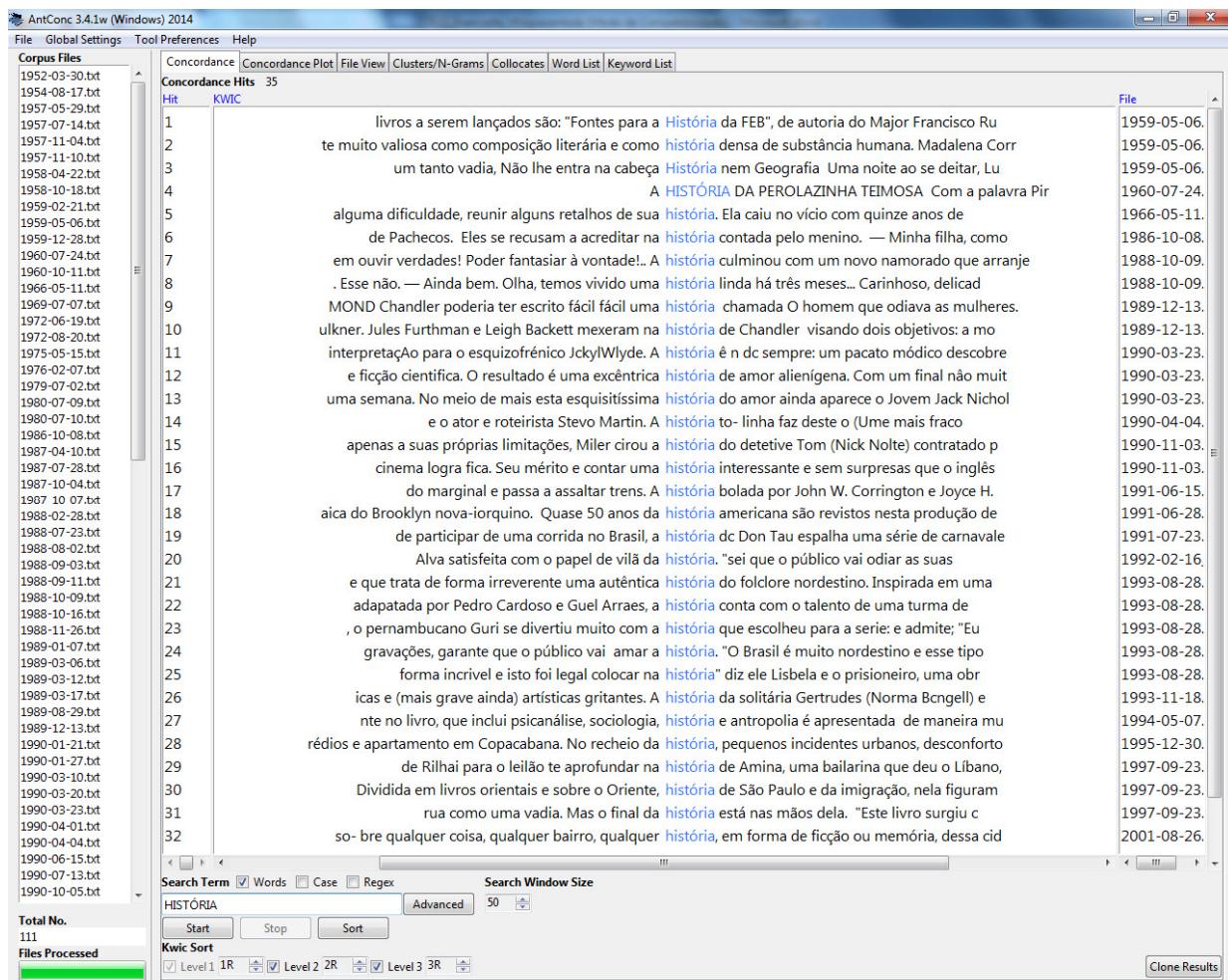


Figura 24. Contextos de uso do lexema HISTÓRIA (1951 – 2010)

VIDA: Este lexema está associado a tipo de vida. Na maioria dos casos, algum tipo de comportamento que transgride a moral. É interessante notar a ocorrência das seguintes expressões que remetem, todas, à ideia de mulher promíscua: *mulher de vida fácil, garota da vida fácil, mulher da vida, vida escandalosa, vida de muitas farras e amores, meninas de vida fácil, vida cigana*.

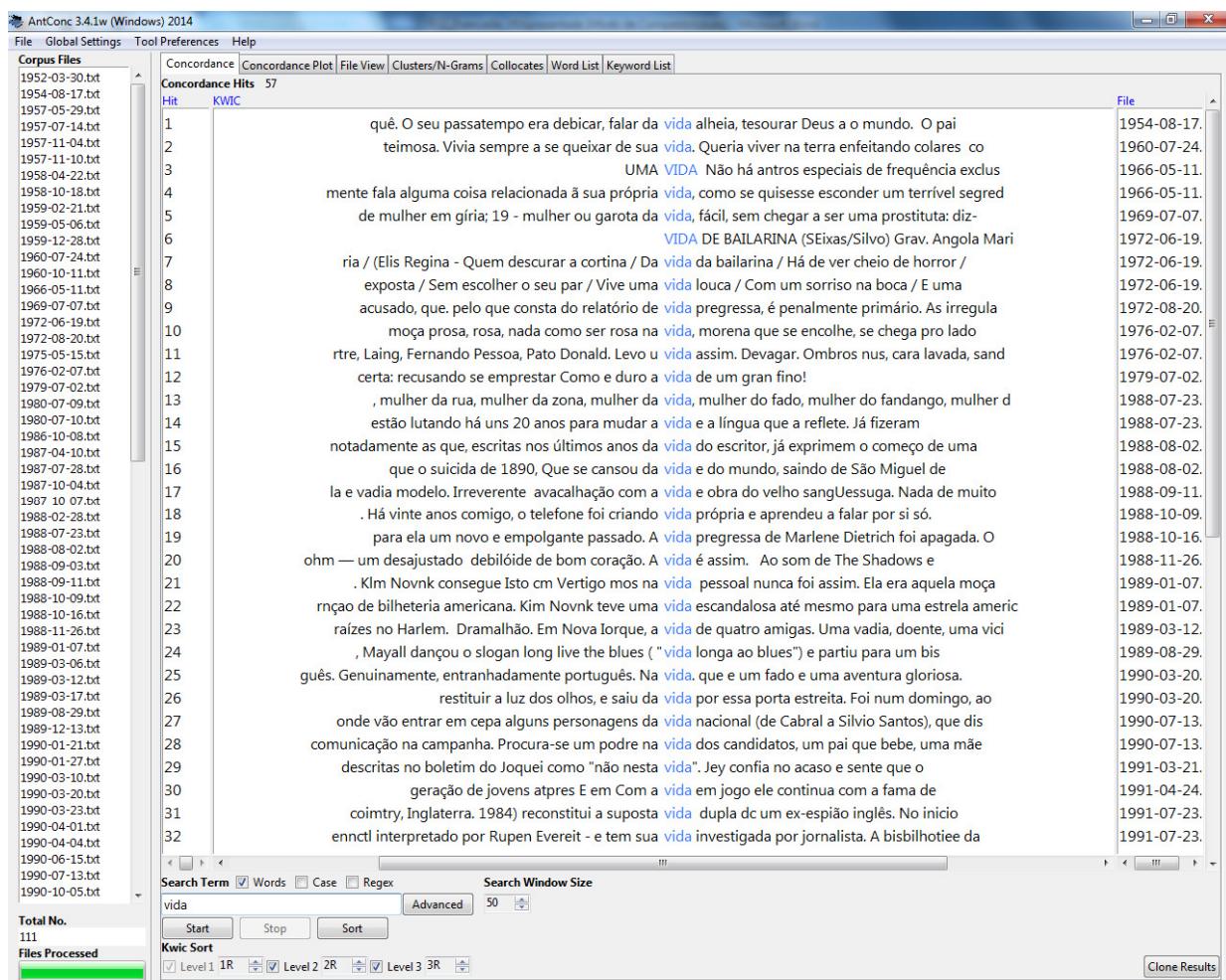


Figura 25. Contextos de uso do lexema VIDA (1951 – 2010)

HOMEM: Este lexema, associado ao lexema de MULHER, em contextos com o termo *vadia* remete a promiscuidade.

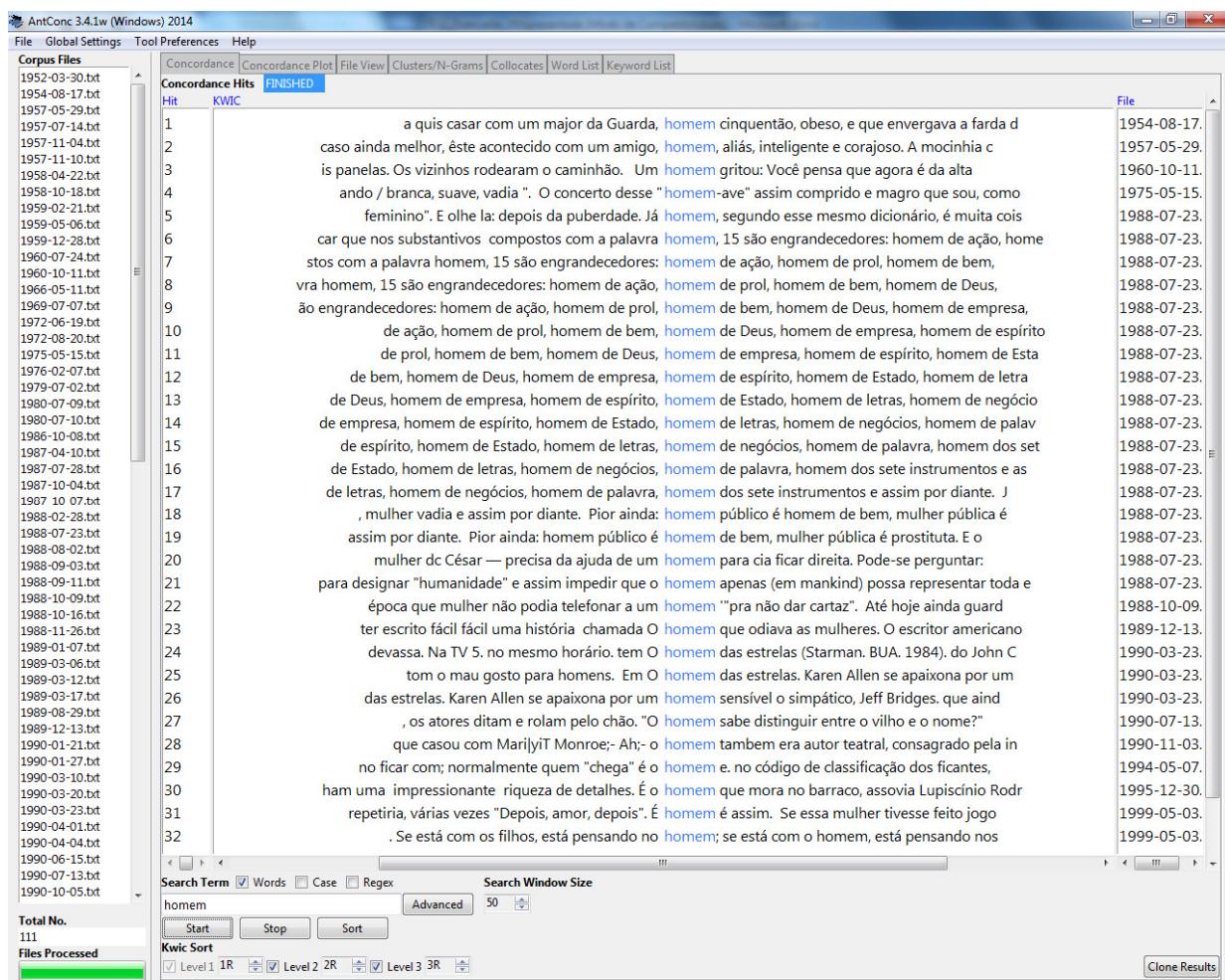


Figura 26. Contextos de uso do lexema HOMEM (1951 – 2010)

MUNDO: É curioso o aparecimento, junto a este lexema, de expressões como *mulher do mundo*, *rainha do burlesco*, *miss mundo exótico* e *vagando pelo mundo*. As três primeiras remetem aos campos semânticos de mulher e vadiagem.

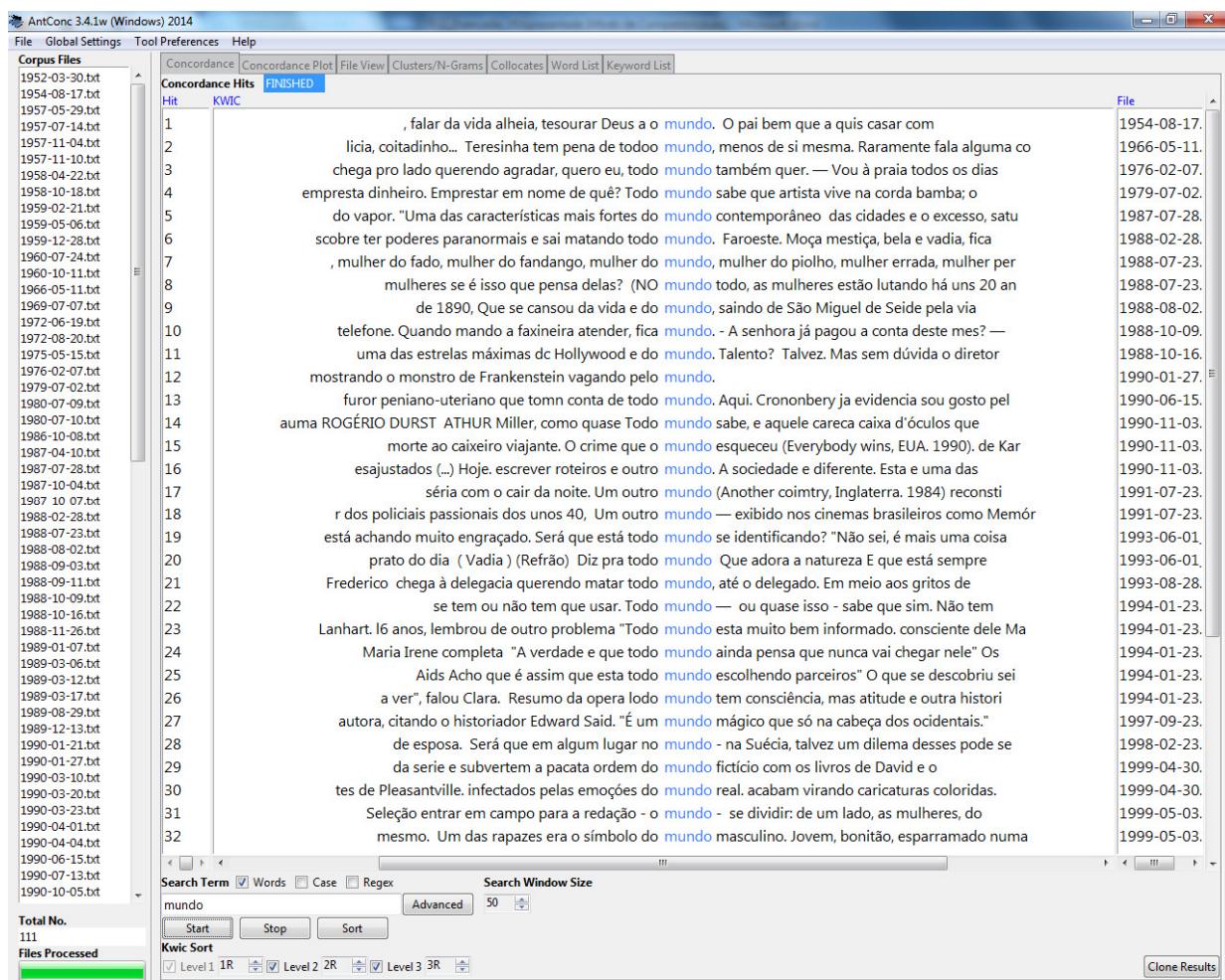


Figura 27. Contextos de uso do lexema MUNDO (1951 – 2010)

COISA: Partilha de características de vários campos semânticos.

Os contextos de usos apontaram para diferentes campos semânticos; porém, pode-se identificar uma relação forte em contextos de uso relacionados à ficção.

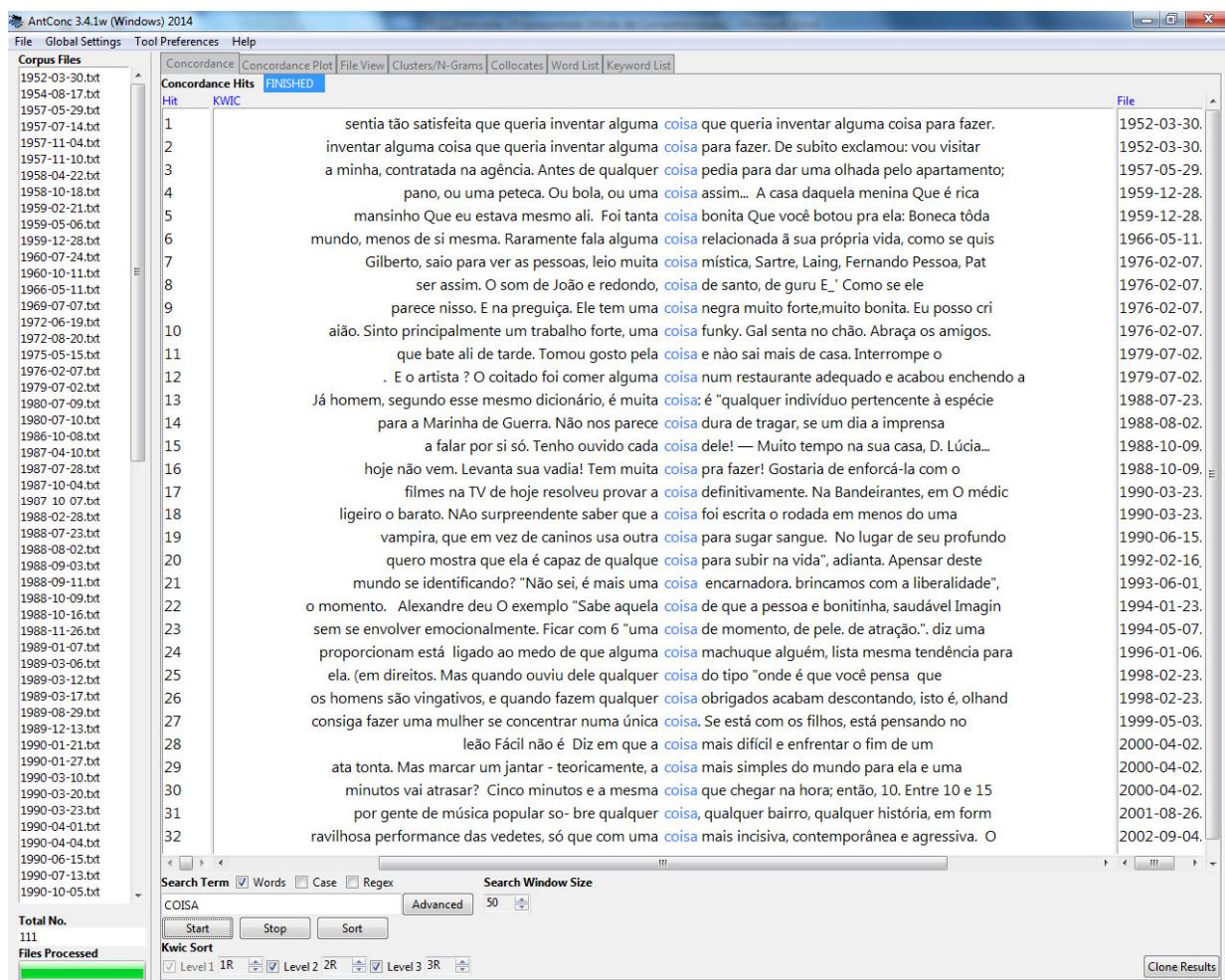


Figura 28. Contextos de uso do lexema COISA (1951 – 2010)

Além dos grupos A e B, é interessante analisarmos o grupo de intercessão. Este grupo comprehende os lexemas que foram recorrentes durante todo período analisado, ou seja, durante os cento e vinte anos (1891-2010).

Nos primeiros sessenta anos (1891-1950), as ocorrências mais frequentes do termo *vadia* fazem parte do campo semântico da vadiagem. Além disso, a associação do termo *vadia* a mulher não é exclusiva, ou seja, esse termo é atribuído também a crianças e coisas. Um exemplo está na linha 2, da figura 29 (abaixo): “uma vadia, só não digo malcriada”.

O contexto relacionado à polícia também é mais recorrente no primeiro período. Na linha 18, da figura 29 há: “com a polícia que a tomardo por uma vadia”. Neste

contexto, *vadia* é considerada a mulher que cometeu o crime de vadiagem¹¹. Para explicar esse fenômeno da vadiagem, Carvalho (2006, p.6) diz:

A identidade marginal do vagabundo é fruto do desenvolvimento do capitalismo, visto que este proporcionou o advento desse personagem ao causar a exclusão de uma vasta parcela de mão-de-obra do processo produtivo hegemonic. A sub-classe trabalhadora vagava pelas ruas do centro exercendo atividades que ofendiam os novos valores da ideologia capitalista, muitas vezes cometiam pequenos delitos, ocasionando o desordenamento do espaço público da cidade. Era alvo do aparato repressivo do Estado, que reprimia os vadios e desordeiros que agrediam o ordenamento social.

Chama ainda a atenção nesse período é a recorrência do lexema MENINA e apenas uma ocorrência do lexema MULHER.

Outro fenômeno interessante são os campos semânticos de *escola* e de *policia* nesse primeiro período: lexemas como MALCRIADA, IGNORANTE, MENINA, COLEGIAL, CASTIGO, ESCOLA, DIABINHA, LEVADA e JULGAR remetem ao campo semântico de *escola*; já os lexemas como DESGRAÇA, PROCESSAR, XADREZ, DESORDEIRA, POLÍCIA, LADRA e DETENÇÃO estão dentro do campo semântico de *policia*.

Neste primeiro, 1890 – 1950, houve a ocorrência de 56 lexemas VADIA (ver figura 29).

¹¹Crime que enquadrava os sub-trabalhadores no artigo 399 do Código Penal de 1890, que tratava da vadiagem. No entanto, essas ocupações eram a forma pela qual o exército de excluídos conseguiu sobreviver diante do quadro de miserabilidade destinado a estes pela nova ordem econômica. (CARVALHO, 2006, p.7).



Figura 29. Contextos de uso de *vadia* no primeiro período (1891–1950)

Para o segundo período analisado, 1951 – 2010, o termo *vadia* sofre uma mudança. O aumento da ocorrência do lexema MULHER é significativo.

Outro aspecto importante para esse período é serem recorrentes lexemas que evidenciam o corpo e a sensualidade da mulher (BELA, SENSUAL, SEREIA, SEDUTORA, MUSA, BELDADE) além dos lexemas que remetem ao universo da promiscuidade (CABARÉ, PROFISSÃO, PERDIDA, QUARTO, PROSTITUTA, VAGABUNDA)e o universo da sexualidade: ENCOCHAR e SAFADA.

Outro dado interessante diz respeito às denominações animalescas para se referir à vadia, tais como os lexemas GALINHA, GATA e CADELA, que representariam práticas promíscuas.

Além disso, os lexemas como ALCOÓLATRA, VICIADA e HOMOSSEXUAL são usados e compartilham o mesmo campo semântico: o da moral.

Os lexemas HISTÉRICA e LUNÁTICA apareceram neste contexto, também e são termos igualmente pejorativos.

Outro fenômeno interessante, neste período, é o aparecimento dos lexemas PRECONCEITO e ESTEREÓTIPO. Tais lexemas podem ser associados à idéia de práticas sociais descriminalizadas.

Para o segundo período analisado, 1951 – 2010 houve a ocorrência de 111 lexemas *vadia*. Comparado ao primeiro período analisado, 1891 – 1950 há um salto significativo. Neste primeiro período. Houve apenas 56 ocorrências de vadia, já no segundo período esse número quase dobrou, sendo 111 ocorrências (ver figura 29 e 30).

File	Global Settings	Tool Preferences	Help				
Corpus Files	Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
	Concordance Hits	125	KWIC	Hit	File		
1952-03-30.txt	1	compreenderia que a cigarra... que você olha como vadia , preguiçosa, não é a formiga — infatigável, nem a mais ligeira aura de Beleza quis. vadia e efemera, perpassar Jamais. Por esta época trabalho forçado. Não desabrocha, irrefletida e vadia , como os passarinhos, entrava a cantar, entr "letra" de música que a indiana, alacre e vadia , como os passarinhos,		1952-03-30.			
1954-08-17.txt	2	Nunes, Olalo - Musa vadia : humorismo e s\xxE1tiras - B\xxE9lem.		1954-08-17.			
1957-05-29.txt	3	NUNES, Olavo — Musa vadia ; humorismos e s\xtiricas. Belém. Livraria Glob		1957-05-29.			
1957-07-14.txt	4	da sabedoria paga na Relação do Pôrto, por vadia , nesta época em que ela não tem que		1957-07-14.			
1957-11-04.txt	5	Policia Militar não podendo assim ser considerada vadia ou desocupada.		1957-11-04.			
1957-10-10.txt	6	usa Talia andava quebrantada, laissa enconchada, vadia e lastimosa, com a sua suficiência pecuniári		1957-11-10.			
1958-04-22.txt	7	, que é bem travessa E lá um tanto vadia , Não lhe entra na cabeça História nem Geogra		1958-04-22.			
1958-10-18.txt	8	vez me procura Não sou má, não sou vadia , levanto-me todo dia, E a rua indo		1958-10-18.			
1959-02-21.txt	9	muito estudosos e obediente, mas Pérrolazinha era vadia e teimosa. Vivia sempre a se queixar de		1959-02-21.			
1959-12-28.txt	10	SUA NEGRA VADIA "Não causou surpresa primeiro ato de Caroll		1959-05-06.			
1960-07-24.txt	11	que agora é da alta classe, sua negra vadia ? Você escreveu sobre nós, ganhou um monte d		1959-12-28.			
1960-10-11.txt	12	companhia de sua mãe? — Ela é muito mais vadia do que eu... chegar a ser uma prostituta: diz-se também vadia ; 20 - quarentão; 21 - Sinistra;22 - Corrupt		1960-07-24.			
1966-05-11.txt	13	/ Como um lirio em lamaçal / uma sereia vadia / Prepara em noite de orgia / O seu drama		1960-10-11.			
1969-07-07.txt	14	justo, pois a acusada não era comprovadamente uma vadia , pois trabalhava numa empresa Industrial. N		1966-05-11.			
1972-06-19.txt	15	a vazia / a criança engatinhando / branca, suave, vadia ". O concerto desse "homem-ave" assim compr		1969-07-07.			
1972-08-20.txt	16	, a força, a importância. E os contrastes. Menina vadia , de vez em quando carrega no batom, na		1972-06-19.			
1975-05-15.txt	17	falsa grá-fina, treinando sua futura profissão de vadia consentida pela distribuição de oportunidade		1972-08-20.			
1976-02-07.txt	18	para a 1º DP, onde foi qualificada como vadia contumaz. Só que ela apresentou documentos,		1975-05-15.			
1979-07-02.txt	19	, foi presa Tânia Carvalho O'Connor, acusada como vadia costumaie por Isso enquadrada no artigo 57		1976-02-07.			
1980-07-09.txt	20	de plantão. Jose Carlos da Silveira Rocha, a vadia contumaz (Tânia) "vagava em ociosidade e não		1979-07-02.			
1980-09-03.txt	21	o Persico — presa na Praça Mauô, anteontem, como vadia contumaz, denunciou, ontem, ao Juiz da 261		1980-07-09.			
1980-10-09.txt	22	sua idade, gostava de baile, mas não era vadia — afirmava a mãe de Kátia, Sebastiana dos Sa		1980-07-09.			
1980-10-16.txt	23	ter aula no colégio, mas que não seja vadia ; que tenha imaginação; que seja feliz. Para:		1980-07-09.			
1988-11-26.txt	24	Gueixa Vadia (Fausto Fawcett/Carlos Laufer) E no meio da		1980-07-09.			
1989-01-07.txt	25	pa. formada por meninhas neoMadonnas. Em Gueixa vadia , um turista cientista marginal americano e s		1987-07-28.			
1989-03-06.txt	26	ter aula no colégio, mas que não seja vadia ; que tenha imaginação; que seja feliz. Para:		1987-07-28.			
1989-03-12.txt	27	ter aula no colégio, mas que não seja vadia , que tenha imaginação, que seja feliz. Para:		1987-10-04.			
1989-03-17.txt	28	tando todo mundo. Faroeste. Moça mestiça, bela e vadia , fica dividida entre o amor de dois irmãos,		1987-10-07.			
1989-04-17.txt	29	do piolho, mulher errada, mulher perdida, mulher vadia e assim por diante. Pior ainda: homem p\xfbl		1988-02-28.			
1989-11-04.txt	30	sabedoria pag\xe1 presa na Relação do Porto, por vadia , nesta época em que ela não tem que		1988-07-23.			
1989-11-10.txt	31	egory Pcck. Produção americana. Mestiça um tanto vadia (Jones) não consegue se decidir entre dois i		1988-08-02.			
1989-10-22.txt	32	Nova Iorque onde se apaixona por bela e vadia modelo. Irreverente avacalhação com a vida		1988-09-03.			
1989-10-18.txt	33	horas. A faxineira hoje não vem. Levanta sua vadia ! Tem muita coisa pra fazer! Gostaria de enfo		1988-09-11.			
1989-05-06.txt	34	professor seduzido por uma dona boa. sensual e vadia . Nenhuma das grandes atrizes locais o agrad		1988-10-09.			
1989-12-28.txt	35	se apaixonar por dançarina de cabaré um tanto vadia . P&B (98).		1988-10-16.			
1989-07-24.txt	36	a canalidade ao sobrenatural — a casa da falecida vadia da cidade o lida como assombrada. A fotogra		1988-11-26.			
1990-01-21.txt	37	estrela "ser uma dama na sala e uma vadia no quarto".		1989-01-07.			
1990-01-27.txt	38	&B. Judy (Scott) tem 18 anos. É mãe, viúva, vadia e rumava para um triste fim. Melodrama barato		1989-03-06.			
1990-03-10.txt	39	Nova Iorque, a vida de quatro amigas. Uma vadia , doente, uma viciada e uma bem sucedida. Dr		1989-03-12.			
1990-03-20.txt	40	agressão ela era fácil, drogada e um tanto vadia . Depois ela se torna responsável, decidida e		1989-03-17.			
1990-04-01.txt	41	triste desde que ela me deixou, amaldiçoou aquela vadia ". Nele estariam os temas fundamentais do		1989-08-29.			
1990-04-01.txt	42	e a mais nova, Carmen (Martha Vickers). É vadia e viciada Um empregado da família desaparece		1989-12-13.			
1990-04-10.txt	43	perfeita só que ela se revela uma grandessíssima vadia . Dramalhão. Após o suicídio dc sua esposa,		1990-01-21.			
1987-04-10.txt	44	(Beals) só que ela se revela um grandíssima vadia . Este é mais ou menos baseado em A		1990-01-27.			
1987-07-28.txt	45	a persona f\xflimica da atriz Warren é tão vadia que fica difícil acreditar que precise de ta		1990-03-10.			
1987-10-07.txt	46	da sabedoria pag\xe1 na Relação do Porto, por vadia , nesta época em que ela não tem que		1990-03-20.			
1988-02-28.txt	47	Turner faz a mocinha o Ingrid Borgman. a vadia . Em comum as beldades tom o mau gosto		1990-03-23.			
1988-07-23.txt	48	Comédia. Cirurgião casa com uma dona vadia e planeja transplantar o cérebro de outra mu		1990-04-01.			
1988-08-02.txt	49	. Cor (Km) . Cirurgião (Martin) casa com uma dona vadia (Turner) o planeja transplantar o cérebro de		1990-04-04.			
1988-09-03.txt	50	cobaia. Portanto, ela era uma tremenda de uma vadia , que passou na cara o. logo. contaminou meio		1990-06-15.			
1988-09-11.txt	51	dos candidatos, um pai que bebe, uma mãe vadia , um irmão homossexual. "Isso não é roubo, é		1990-07-13.			
1988-10-09.txt	52	Murray afundam a atriz no infeliz estereótipo da vadia de bom coração.		1990-10-05.			
1988-10-16.txt	53	acontece. O herói é um idiota; a mocinha, uma vadia histérica; o vilão, inexistente; e o psicop		1990-11-03.			
1988-10-26.txt	54	foras da lei, um bateote (Tesli), uma mocinha vadia (Frederick) um alcoólatra (PolLard) e um psic		1991-01-24.			
1988-11-26.txt	55	a considerá-lo um santo. E, claro, a vadia , Vickes (jenifer Tilly, irmã da Meg Tilly de		1991-03-21.			
1989-01-07.txt	56	da stripper interpretada por Diane Lane. Mas a vadia e cacho de seu maior rival na mesa		1991-04-24.			
1989-03-06.txt	57	inicia romance com balconista (Cowper) um tanto vadia . Ele parte para a guerra e, do front.		1991-05-19.			
1989-03-12.txt	58	. No inicio da década de 30, garota interiorana e vadia reencontra um velho namorado, hoje envolvido		1991-06-14.			
1989-08-29.txt	59	dc 30. garota (Hershey) do interior e duble de vadia reencontra um antigo namorado (Carradinc), u		1991-06-15.			
1989-12-13.txt	60	filme de Tourneur fazia o papel da garota vadia		1991-07-23.			
1990-01-21.txt	61						
1990-01-27.txt	62						
1990-03-10.txt	63						
1990-03-20.txt	64						
1990-03-23.txt							
1990-04-01.txt							
1990-04-04.txt							



Figura 30. Contextos de uso de *vadia* no segundo período (1951 – 2010)

A partir dos dados apresentados, verifica-se que, no primeiro período analisado, o termo *vadia* parece estar associado, não diretamente a mulher, e sim a pessoas que transgrediam alguma expectativa a elas atribuída ou cometiam crimes de

vadiagem; já no segundo período, essa concepção ainda está presente, porém outros campos semânticos se sobressaem: são eles os campos semânticos que remetem à sensualidade, ao corpo, a sexualidade, a promiscuidade, e também ao universo das bebidas e das drogas.

7.0 CONSIDERAÇÕES FINAIS

A partir da análise realizada, foi possível confirmar a hipótese de que houve mudança semântica no termo *vadia* ao longo dos séculos XIX a XXI motivada por mudanças sociais.

Nesse período, o Brasil foi palco de diversos movimentos sociais liderados por mulheres, que reivindicavam direitos iguais entre gêneros, o fim da violência e exploração das mulheres e o desvencilhamento da subordinação de um gênero a outro. A cada conquista dos grupos feministas, no âmbito social, promoveu-se mudança gradativa de paradigmas sociais e novas identidades sociais foram atribuídas às mulheres. À medida que as mulheres se inseriam nos grupos de liderança política e adquiriram direitos e voz social, comportamentos sociais atribuídos às mulheres eram modificados, criando novos comportamentos. Esse processo gerou certa inadequação aos modelos canônicos comportamentais de gênero. Devido a essa mudança paradigmática de comportamento a percepção da sociedade em relação às mulheres sofreu mudanças.

Essas mudanças refletiram-se no uso do termo *vadia*, empregado como referência a valores e práticas sociais que transgrediam a ordem social da época.

Além da investigação de um neologismo semântico do PB, outra proposta deste trabalho era apresentar um modelo que fosse capaz de realizar a lematização automática de lexias, ou seja, um processo de lematização de modo automático. Neste trabalho, devido à amplitude deste desafio e às limitações de um trabalho de monografia, foi proposto apenas um modelo teórico a fim de servir de modelo para projetos futuros. Para trabalhos futuros, um das questões devem ser implementadas, ou seja, tornar o modelo teórico em uma aplicação para o usuário, ou seja, um sistema especialista. Segundo Mendes (1997), “estrutura básica para um sistema especialista, é constituída por três elementos fundamentais: base de conhecimento, motor de inferência e interface com o usuário”.

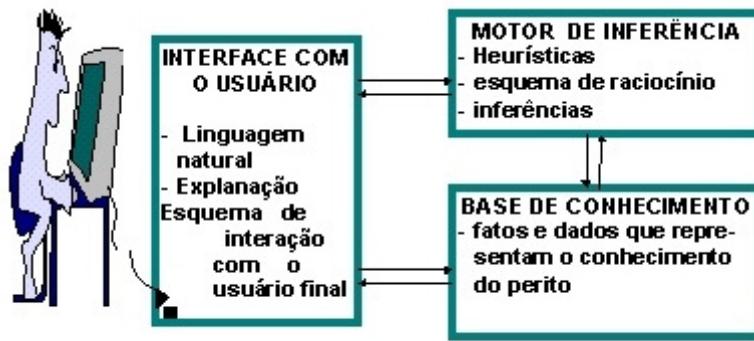


Figura 31. Estrutura básica de um sistema especialista.

No modelo teórico aqui apresentado, tratamos de duas das etapas que abarcam o desenvolvimento de um sistema especialista: base de conhecimento e motor de inferência. Para um trabalho futuro, a fase de interface com o usuário deve ser desenvolvida.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBADE, C. M. de S. A lexicologia e a teoria dos campos lexicais. **Cadernos do CNLF**, Rio de Janeiro, v. XV, n. 5. t 2, p. 1332-1343. 2011. Disponível em <http://www.filologia.org.br/xv_cnlf/tomo_2/105.pdf>. Acesso em 22 mar 2014.

AITCHISON, J. **Words in the mind: an introduction to the mental lexicon**. Oxford: Brasil Blackwell, 1987.

ALAN Turing Brasil 2012. Porto Alegre, 2012. Disponível em <<http://www.ufrgs.br/alanturingbrasil2012/>>. Acesso em 13 mai 2014.

ALENCAR, L. F. de. **Linguagem e Inteligência Artificial: na coletânea Linguagens. As expressões do Múltiplo**. Fortaleza: Premius, 2006.

ALLEN, J. F. **Natural language understanding**. Menlo Park: Benjamin Cummings, 1987.

ALVES, I. M. A integração dos neologismos por empréstimo ao léxico do português. **Alfa**, São Paulo, n. 28 (supl.), p. 119-126, 1984.

BAKHTIN, M. **Marxismo e filosofia da Linguagem**. São Paulo: Hucitec, 1988.

BASÍLIO, M. **Dissecando a palavra: formação e classes de palavras no português do Brasil**. São Paulo: Contexto, 2004.

BERBER SARDINHA, T. Linguística de *corpus*: histórico e problemática. Retrospectiva. **DELTA**, v. 16, n. 2, p. 323-367, 2000. Disponível em <<http://www.scielo.br/pdf/delta/v16n2/a05v16n2.pdf>> Acesso em 14 mar 2014.

BIDERMAN, M. T. C. Conceito linguístico da palavra. **Palavra**, Rio de Janeiro, v. 5, p.81-97, 1999.

BIDERMAN, M. T. **Teoria linguística: teoria lexical e linguística computacional**. 2. ed. São Paulo: Martins Fontes, 2001.

BIDERMAN, Maria Tereza Camargo. **Teoria Lingüística quantitativa computacional**. Rio de Janeiro: Livros técnicos e científicos, 1978. p.71 – 166.

BOURDIEU, P. **O poder simbólico**. Lisboa: Difel, 1989.

BRASIL. **Movimento feminista**. Governo Federal: Portal Brasil. 2013. Disponível em <<http://www.brasil.gov.br/cidadania-e-justica/2012/02/feminismo-pela-igualdade-dos-direitos>>. Acesso em 01 jun 2014.

BYBEE, J. **Morphology: a study of the relation between meaning and form**. Philadelphia: Benjamins, 1985.

CAMBRAIA, C. N. Da lexicologia social a uma lexicologia sócio-histórica: caminhos possíveis. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 21, n. 1, p. 157-

188, jan./jun. 2013 Disponível em <<http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/5096/>>. Acesso em 03 abr 2014.

CAMLONG, A. Lexicométrie. In: **Stablex version PC**. São Paulo: Pirus Tecnologia, 2004. CD-ROM.

CAMLONG, A. **Méthode d'analyse lexicale textuelle et discursive**. Paris: C.R.I.C. & OPHRYS, 1996.

CAMLONG, A. **Stablex pratique**. Toulouse: Teknea, 1991.

CARVALHO, M. V. **Vadiagem e criminalização: a formação da marginalidade social do Rio de Janeiro de 1888 a 1902**. Usos do Passado — XII Encontro Regional de História ANPUH-RJ. 2006.

CAVALCANTI, E. G. **A estatística e a semiótica: imbricação de olhares sobre textos neotestamentários**. São Paulo, 2005. Dissertação (Mestrado em Letras - Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

CORREIA, M. Os dicionários são todos iguais? In: **Os dicionários portugueses**. Bahia: Caminho, 2009.

CÚRCIO, V. R. Estudos estatísticos de textos literários. **Texto Digital**, Florianópolis, v. 2, n. 2, 2006. Disponível em <<http://www.textodigital.ufsc.br/num03/veronica.htm>> Acesso em 3 mar 2014.

DI FELIPPO, A.; DIAS-DA-SILVA, B. C. (2006). Dos olhares sobre o léxico: diferenças e semelhanças. In: DIAS-DA-SILVA, B.C.; LONGO, B.N.O. (Orgs.) **A construção de dicionários e bases de conhecimento lexical**. Araraquara/São Paulo: Laboratório Editorial FCL-UNESP/Cultura Acadêmica, 2006.p. 169-185, (Trilhas Linguísticas, 9). Disponível em <http://www.geterm.ufscar.br/ariani/Dos_olhares_sobre_o_lexico.pdf>. Acesso em 10 mai 2014.

DIAS-DA-SILVA, B. C. **A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais**. Tese de doutorado em Letras na área de concentração linguística e língua portuguesa – Faculdade de Ciências e Letras da Universidade Estadual Paulista. Araraquara, 1996. 272p. Disponível em <<http://wiki.icmc.usp.br/images/a/ad/DiasDaSilva1996.pdf>>. Acesso em 03 mar 2014.

DUQUE, J. I. **Um processo baseado em parágrafos para extração de tratamentos de artigos científicos do domínio biomédico**. 121f. Dissertação (Mestrado em Ciência da Computação). Departamento de Ciência da Computação, Universidade Federal de São Carlos. São Paulo. 2012. Disponível em <http://www.bdtd.ufscar.br/htdocs/tedeSimplificado/tde_arquivos/3/TDE-2012-05-16T120945Z-4409/Publico/4310.pdf>. Acesso em 13 mar 2014.

EBECKEN, N.F.F; LOPES, M.C.S; COSTA, M.C.D. A mineração de textos. In: REZENDE, S. O. (Ed.). **Sistemas inteligentes: fundamentos e aplicações**. São Carlos: Manole, 2003. p.337-370.

FLICKINGER, D.; POLLARD, C.; WASOW, T. Structure sharing in lexical representation. In: **Proceedings of the 23rd annual meeting of the Association for Computational Linguistics**. Morristown, N.J.: Association for Computational Linguistics, 1985.

GRITTI, L. L. Os campos semânticos e o processamento cognitivo. **Cadernos de Letras da UFF**, Niterói, n. 41, p. 137-148. 2010. Disponível em <<http://www.uff.br/cadernosdeletrasuff/41/artigo6.pdf>>. Acesso em 20 abr 2014.

GUARIGLIA, R. Palavras e práticas discursivas. **Estudos Linguísticos**. n. XXXV, p. 1695-1699, 2006. Disponível em <<http://www.gel.org.br/estudoslinguisticos/edicoesanteriores/4publica-estudos-2006/sistema06/42.pdf>>. Acesso em 20 abr 2014.

LAROCA, M. N. de C. **Manual de morfologia do português**. Juiz de Fora/Campinas: Editora da UFJF/Pontes, 2003.

LÉON, J. **A linguística de corpus: história problemas, legitimidade**. Filologia e Linguística Portuguesa, n.8, p. 51-81, 2006. Disponível em <<http://revistas.usp.br/flp/article/view/59745>>. Acesso em 6 mai 2014.

LYONS, J. **Semantics**. New York: Cambridge University Press, 1977.

MARCIEIRA, G. 'Marcha das Vadias' cobra fim de exploração sexual no RJ. **Jornal do Brasil**, Rio de Janeiro, 2011. Disponível em <<http://www.jb.com.br/rio/noticias/2011/07/02/marcha-das-vadias-cobra-fim-da-exploracao-sexual-no-rj>>. Acesso em 14 mai 2014.

MATORÉ, G. **La méthode en lexicologie: domaine français**. Nouv.éd. Paris: Didier, 1973.

MEL'ČUK, I. **Dependency syntax: theory and practice**. Albany, N.Y: The SUNY Press, 1988.

MENDES, R. D. Inteligência artificial: sistemas especialistas no gerenciamento da informação. **Ciência da Informação**, Brasilia, v. 26, n. 1, 1997. Disponível em <http://www.scielo.br/scielo.php?pid=S0100-19651997000100006&script=sci_arttext>. Acesso em 16 abr 2014.

MENEZES, C. de. Marcha das Mulheres arrasta multidão pelas ruas do Rio de Janeiro. **Jornal do Brasil**, Rio de Janeiro, 18 jun 2012. Disponível em <http://www.jb.com.br/ambiental/noticias/2012/06/18/marcha-das-mulheres-arrasta-multidao-pelas-ruas-do-rio-de-janeiro/>; acessado em 06.06.2014.

MICROSOFT, O. **Sobre o OCR (reconhecimento ótico de caractere)**. Microsoft Corporation. 2014. Disponível em <<http://office.microsoft.com/pt-br/help/sobre-o-ocr-reconhecimento-optico-de-caractere-HP003081255.aspx>> Acesso em 23 dez 2013.

NUNES, J. H. Lexicologia e lexicografia. In: GUIMARÃES, E.; ZOPPI-FONTANA, M. (orgs.). **Introdução às ciências da linguagem: a palavra e a frase**. Campinas: Pontes, 2006. p. 149-150.

OTHERO, G. A.; MENUZZI, S. M. **Linguística computacional: teoria e prática**. São Paulo: Parábola, 2005.

OTTO, C. O feminismo no Brasil: suas múltiplas faces. **Rev. Estud. Fem.**, v.12, n.2, p. 238-241,2004. Disponível em <<http://www.scielo.br/pdf/ref/v12n2/23971.pdf>>. Acesso em 01 jun 2014.

KNUTH, D. E. **The art of computer programming: fundamental algorithms**. Vol. 1. Massachusetts, USA: Addison-Wesley publishing company inc, 1973.

PEDRO, J. M. A experiência com contraceptivos no Brasil: uma questão de geração. **Revista Brasileira de História**, v.23, n.45, p. 239-260, 2003. Disponível em <<http://www.scielo.br/pdf/rbh/v23n45/16527.pdf>>. Acesso em 01 jun 2014.

PERINI, M. A. ReVEL na escola: semântica lexical. **ReVEL**, Belo Horizonte, v. 11, n. 20, 2013. Disponível em <<http://www.revel.inf.br/files/d82d8b71aefb021e101f3dd8a53f9af9.pdf>>. Acesso em 01 jun 2014.

RANCHHOD, E. O uso de dicionários e de autômatos finitos na representação lexical das línguas naturais. In: _____. (Org.). **Tratamento das línguas por computador: uma introdução à linguística computacional e suas aplicações**. Lisboa: Caminho, 2001. p. 13-47.

ROBIN, R. **História e linguística**. São Paulo: Cultrix, 1977.

RUSH, C. Cop apologizes for 'sluts' remark at law school. **Toronto Star**, Toronto, 18 fev. 2011. Disponível em <http://www.thestar.com/news/gta/2011/02/18/cop_apologizes_for_sluts_remark_at_la_w_school.html>. Acesso em 05jun2014.

SAFFIOTTI, H. I. B. **Gênero, patriarcado, violência**. São Paulo: Fundação Perseu Abramo, 2004.

SANTOS, C. M.; IZUMINO, W. P. Violência contra as mulheres e violência de gênero: notas sobre estudos feministas no Brasil. **Revista E.I.A.L. Estudios Interdisciplinarios de América Latina y El Caribe**, TelAviv, 2005. Disponível em <<http://www.nevusp.org/downloads/down083.pdf>>. Acesso em 07 mai 14.

SAUSSURE, F. de. **Curso de linguística geral**. 2. ed. Tradução de A. Chelini et al. São Paulo: Cultrix/Editora da USP, 1972.

SILVA, Bento Carlos Dias da; MONTILHA, Gisele; SPECIA, Lucia Helena Machado Rino; NUNES, Lucia Maria das Graças Volpe; JUNIOR, Osvaldo Novais de Oliveira; MARTINS, Ronaldo Teixeira; PARDO, Thiago Alexandre Salgueiro. **Introdução ao processamento de línguas naturais e algumas aplicações.** São Carlos: Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional NILC - ICMC-USP, 2007. Disponível em <<http://www.letras/etc.br/ebralc/NILCTR0710-DiasDaSilvaEtAl.pdf>>. Acesso em 07 maio 2014.

TIMOTEO, C. Q. As transformações do movimento feminista no Brasil e sua relação com a América Latina. **Anais do V Simpósio Internacional Lutas Sociais na América Latina “Revoluções nas Américas: passado, presente e futuro”.** GT 7. Feminismos, sexualidades e marxismos na América Latina. 2013. Disponível em <http://www.uel.br/grupo-pesquisa/gepal/v7_carolina_GVII.pdf>. Acesso em 15 abr 2014.

ZAPPAROLI, Z. M.; CAMLONG, A. **Do léxico ao discurso pela informática.** São Paulo: EDUSP/FAPESP, 2002.

Anexo I : Lista das 1400 lexias mais frequentes no *corpus*

Ordem	Lexias	Frequência
1.	de	2439
2.	a	2189
3.	e	1857
4.	o	1703
5.	que	1543
6.	do	841
7.	da	755
8.	um	733
9.	se	684
10.	não	663
11.	uma	638
12.	com	623
13.	para	581
14.	em	580
15.	no	476
16.	é	452
17.	na	437
18.	os	431
19.	por	373
20.	as	312
21.	como	296
22.	mais	286
23.	mas	268
24.	ao	260
25.	dos	222
26.	sua	219
27.	seu	192
28.	vadia	184
29.	foi	181
30.	era	165
31.	me	157
32.	quando	150
33.	das	147

34.	ela	144
35.	tem	142
36.	nos	141
37.	eu	137
38.	Ihe	134
39.	muito	127
40.	à	121
41.	ser	117
42.	está	114
43.	ou	114
44.	ele	111
45.	anos	108
46.	pela	108
47.	sem	106
48.	bem	104
49.	pelo	103
50.	dia	101
51.	só	101
52.	mulher	100
53.	já	99
54.	quem	98
55.	casa	97
56.	mesmo	94
57.	depois	92
58.	são	92
59.	ainda	91
60.	até	90
61.	nem	84
62.	tinha	81
63.	vida	81
64.	meu	79
65.	onde	77
66.	vez	77
67.	todos	74

68.	minha	73
69.	tudo	73
70.	isso	71
71.	assim	70
72.	seus	70
73.	disse	69
74.	ella	68
75.	nada	67
76.	sobre	67
77.	então	66
78.	vai	66
79.	aos	65
80.	também	65
81.	ter	65
82.	você	64
83.	porque	63
84.	nas	62
85.	entre	61
86.	noite	60
87.	agora	59
88.	esta	59
89.	grande	58
90.	há	58
91.	sempre	58
92.	tão	58
93.	mundo	57
94.	mãe	55
95.	dois	54
96.	estava	53
97.	homem	52
98.	tempo	52
99.	todo	52
100.	dc	51
101.	ficar	51
102.	filme	51
103.	outro	51
104.	suas	51
105.	fazer	50
106.	meio	50
107.	lá	49
108.	olhos	49
109.	num	47
110.	rio	47
111.	nunca	45
112.	pouco	45
113.	zizina	45

114.	deu	44
115.	dias	43
116.	apenas	42
117.	dar	42
118.	este	42
119.	pode	42
120.	coisa	41
121.	duas	41
122.	essa	41
123.	sou	41
124.	diz	40
125.	outros	40
126.	rua	40
127.	sabe	40
128.	vezes	40
129.	às	39
130.	ha	39
131.	hoje	39
132.	menina	39
133.	menos	39
134.	tanto	39
135.	antes	38
136.	faz	38
137.	numa	38
138.	pra	38
139.	todas	38
140.	bom	37
141.	deus	37
142.	história	37
143.	outra	37
144.	ver	37
145.	hora	36
146.	mulheres	35
147.	primeiro	35
148.	aqui	34
149.	esse	34
150.	mal	34
151.	toda	34
152.	amor	33
153.	havia	33
154.	melhor	33
155.	sim	33
156.	ali	32
157.	artista	32
158.	caso	32
159.	filha	32

160.	la	32
161.	nome	32
162.	pelos	32
163.	qualquer	32
164.	quer	32
165.	te	32
166.	maria	31
167.	parte	31
168.	quanto	31
169.	ah	30
170.	alguns	30
171.	amigo	30
172.	ano	30
173.	brasil	30
174.	cidade	30
175.	eles	30
176.	elle	30
177.	h	30
178.	livro	30
179.	maior	30
180.	pequena	30
181.	seja	30
182.	tenho	30
183.	tv	30
184.	vem	30
185.	desde	29
186.	fica	29
187.	marido	29
188.	mim	29
189.	quase	29
190.	talvez	29
191.	três	29
192.	c	28
193.	diretor	28
194.	dizer	28
195.	enquanto	28
196.	gente	28
197.	ir	28
198.	pessoas	28
199.	s	28
200.	conta	27
201.	lado	27
202.	loba	27
203.	logo	27
204.	oh	27
205.	pai	27

206.	parece	27
207.	pobre	27
208.	boa	26
209.	certo	26
210.	contra	26
211.	corpo	26
212.	mesma	26
213.	primeira	26
214.	cartouche	25
215.	estão	25
216.	fez	25
217.	ia	25
218.	longe	25
219.	mil	25
220.	momento	25
221.	nova	25
222.	º	25
223.	obra	25
224.	paulo	25
225.	policia	25
226.	senhor	25
227.	sob	25
228.	dinheiro	24
229.	ficou	24
230.	filho	24
231.	fim	24
232.	nós	24
233.	porta	24
234.	qual	24
235.	santos	24
236.	tal	24
237.	telefone	24
238.	trabalho	24
239.	annos	23
240.	camilo	23
241.	cara	23
242.	fausto	23
243.	homens	23
244.	junto	23
245.	lugar	23
246.	mão	23
247.	nao	23
248.	novo	23
249.	romance	23
250.	sei	23
251.	uns	23

252.	verdade	23
253.	á	22
254.	americana	22
255.	branco	22
256.	cada	22
257.	chega	22
258.	d	22
259.	dá	22
260.	falta	22
261.	música	22
262.	ninguém	22
263.	nosso	22
264.	personagem	22
265.	pois	22
266.	será	22
267.	velha	22
268.	voz	22
269.	algumas	21
270.	brasileira	21
271.	casamento	21
272.	cinema	21
273.	coração	21
274.	diante	21
275.	drama	21
276.	grupo	21
277.	isto	21
278.	lo	21
279.	nossa	21
280.	ouro	21
281.	outras	21
282.	pelas	21
283.	saber	21
284.	sol	21
285.	the	21
286.	vamos	21
287.	velho	21
288.	aquela	20
289.	claro	20
290.	criança	20
291.	dele	20
292.	foram	20
293.	horas	20
294.	jovem	20
295.	olympia	20
296.	semana	20
297.	sendo	20

298.	sociedade	20
299.	alguma	19
300.	carlos	19
301.	cláudio	19
302.	dona	19
303.	fora	19
304.	fosse	19
305.	globo	19
306.	john	19
307.	morte	19
308.	muitos	19
309.	nesse	19
310.	neste	19
311.	quatro	19
312.	quero	19
313.	rádio	19
314.	sr	19
315.	teve	19
316.	vão	19
317.	alli	18
318.	cabeça	18
319.	coisas	18
320.	crianças	18
321.	dela	18
322.	dr	18
323.	feito	18
324.	festa	18
325.	juiz	18
326.	não	18
327.	podia	18
328.	saiu	18
329.	si	18
330.	tenha	18
331.	tom	18
332.	tu	18
333.	amigos	17
334.	aquele	17
335.	caminho	17
336.	copacabana	17
337.	eua	17
338.	feira	17
339.	fino	17
340.	forma	17
341.	gosto	17
342.	jornal	17
343.	passou	17

344.	pedro	17
345.	pound	17
346.	regina	17
347.	tarde	17
348.	vou	17
349.	ai	16
350.	apesar	16
351.	após	16
352.	causa	16
353.	chegou	16
354.	cinco	16
355.	colinette	16
356.	desse	16
357.	escola	16
358.	estou	16
359.	filmes	16
360.	formiga	16
361.	frente	16
362.	frio	16
363.	meses	16
364.	murette	16
365.	muita	16
366.	n	16
367.	nenhum	16
368.	nesta	16
369.	olhar	16
370.	pae	16
371.	perguntou	16
372.	pessoa	16
373.	poder	16
374.	ponto	16
375.	quarto	16
376.	rigolo	16
377.	rosto	16
378.	segundo	16
379.	sentia	16
380.	sucesso	16
381.	tambem	16
382.	viu	16
383.	volta	16
384.	acaba	15
385.	além	15
386.	amanhã	15
387.	crime	15
388.	dessa	15
389.	desta	15

390.	estar	15
391.	ex	15
392.	falar	15
393.	final	15
394.	forte	15
395.	idade	15
396.	ja	15
397.	linda	15
398.	meia	15
399.	morreu	15
400.	muitas	15
401.	palavra	15
402.	palavras	15
403.	papel	15
404.	passa	15
405.	passar	15
406.	personagens	15
407.	problema	15
408.	processo	15
409.	produção	15
410.	programa	15
411.	r	15
412.	seguinte	15
413.	sido	15
414.	tomar	15
415.	vadrille	15
416.	vista	15
417.	voce	15
418.	alguém	14
419.	avenida	14
420.	clube	14
421.	cm	14
422.	condessa	14
423.	consegue	14
424.	cor	14
425.	cousa	14
426.	deste	14
427.	ê	14
428.	exemplo	14
429.	familia	14
430.	família	14
431.	filhos	14
432.	jogo	14
433.	justiça	14
434.	liberdade	14
435.	língua	14

436.	mãos	14
437.	mccain	14
438.	mestre	14
439.	nacional	14
440.	paris	14
441.	peça	14
442.	porém	14
443.	ribeiro	14
444.	seria	14
445.	sexo	14
446.	so	14
447.	terra	14
448.	tipo	14
449.	visto	14
450.	algum	13
451.	alva	13
452.	ar	13
453.	beleza	13
454.	cena	13
455.	certa	13
456.	chegar	13
457.	deixa	13
458.	dentro	13
459.	durante	13
460.	eneida	13
461.	época	13
462.	esses	13
463.	exclamou	13
464.	feliz	13
465.	força	13
466.	fui	13
467.	garota	13
468.	medo	13
469.	moça	13
470.	morta	13
471.	ontem	13
472.	pais	13
473.	pedra	13
474.	pena	13
475.	pequeno	13
476.	poderia	13
477.	próprio	13
478.	queria	13
479.	rica	13
480.	rosa	13
481.	silva	13

482.	tempos	13
483.	tres	13
484.	vae	13
485.	vive	13
486.	achou	12
487.	ahi	12
488.	algodão	12
489.	americano	12
490.	anda	12
491.	aquella	12
492.	baixo	12
493.	braços	12
494.	brasileiro	12
495.	campanha	12
496.	cigarra	12
497.	companhia	12
498.	costa	12
499.	costas	12
500.	década	12
501.	deixar	12
502.	delegado	12
503.	destino	12
504.	diferente	12
505.	dizia	12
506.	domingo	12
507.	dous	12
508.	embora	12
509.	entrar	12
510.	escritor	12
511.	estado	12
512.	fala	12
513.	fazendo	12
514.	fundo	12
515.	grandes	12
516.	helena	12
517.	historia	12
518.	janeiro	12
519.	livros	12
520.	longo	12
521.	meninas	12
522.	meus	12
523.	miss	12
524.	nenhuma	12
525.	peito	12
526.	pernas	12
527.	presa	12

528.	preto	12
529.	professor	12
530.	quê	12
531.	questão	12
532.	rapaz	12
533.	realmente	12
534.	relação	12
535.	ruas	12
536.	sair	12
537.	santa	12
538.	seis	12
539.	senhora	12
540.	sexual	12
541.	sul	12
542.	teatro	12
543.	têm	12
544.	unidos	12
545.	vestido	12
546.	acho	11
547.	alma	11
548.	anúncio	11
549.	atriz	11
550.	autor	11
551.	autora	11
552.	blues	11
553.	bonita	11
554.	céu	11
555.	começou	11
556.	cujo	11
557.	daquela	11
558.	deixou	11
559.	delegacia	11
560.	desses	11
561.	disco	11
562.	disso	11
563.	dormir	11
564.	encontrar	11
565.	espaço	11
566.	estados	11
567.	estavam	11
568.	estilo	11
569.	fácil	11
570.	fawcett	11
571.	federal	11
572.	grã	11
573.	humana	11

574.	jorge	11
575.	machado	11
576.	maneira	11
577.	manhã	11
578.	mayall	11
579.	menor	11
580.	minutos	11
581.	mora	11
582.	morrer	11
583.	morto	11
584.	ora	11
585.	ouvir	11
586.	passagem	11
587.	pé	11
588.	pensa	11
589.	português	11
590.	praça	11
591.	precisa	11
592.	público	11
593.	querida	11
594.	resultado	11
595.	rocha	11
596.	sangue	11
597.	santo	11
598.	segunda	11
599.	série	11
600.	soube	11
601.	televisão	11
602.	texto	11
603.	tinham	11
604.	vê	11
605.	veiu	11
606.	vivo	11
607.	abriu	10
608.	andava	10
609.	antonio	10
610.	apartamento	10
611.	arthur	10
612.	banda	10
613.	camisinha	10
614.	cantora	10
615.	carnaval	10
616.	chão	10
617.	contar	10
618.	continua	10
619.	conto	10

620.	correu	10
621.	dando	10
622.	david	10
623.	demais	10
624.	elles	10
625.	empregada	10
626.	estamos	10
627.	for	10
628.	fugir	10
629.	graça	10
630.	james	10
631.	joão	10
632.	josé	10
633.	jovens	10
634.	lamina	10
635.	lei	10
636.	levar	10
637.	luz	10
638.	meninos	10
639.	mesa	10
640.	namorada	10
641.	namorado	10
642.	naquela	10
643.	oito	10
644.	ordem	10
645.	perto	10
646.	policiais	10
647.	posso	10
648.	preta	10
649.	própria	10
650.	quis	10
651.	razão	10
652.	respondeu	10
653.	rural	10
654.	sete	10
655.	social	10
656.	triste	10
657.	vadiagem	10
658.	veio	10
659.	vinha	10
660.	ada	9
661.	afinal	9
662.	alegria	9
663.	almeida	9
664.	alto	9
665.	andar	9

666.	aquelle	9
667.	assunto	9
668.	atrás	9
669.	aula	9
670.	bairro	9
671.	bela	9
672.	bruyére	9
673.	cabral	9
674.	caderneta	9
675.	canto	9
676.	carioca	9
677.	carro	9
678.	certeza	9
679.	chamava	9
680.	cheia	9
681.	cima	9
682.	claudio	9
683.	clima	9
684.	comida	9
685.	conhecida	9
686.	dansar	9
687.	deve	9
688.	devia	9
689.	direito	9
690.	dollar	9
691.	eça	9
692.	elas	9
693.	entanto	9
694.	eram	9
695.	escolha	9
696.	espera	9
697.	estrelas	9
698.	finalmente	9
699.	flor	9
700.	fome	9
701.	francine	9
702.	grá	9
703.	humor	9
704.	igreja	9
705.	irmão	9
706.	jacqueline	9
707.	jamais	9
708.	jardim	9
709.	joaquim	9
710.	leva	9
711.	leve	9

712.	lhes	9
713.	loura	9
714.	m	9
715.	madrugada	9
716.	mar	9
717.	menino	9
718.	monstro	9
719.	negra	9
720.	nele	9
721.	nessa	9
722.	noites	9
723.	palco	9
724.	papai	9
725.	passado	9
726.	passo	9
727.	pedir	9
728.	pior	9
729.	praia	9
730.	preciso	9
731.	prostituta	9
732.	quinze	9
733.	roupa	9
734.	sabendo	9
735.	sabia	9
736.	sai	9
737.	seguir	9
738.	serviço	9
739.	simples	9
740.	soldados	9
741.	sra	9
742.	talento	9
743.	tendo	9
744.	tivesse	9
745.	tomou	9
746.	trama	9
747.	trata	9
748.	tua	9
749.	ultima	9
750.	v	9
751.	vários	9
752.	vinte	9
753.	viver	9
754.	vocês	9
755.	alasca	8
756.	alegre	8
757.	aliás	8

758.	alta	8
759.	ana	8
760.	art	8
761.	artigo	8
762.	bonito	8
763.	camiliana	8
764.	capaz	8
765.	caricatura	8
766.	carreira	8
767.	casal	8
768.	central	8
769.	chamar	8
770.	chegaram	8
771.	chiquinha	8
772.	chorar	8
773.	clara	8
774.	código	8
775.	colégio	8
776.	comédia	8
777.	comer	8
778.	conselho	8
779.	daqui	8
780.	diferentes	8
781.	dizem	8
782.	dizendo	8
783.	doente	8
784.	encontro	8
785.	encontrou	8
786.	entra	8
787.	entrada	8
788.	escreveu	8
789.	esposa	8
790.	estranho	8
791.	estudante	8
792.	expressão	8
793.	fallar	8
794.	fazia	8
795.	fernando	8
796.	festival	8
797.	fina	8
798.	flagrante	8
799.	graças	8
800.	guarda	8
801.	guerra	8
802.	inicio	8
803.	interior	8

804.	iorque	8
805.	irmãos	8
806.	jornalista	8
807.	lagrimas	8
808.	largo	8
809.	leitor	8
810.	levou	8
811.	lilia	8
812.	literatura	8
813.	madalena	8
814.	major	8
815.	marcilley	8
816.	martin	8
817.	meios	8
818.	minhas	8
819.	minuto	8
820.	moças	8
821.	modo	8
822.	motivo	8
823.	nariz	8
824.	noel	8
825.	páginas	8
826.	passada	8
827.	pediu	8
828.	pergunta	8
829.	pitou	8
830.	pobres	8
831.	pôde	8
832.	policial	8
833.	prazer	8
834.	presente	8
835.	preso	8
836.	principalmente	8
837.	rapariga	8
838.	respeito	8
839.	resposta	8
840.	romances	8
841.	século	8
842.	show	8
843.	silencio	8
844.	som	8
845.	surpresa	8
846.	teria	8
847.	tive	8
848.	umas	8
849.	único	8

850.	uso	8
851.	vadio	8
852.	verde	8
853.	via	8
854.	vice	8
855.	voltar	8
856.	zona	8
857.	acabou	7
858.	ação	7
859.	acha	7
860.	acima	7
861.	afirmou	7
862.	agente	7
863.	aí	7
864.	ambos	7
865.	amiga	7
866.	andré	7
867.	apresenta	7
868.	ás	7
869.	assassino	7
870.	b	7
871.	bahia	7
872.	banco	7
873.	bateria	7
874.	boca	7
875.	bons	7
876.	branca	7
877.	cabelo	7
878.	cadeia	7
879.	caes	7
880.	cães	7
881.	cahir	7
882.	cair	7
883.	cartas	7
884.	casar	7
885.	castelo	7
886.	chama	7
887.	chamada	7
888.	chouriço	7
889.	colega	7
890.	comigo	7
891.	comportamento	7
892.	connor	7
893.	contrario	7
894.	conversar	7
895.	corrida	7

896.	criminal	7
897.	culpa	7
898.	della	7
899.	difícil	7
900.	direção	7
901.	distrito	7
902.	duque	7
903.	êle	7
904.	enorme	7
905.	entrou	7
906.	és	7
907.	escrever	7
908.	escuro	7
909.	especial	7
910.	espetáculo	7
911.	estas	7
912.	existe	7
913.	faço	7
914.	falando	7
915.	fato	7
916.	favor	7
917.	feia	7
918.	felicidade	7
919.	folhetim	7
920.	francisco	7
921.	francos	7
922.	fuga	7
923.	futuro	7
924.	general	7
925.	georges	7
926.	gonçalves	7
927.	guitarrista	7
928.	iate	7
929.	ideal	7
930.	ideia	7
931.	inferno	7
932.	informações	7
933.	inglês	7
934.	irmã	7
935.	jeito	7
936.	joinville	7
937.	juntos	7
938.	ladrões	7
939.	leitores	7
940.	leitura	7
941.	lembrar	7

942.	levada	7
943.	lolita	7
944.	louco	7
945.	lustro	7
946.	mae	7
947.	mamãe	7
948.	mandou	7
949.	marianna	7
950.	marlowe	7
951.	médico	7
952.	megera	7
953.	miguel	7
954.	min	7
955.	mocinha	7
956.	multidão	7
957.	musa	7
958.	namoro	7
959.	nela	7
960.	noiva	7
961.	novela	7
962.	olho	7
963.	p	7
964.	papo	7
965.	partido	7
966.	penal	7
967.	pensão	7
968.	pensava	7
969.	perceber	7
970.	pérola	7
971.	podem	7
972.	portas	7
973.	porto	7
974.	portuguesa	7
975.	presidente	7
976.	príncipe	7
977.	prisão	7
978.	procura	7
979.	prova	7
980.	publica	7
981.	quasi	7
982.	queiroz	7
983.	quiz	7
984.	receber	7
985.	relações	7
986.	resolveu	7
987.	richard	7

988.	roteiro	7
989.	sahir	7
990.	saint	7
991.	segredo	7
992.	sessão	7
993.	sintético	7
994.	somente	7
995.	sorte	7
996.	tanta	7
997.	temos	7
998.	teu	7
999.	tio	7
1000.	tocar	7
1001.	trato	7
1002.	única	7
1003.	vagas	7
1004.	vento	7
1005.	voltou	7
1006.	volume	7
1007.	vontade	7
1008.	á	7
1009.	ã	6
1010.	abelha	6
1011.	abre	6
1012.	acabava	6
1013.	acústico	6
1014.	afim	6
1015.	agua	6
1016.	alda	6
1017.	alem	6
1018.	algo	6
1019.	alves	6
1020.	amores	6
1021.	andou	6
1022.	apaixona	6
1023.	aparece	6
1024.	aqueles	6
1025.	aquilo	6
1026.	assis	6
1027.	assistencia	6
1028.	ator	6
1029.	augusto	6
1030.	bandeirantes	6
1031.	banho	6
1032.	base	6
1033.	básico	6

1034.	bastante	6
1035.	batalha	6
1036.	bate	6
1037.	big	6
1038.	cabellos	6
1039.	café	6
1040.	caixa	6
1041.	cama	6
1042.	camille	6
1043.	cantava	6
1044.	capataz	6
1045.	capital	6
1046.	certamente	6
1047.	civil	6
1048.	collegio	6
1049.	collinette	6
1050.	comadre	6
1051.	começo	6
1052.	commissario	6
1053.	compromisso	6
1054.	comum	6
1055.	conclusão	6
1056.	conversa	6
1057.	corrêa	6
1058.	cozinha	6
1059.	cuja	6
1060.	dados	6
1061.	daniel	6
1062.	daquella	6
1063.	dava	6
1064.	deixe	6
1065.	deles	6
1066.	descobre	6
1067.	desejo	6
1068.	dessas	6
1069.	diabo	6
1070.	duro	6
1071.	durst	6
1072.	entrega	6
1073.	escritores	6
1074.	espécie	6
1075.	espectador	6
1076.	essas	6
1077.	estaçao	6
1078.	estes	6
1079.	estranya	6

1080.	estrela	6
1081.	estudos	6
1082.	falla	6
1083.	falsa	6
1084.	fama	6
1085.	famoso	6
1086.	fanny	6
1087.	fase	6
1088.	feitor	6
1089.	ferreira	6
1090.	ficam	6
1091.	ficantes	6
1092.	ficava	6
1093.	ficção	6
1094.	fiz	6
1095.	fizeram	6
1096.	forças	6
1097.	fortuna	6
1098.	gosta	6
1099.	governo	6
1100.	gritou	6
1101.	guy	6
1102.	haver	6
1103.	hooker	6
1104.	horário	6
1105.	hospital	6
1106.	houve	6
1107.	i	6
1108.	ii	6
1109.	ilha	6
1110.	iniciativa	6
1111.	inteligente	6
1112.	interessante	6
1113.	jennifer	6
1114.	justamente	6
1115.	juventude	6
1116.	labios	6
1117.	laura	6
1118.	letras	6
1119.	lisbela	6
1120.	longa	6
1121.	louras	6
1122.	lucita	6
1123.	luis	6
1124.	má	6
1125.	maioria	6

1126.	marca	6
1127.	marginal	6
1128.	marinheiro	6
1129.	melhores	6
1130.	menores	6
1131.	mente	6
1132.	merece	6
1133.	minerva	6
1134.	mostra	6
1135.	movimento	6
1136.	mr	6
1137.	multo	6
1138.	murmurou	6
1139.	naquella	6
1140.	natural	6
1141.	ninguem	6
1142.	novos	6
1143.	numero	6
1144.	obras	6
1145.	obrigada	6
1146.	of	6
1147.	opinião	6
1148.	orgulhosa	6
1149.	paixão	6
1150.	par	6
1151.	pareciam	6
1152.	passava	6
1153.	passos	6
1154.	paul	6
1155.	pedindo	6
1156.	pegar	6
1157.	pensando	6
1158.	pensar	6
1159.	pequenos	6
1160.	perguntas	6
1161.	perigo	6
1162.	personalidade	6
1163.	peso	6
1164.	pessoal	6
1165.	poesia	6
1166.	população	6
1167.	popular	6
1168.	portanto	6
1169.	poucos	6
1170.	povo	6
1171.	preferência	6

1172.	presos	6
1173.	principal	6
1174.	problemas	6
1175.	prosa	6
1176.	próxima	6
1177.	próximo	6
1178.	pública	6
1179.	punha	6
1180.	qualidade	6
1181.	quartos	6
1182.	real	6
1183.	recebeu	6
1184.	reconhecer	6
1185.	relacionamento	6
1186.	revela	6
1187.	revista	6
1188.	rico	6
1189.	rock	6
1190.	rogério	6
1191.	romântica	6
1192.	romântico	6
1193.	sabem	6
1194.	sala	6
1195.	seguida	6
1196.	sentido	6
1197.	sexy	6
1198.	silvio	6
1199.	sinto	6
1200.	suzane	6
1201.	talavera	6
1202.	terça	6
1203.	terceiro	6
1204.	tirar	6
1205.	titulo	6
1206.	to	6
1207.	torno	6
1208.	totalmente	6
1209.	trabalhar	6
1210.	trecho	6
1211.	u	6
1212.	ultimo	6
1213.	usar	6
1214.	vagabundos	6
1215.	vara	6
1216.	venha	6
1217.	vi	6

1218.	violão	6
1219.	virgem	6
1220.	william	6
1221.	xadrez	6
1222.	fossem	5
1223.	gênero	5
1224.	geral	5
1225.	gostou	5
1226.	grave	5
1227.	grêmio	5
1228.	grito	5
1229.	gritos	5
1230.	gueixa	5
1231.	habilidade	5
1232.	habitantes	5
1233.	hawks	5
1234.	histórias	5
1235.	hontem	5
1236.	hospício	5
1237.	hotel	5
1238.	iam	5
1239.	ídéias	5
1240.	imaginação	5
1241.	importância	5
1242.	importante	5
1243.	inclusive	5
1244.	incrível	5
1245.	infantil	5
1246.	inglesa	5
1247.	irreverente	5
1248.	jantar	5
1249.	jey	5
1250.	joga	5
1251.	jornais	5
1252.	l	5
1253.	late	5
1254.	lembra	5
1255.	letra	5
1256.	levando	5
1257.	levantar	5
1258.	levantou	5
1259.	lições	5
1260.	linha	5
1261.	lista	5
1262.	loja	5
1263.	marcado	5

1264.	marcas	5
1265.	marcello	5
1266.	marcial	5
1267.	marinha	5
1268.	mario	5
1269.	marlboro	5
1270.	martel	5
1271.	mato	5
1272.	maxime	5
1273.	memórias	5
1274.	mendigo	5
1275.	mês	5
1276.	migalhas	5
1277.	miller	5
1278.	mistério	5
1279.	mistura	5
1280.	monroe	5
1281.	monte	5
1282.	moraes	5
1283.	moreno	5
1284.	mudou	5
1285.	municipal	5
1286.	necessidade	5
1287.	negro	5
1288.	netto	5
1289.	neve	5
1290.	ninfeta	5
1291.	nomes	5
1292.	notável	5
1293.	novamente	5
1294.	novas	5
1295.	nove	5
1296.	novidades	5
1297.	oficial	5
1298.	oliveira	5
1299.	ótima	5
1300.	paes	5
1301.	página	5
1302.	país	5
1303.	passando	5
1304.	passaram	5
1305.	paulinho	5
1306.	pegou	5
1307.	pensou	5
1308.	perdeu	5
1309.	perdida	5

1310.	pereira	5
1311.	perfeitamente	5
1312.	perguntar	5
1313.	pesquisa	5
1314.	piano	5
1315.	pinto	5
1316.	pintura	5
1317.	plano	5
1318.	planos	5
1319.	plateia	5
1320.	poderá	5
1321.	poeta	5
1322.	pondô	5
1323.	pontos	5
1324.	possa	5
1325.	prato	5
1326.	preocupação	5
1327.	primeiros	5
1328.	princesa	5
1329.	processos	5
1330.	profissão	5
1331.	programação	5
1332.	propriedade	5
1333.	provas	5
1334.	pudesse	5
1335.	quadro	5
1336.	radio	5
1337.	rainha	5
1338.	rap	5
1339.	rapidamente	5
1340.	raul	5
1341.	refrão	5
1342.	rei	5
1343.	representantes	5
1344.	republicana	5
1345.	resplendor	5
1346.	riqueza	5
1347.	risco	5
1348.	rival	5
1349.	robert	5
1350.	sarah	5
1351.	scorsese	5
1352.	sebastião	5
1353.	seixas	5
1354.	selvagem	5
1355.	semelhante	5

1356.	senador	5
1357.	sentidos	5
1358.	sentir	5
1359.	sentiu	5
1360.	sério	5
1361.	soltou	5
1362.	sombra	5
1363.	sorriso	5
1364.	sotaque	5
1365.	souza	5
1366.	tânia	5
1367.	temas	5
1368.	tens	5
1369.	tenta	5
1370.	termos	5
1371.	thesouro	5
1372.	tiro	5
1373.	total	5
1374.	tratamento	5
1375.	tratava	5
1376.	tribunal	5
1377.	trinta	5
1378.	tristeza	5

1379.	turma	5
1380.	turner	5
1381.	últimos	5
1382.	vejam	5
1383.	venda	5
1384.	vendo	5
1385.	versão	5
1386.	vestia	5
1387.	vestida	5
1388.	vídeo	5
1389.	vigilância	5
1390.	villemer	5
1391.	violência	5
1392.	vira	5
1393.	visita	5
1394.	viuva	5
1395.	vivia	5
1396.	vocalista	5
1397.	volumes	5
1398.	votos	5
1399.	york	5
1400.	you	5

Anexo II: Lista de lexemas a partir das 1400 lexias mais frequentes no *corpus*

Ordem	Lexema	Frequência
1.	TER	414
2.	ESTAR	228
3.	VADIA	184
4.	DIZER	165
5.	FAZER	151
6.	MULHER	135
7.	FICAR	116
8.	PODER	109
9.	DIA	101
10.	CASA	97
11.	SABER	85
12.	HAVER	84
13.	VIDA	81
14.	DAR	80
15.	HOMEM	75
16.	IR	75
17.	QUERER	74
18.	PASSAR	72
19.	CHEGAR	67
20.	FILME	67
21.	TEMPO	65
22.	DAR	62
23.	MÃE	62
24.	PAI	61
25.	NOITE	60
26.	COISA	59
27.	MUNDO	57
28.	OLHOS	56
29.	VIR	56
30.	HISTÓRIA	54
31.	RUA	52
32.	MENINA	51
33.	SER	50
34.	FALAR	48
35.	AMIGO	47
36.	RIO	47
37.	SAIR	46
38.	BOM	44
39.	DEIXAR	42
40.	LIVRO	42
41.	PAI	40
42.	AMOR	39
43.	CRIANÇA	38
44.	FILHO	38
45.	VER	38
46.	DEUS	37
47.	MÃO	37
48.	PERSONAGEM	37

49.	PENSAR	35
50.	POBRE	35
51.	POLICIA	35
52.	MAL	34
53.	PARECER	33
54.	ARTISTA	32
55.	CASO	32
56.	FAZER	32
57.	FILHA	32
58.	OBRAS	31
59.	PORTA	31
60.	ROMANCE	31
61.	VOLTAR	31
62.	ACHAR	30
63.	BRASIL	30
64.	CIDADE	30
65.	JOVEM	30
66.	PEQUENA	30
67.	TV	30
68.	MARIDO	29
69.	NOVO	29
70.	ACABAR	28
71.	DIRETOR	28
72.	GOSTAR	28
73.	LEVANTAR	28
74.	CONTA	27
75.	ENCONTRAR	27
76.	ENTRAR	27
77.	LOBA	27
78.	VIVER	27
79.	BOA	26
80.	CERTO	26
81.	CORPO	26
82.	MORRER	26
83.	SENTIR	26
84.	ANDAR	25
85.	NOVA	25
86.	SENHOR	25
87.	CHAMAR	24
88.	COSTA	24
89.	DINHEIRO	24
90.	TELEFONE	24
91.	TOMAR	24
92.	TRABALHO	24
93.	CARA	23
94.	ESTADO	23
95.	PEDIR	23
96.	VERDADE	23
97.	AMERICANA	22

98.	BRANCO	22
99.	FALTAR	22
100.	JORNAL	22
101.	MÚSICA	22
102.	QUARTO	22
103.	VELHA	22
104.	VOZ	22
105.	BRASILEIRA	21
106.	CASAMENTO	21
107.	CINEMA	21
108.	CORAÇÃO	21
109.	DRAMA	21
110.	GRUPO	21
111.	MOÇA	21
112.	OURO	21
113.	PERGUNTAR	21
114.	PROBLEMA	21
115.	SOL	21
116.	VELHO	21
117.	CLARO	20
118.	DIFERENTE	20
119.	PRECISAR	20
120.	PROCESSOS	20
121.	SOCIEDADE	20
122.	DONA	19
123.	FORÇA	19
124.	GLOBO	19
125.	MENINO	19
126.	MORTE	19
127.	PEQUENO	19
128.	RÁDIO	19
129.	CABEÇA	18
130.	ESCRITOR	18
131.	FESTA	18
132.	GRAÇA	18
133.	JUIZ	18
134.	CAMINHO	17
135.	CONTAR	17
136.	COPACABANA	17
137.	FEIRA	17
138.	FINO	17
139.	FORMA	17
140.	IRMÃO	17
141.	JORNAL	17
142.	MENOR	17
143.	ABRIR	16
144.	ESCOLA	16
145.	FILMES	16
146.	FORMIGA	16
147.	FRIO	16
148.	OLHAR	16
149.	ROSTO	16

150.	SUCESSO	16
151.	CANTAR	15
152.	CRIME	15
153.	ESCREVER	15
154.	ESTRELA	15
155.	FINAL	15
156.	FORTE	15
157.	LEITOR	15
158.	LINDA	15
159.	LOURA	15
160.	MEIA	15
161.	PALAVRA	15
162.	PAPEL	15
163.	PRODUÇÃO	15
164.	PROGRAMA	15
165.	AVENIDA	14
166.	CAIR	14
167.	CÃO	14
168.	CLUBE	14
169.	CONDESSA	14
170.	CONSEGUIR	14
171.	COR	14
172.	COUSA	14
173.	EXEMPLO	14
174.	FAMILIA	14
175.	FAMÍLIA	14
176.	JOGO	14
177.	JUSTIÇA	14
178.	LEVAR	14
179.	LIBERDADE	14
180.	LÍNGUA	14
181.	MESTRE	14
182.	NACIONAL	14
183.	PARIS	14
184.	PEÇA	14
185.	PERGUNTA	14
186.	PRESO	14
187.	SEXO	14
188.	TERRA	14
189.	TRATAR	14
190.	USAR	14
191.	ALVA	13
192.	AR	13
193.	BELEZA	13
194.	CENA	13
195.	ÉPOCA	13
196.	EXCLAMAR	13
197.	FELIZ	13
198.	GAROTA	13
199.	MEDO	13
200.	MORTA	13
201.	PÁGINA	13

202.	PEDRA	13
203.	PENA	13
204.	RECEBER	13
205.	RELACIONAMENTO	13
206.	RICA	13
207.	ROSA	13
208.	ALGODÃO	12
209.	AMERICANO	12
210.	BRAÇO	12
211.	BRASILEIRO	12
212.	CAMPANHA	12
213.	CIGARRA	12
214.	COMPANHIA	12
215.	DÉCADA	12
216.	DELEGADO	12
217.	DESTINO	12
218.	DOMINGO	12
219.	FUNDO	12
220.	IDEIA	12
221.	JANEIRO	12
222.	LEMBRAR	12
223.	PEITO	12
224.	PERNA	12
225.	PRESA	12
226.	PRETO	12
227.	PROFESSOR	12
228.	PROVA	12
229.	QUESTÃO	12
230.	RAPAZ	12
231.	RELAÇÃO	12
232.	SANTA	12
233.	SENHORA	12
234.	SEXUAL	12
235.	SUL	12
236.	TEATRO	12
237.	UNIDO	12
238.	VESTIDO	12
239.	VOLUME	12
240.	ALMA	11
241.	ANÚNCIO	11
242.	ATRIZ	11
243.	AUTOR	11
244.	AUTORA	11
245.	BONITA	11
246.	CÉU	11
247.	COMEÇAR	11
248.	DELEGACIA	11
249.	DISCO	11
250.	DORMIR	11
251.	ESPAÇO	11
252.	ESTILO	11
253.	FÁCIL	11

254.	FEDERAL	11
255.	HUMANA	11
256.	LETRA	11
257.	MACHADO	11
258.	MARCA	11
259.	MORAR	11
260.	MORTO	11
261.	OUVIR	11
262.	PASSAGEM	11
263.	PÉ	11
264.	PEGAR	11
265.	PORTUGUÊS	11
266.	PRAÇA	11
267.	PÚBLICO	11
268.	QUERIDA	11
269.	RESULTADO	11
270.	ROCHA	11
271.	SANGUE	11
272.	SANTO	11
273.	SEGUNDA	11
274.	SENTIDO	11
275.	SÉRIE	11
276.	TELEVISÃO	11
277.	TEXTO	11
278.	ULTIMO	11
279.	VIVO	11
280.	APARTAMENTO	10
281.	BANDA	10
282.	CAMISINHA	10
283.	CANTORA	10
284.	CARNAVAL	10
285.	CHÃO	10
286.	CONTINUAR	10
287.	CORRER	10
288.	EMPREGADA	10
289.	FUGIR	10
290.	GRITO	10
291.	LAMINA	10
292.	LEI	10
293.	LUZ	10
294.	MESA	10
295.	NAMORADA	10
296.	NAMORADO	10
297.	ORDEM	10
298.	PLANO	10
299.	PRETA	10
300.	RAZÃO	10
301.	RESPONDER	10
302.	RURAL	10
303.	SOCIAL	10
304.	TRISTE	10
305.	VADIAGEM	10

306.	ALEGRIA	9
307.	ALTO	9
308.	ASSUNTO	9
309.	AULA	9
310.	BAIRRO	9
311.	BELA	9
312.	CADERNETA	9
313.	CARIOWA	9
314.	CARRO	9
315.	CERTEZA	9
316.	CLIMA	9
317.	COMIDA	9
318.	CONHECIDA	9
319.	DANÇAR	9
320.	DIREITO	9
321.	DOLLAR	9
322.	ESCOLHA	9
323.	ESPERAR	9
324.	FINALMENTE	9
325.	FLOR	9
326.	FOME	9
327.	HUMOR	9
328.	IGREJA	9
329.	JARDIM	9
330.	LEVE	9
331.	MADRUGADA	9
332.	MAR	9
333.	MONSTRO	9
334.	NEGRA	9
335.	NOITES	9
336.	PALCO	9
337.	PAPAI	9
338.	PASSADO	9
339.	PIOR	9
340.	PRAIA	9
341.	PROSTITUTA	9
342.	QUINZE	9
343.	ROUPA	9
344.	SEGUIR	9
345.	SERVIÇO	9
346.	SIMPLES	9
347.	SOLDADO	9
348.	TALENTO	9
349.	TRAMA	9
350.	ALASCA	8
351.	ALEGRE	8
352.	ALTA	8
353.	ARTIGO	8
354.	BONITO	8
355.	CAPAZ	8
356.	CARICATURA	8
357.	CARREIRA	8

358.	CASAL	8
359.	CENTRAL	8
360.	CLARA	8
361.	CÓDIGO	8
362.	COLÉGIO	8
363.	COMÉDIA	8
364.	COMER	8
365.	CONSELHO	8
366.	DOENTE	8
367.	ENTRADA	8
368.	ESPOSA	8
369.	ESTRANHO	8
370.	ESTUDANTE	8
371.	EXPRESSÃO	8
372.	FESTIVAL	8
373.	FINA	8
374.	FLAGRANTE	8
375.	GRAÇAS	8
376.	GUARDA	8
377.	GUERRA	8
378.	INICIO	8
379.	INTERIOR	8
380.	IORQUE	8
381.	JORNALISTA	8
382.	LÁGRIMA	8
383.	LARGO	8
384.	LITERATURA	8
385.	MAJOR	8
386.	MEIO	8
387.	MINUTO	8
388.	MOTIVO	8
389.	NARIZ	8
390.	NOEL	8
391.	PASSADA	8
392.	PITAR	8
393.	POLICIAL	8
394.	PRAZER	8
395.	PRESENTE	8
396.	PRINCIPALMENTE	8
397.	RAPARIGA	8
398.	RESPEITO	8
399.	RESPOSTA	8
400.	SÉCULO	8
401.	SHOW	8
402.	SILENCIO	8
403.	SOM	8
404.	SURPRESA	8
405.	ÚNICO	8
406.	VADIO	8
407.	VERDE	8
408.	VICE	8
409.	ZONA	8

410.	AÇÃO	7
411.	AFIRMAR	7
412.	AGENTE	7
413.	AMIGA	7
414.	APRESENTAR	7
415.	ASSASSINO	7
416.	BAHIA	7
417.	BANCO	7
418.	BATER	7
419.	BATERIA	7
420.	BOCA	7
421.	BRANCA	7
422.	CABELO	7
423.	CADEIA	7
424.	CARTA	7
425.	CASAR	7
426.	CASTELO	7
427.	CHAMADA	7
428.	CHOURIÇO	7
429.	COLEGA	7
430.	COMPORTAMENTO	7
431.	CONTRARIO	7
432.	CONVERSAR	7
433.	CORRIDA	7
434.	CRIMINAL	7
435.	CULPA	7
436.	DIFÍCIL	7
437.	DIREÇÃO	7
438.	DISTRICTO	7
439.	DUQUE	7
440.	ENORME	7
441.	ESCURO	7
442.	ESPECIAL	7
443.	ESPETÁCULO	7
444.	EXISTIR	7
445.	FALSO	7
446.	FATO	7
447.	FAVOR	7
448.	FAZER	7
449.	FEIA	7
450.	FELICIDADE	7
451.	FOLHETIM	7
452.	FRANCO	7
453.	FUGA	7
454.	FUTURO	7
455.	GENERAL	7
456.	GUITARRISTA	7
457.	IATE	7
458.	IDEAL	7
459.	INFERNO	7
460.	INFORMAÇÃO	7
461.	INGLÊS	7

462.	IRMÃ	7
463.	JEITO	7
464.	JOINVILLE	7
465.	LADRÃO	7
466.	LEITURA	7
467.	LEVADA	7
468.	LOUCO	7
469.	LUSTRO	7
470.	MAMÃE	7
471.	MANDAR	7
472.	MÉDICO	7
473.	MEGERA	7
474.	MOCINHA	7
475.	MULTIDÃO	7
476.	MUSA	7
477.	NAMORO	7
478.	NOIVA	7
479.	NOVELA	7
480.	PAPO	7
481.	PARTIDO	7
482.	PENAL	7
483.	PENSÃO	7
484.	PERCEBER	7
485.	PÉROLA	7
486.	PORTO	7
487.	PORTUGUESA	7
488.	PRESIDENTE	7
489.	PRÍNCIPE	7
490.	PRISÃO	7
491.	PROCURA	7
492.	PUBLICA	7
493.	QUIZ	7
494.	RESOLVER	7
495.	ROTEIRO	7
496.	SEGREDO	7
497.	SESSÃO	7
498.	SINTÉTICO	7
499.	SORTE	7
500.	TIO	7
501.	TOCAR	7
502.	TRATO	7
503.	VAGA	7
504.	VENTO	7
505.	VONTADE	7
506.	ABELHA	6
507.	ACÚSTICO	6
508.	ÁGUA	6
509.	APAIXONAR	6
510.	APARECE	6
511.	ASSISTENCIA	6
512.	ATOR	6
513.	BANDEIRANTE	6

514.	BANHO	6
515.	BASE	6
516.	BÁSICO	6
517.	BATALHA	6
518.	CABELLO	6
519.	CAFÉ	6
520.	CAIXA	6
521.	CAMA	6
522.	CAPATAZ	6
523.	CAPITAL	6
524.	CIVIL	6
525.	COLLEGIO	6
526.	COMADRE	6
527.	COMEÇO	6
528.	COMMISSARIO	6
529.	COMPROMISSO	6
530.	COMUM	6
531.	CONCLUSÃO	6
532.	CONVERSA	6
533.	COZINHA	6
534.	DADO	6
535.	DESCOBRIR	6
536.	DESEJO	6
537.	DIABO	6
538.	DURO	6
539.	ENTREGA	6
540.	ESPÉCIE	6
541.	ESPECTADOR	6
542.	ESTAÇÃO	6
543.	ESTRANHA	6
544.	ESTUDO	6
545.	FALSA	6
546.	FAMA	6
547.	FAMOSO	6
548.	FASE	6
549.	FEITOR	6
550.	FICANTE	6
551.	FICÇÃO	6
552.	FORTUNA	6
553.	GOVERNO	6
554.	GRITAR	6
555.	HORÁRIO	6
556.	HOSPITAL	6
557.	ILHA	6
558.	INICIATIVA	6
559.	INTELIGENTE	6
560.	INTERESSANTE	6
561.	JUVENTUDE	6
562.	LÁBIO	6
563.	LONGA	6
564.	MÁ	6
565.	MARGINAL	6

566.	MARINHEIRO	6
567.	MELHOR	6
568.	MENTE	6
569.	MERECER	6
570.	MINERVA	6
571.	MOSTRAR	6
572.	MOVIMENTO	6
573.	MULTO	6
574.	MURMURAR	6
575.	NATURAL	6
576.	NUMERO	6
577.	OPINIÃO	6
578.	ORGULHOSA	6
579.	PAIXÃO	6
580.	PAR	6
581.	PASSO	6
582.	PERIGO	6
583.	PERSONALIDADE	6
584.	PESO	6
585.	PESSOAL	6
586.	POESIA	6
587.	POPULAÇÃO	6
588.	POPULAR	6
589.	POR	6
590.	POVO	6
591.	PREFERÊNCIA	6
592.	PRINCIPAL	6
593.	PROSA	6
594.	PRÓXIMO	6
595.	PÚBLICA	6
596.	QUALIDADE	6
597.	REAL	6
598.	RECONHECER	6
599.	REVELAR	6
600.	REVISTA	6
601.	RICO	6
602.	ROCK	6
603.	ROMÂNTICA	6
604.	ROMÂNTICO	6
605.	SALA	6
606.	SEGUIDA	6
607.	SEXY	6
608.	TALAVERA	6
609.	TERÇA	6
610.	TIRAR	6
611.	TÍTULO	6
612.	TOTALMENTE	6
613.	TRABALHAR	6
614.	TRECHO	6
615.	VAGABUNDO	6
616.	VARA	6
617.	VIOLÃO	6

618.	VIRGEM	6
619.	XADREZ	6
620.	GÊNERO	5
621.	GERAL	5
622.	GRAVE	5
623.	GRÊMIO	5
624.	GUEIXA	5
625.	HABILIDADE	5
626.	HABITANTE	5
627.	HONTEM	5
628.	HOSPÍCIO	5
629.	HOTEL	5
630.	IATE	5
631.	IMAGINAÇÃO	5
632.	IMPORTÂNCIA	5
633.	IMPORTANTE	5
634.	INCRÍVEL	5
635.	INFANTIL	5
636.	INGLESA	5
637.	IRREVERENTE	5
638.	JANTAR	5
639.	JOGAR	5
640.	LIÇAO	5
641.	LINHA	5
642.	LISTA	5
643.	LOJA	5
644.	MARCADO	5
645.	MARCIAL	5
646.	MARINHA	5
647.	MATO	5
648.	MEMÓRIA	5
649.	MENDIGO	5
650.	MÊS	5
651.	MIGALHA	5
652.	MISTÉRIO	5
653.	MISTURA	5
654.	MONTE	5
655.	MORENO	5
656.	MUDOU	5
657.	MUNICIPAL	5
658.	NECESSIDADE	5
659.	NEGRO	5
660.	NEVE	5
661.	NIINFETA	5
662.	NOTÁVEL	5
663.	NOVAMENTE	5
664.	NOVAS	5
665.	NOVIDADE	5
666.	OFICIAL	5
667.	ÓTIMA	5
668.	PAÍS	5
669.	PERDER	5

670.	PERDIDA	5
671.	PERFEITAMENTE	5
672.	PESQUISA	5
673.	PIANO	5
674.	PINTO	5
675.	PINTURA	5
676.	PLATEIA	5
677.	POETA	5
678.	PONTO	5
679.	PRATO	5
680.	PREOCUPAÇÃO	5
681.	PRINCESA	5
682.	PROFISSÃO	5
683.	PROGRAMAÇÃO	5
684.	PROPRIEDADE	5
685.	QUADRO	5
686.	RADIO	5
687.	RAINHA	5
688.	RAP	5
689.	RAPIDAMENTE	5
690.	REFRÃO	5
691.	REI	5
692.	REPRESENTANTE	5
693.	REPUBLICANA	5
694.	RESPLendor	5
695.	RIQUEZA	5
696.	RISCO	5
697.	RIVAL	5
698.	SEIXA	5
699.	SELVAGEM	5
700.	SEMELHANTE	5
701.	SENADOR	5
702.	SÉRIO	5
703.	SOLTAR	5
704.	SOMBRA	5
705.	SORRISO	5
706.	SOTAQUE	5
707.	TEMA	5
708.	THESOURO	5
709.	TIRO	5
710.	TOTAL	5
711.	TRATAMENTO	5
712.	TRIBUNAL	5
713.	TRISTEZA	5
714.	TURMA	5
715.	VENDA	5
716.	VERSÃO	5
717.	VESTIDA	5
718.	VESTIR	5
719.	VÍDEO	5
720.	VIGILÂNCIA	5
721.	VIOLÊNCIA	5

722.	VISITAR	5
723.	VIUVA	5

724.	VOCALISTA	5
725.	VOTO	5