



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

GUILHERME APARECIDO DE SOUZA

***GAY LANGUAGE: ANÁLISE COLOCACIONAL E A PROPOSTA DE
UM GLOSSÁRIO BILÍNGUE DE COLOCAÇÕES BASEADO EM
CORPUS***

SÃO JOSÉ DO RIO PRETO
2018



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

GUILHERME APARECIDO DE SOUZA

***GAY LANGUAGE: ANÁLISE COLOCACIONAL E A PROPOSTA DE
UM GLOSSÁRIO BILÍNGUE DE COLOCAÇÕES BASEADO EM
CORPUS***

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de São José do Rio Preto. (Área de concentração: Linguística Aplicada)

Orientadora: Prof^ª. Dr^ª. Adriane Orenha-Ottaiano

SÃO JOSÉ DO RIO PRETO
2018

Souza, Guilherme Aparecido de.

Gay language: análise colocacional e a proposta de um glossário bilíngue de colocações baseado em corpus / Guilherme Aparecido de Souza. -- São José do Rio Preto, 2018.

129 f. : il.

Orientador: Adriane Orenha-Ottaiano

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Linguística. 2. Linguística de corpus. 3. Fraseologia. 4. Homossexuais - Linguagem. I. Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas. II. Título.

CDU - 41

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Campus de São José do Rio Preto

GUILHERME APARECIDO DE SOUZA

***GAY LANGUAGE: ANÁLISE COLOCACIONAL E A PROPOSTA DE
UM GLOSSÁRIO BILÍNGUE DE COLOCAÇÕES BASEADO EM
CORPUS***

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto. (Área de concentração: Linguística Aplicada)

COMISSÃO EXAMINADORA

Titulares

Prof^a. Dr^a. Adriane Orenha-Ottaiano – Orientadora
IBILCE/UNESP – São José do Rio Preto

Prof^a. Dr^a. Ieda Maria Alves
Universidade de São Paulo – São Paulo

Prof^a. Dr^a. Maria Helena de Paula
Universidade Federal de Goiás – Regional Catalão

Suplentes

Prof^a Dr^a Marilei Amadeu Sabino
IBILCE/UNESP – São José do Rio Preto

Prof. Dr. Odair Luiz Nadin da Silva
FCLAr/UNESP – Araraquara

SÃO JOSÉ DO RIO PRETO
31 de janeiro de 2018

A todos que amo, pelo apoio,
amor e compreensão.

AGRADECIMENTOS

Aos Deuses Africanos.

A todos meus familiares, em especial minha mãe, Aparecida, cujo estímulo e paciência foram indispensáveis para o alcance e êxito deste projeto.

À Profa. Dra. Adriane Orenha-Ottaiano, por todas as qualidades que fizeram dela uma querida professora, orientadora, amiga e exemplo constante de dedicação e competência.

Às amigas Claudia Garcia, Elaine Oliveira, Gláucia Gomes e Graciele Cucolo, pela amizade e apoio constantes.

Ao colega Jean Michel pela ajuda dispensada no desenvolvimento desta pesquisa.

Aos professores Dr. Odair Luiz Nadin da Silva e Dra. Maria Helena de Paula pelas ricas contribuições e pelos elogios durante o VIII e IX SELin, respectivamente.

Às professoras Dra. Maria Cristina Parreira da Silva e Dra. Maria Emília Pereira Chanut por terem aceito o convite para participar da banca de qualificação, trazendo inestimáveis contribuições.

A todos os integrantes da Banca Examinadora, que, gentilmente, aceitaram o convite para avaliar esta dissertação.

RESUMO

Esta pesquisa tem como apoio o arcabouço teórico-metodológico da Linguística de *Corpus* (SINCLAIR, 1991, 1996, 2003; TOGNINI-BONELLI, 1996, 2001, 2004, 2010; HUNSTON, 2002; MEYER, 2004; MCENERY, HARDIE, 2012) e da Fraseologia (HAUSMANN, 1984, 1985, 1990; COWIE, 1998, 1999; MOON, 1998; WRAY, 2002; e ORENHA-OTTAIANO, 2004, 2009), área da Linguística voltada para combinações lexicais recorrentes, uma vez que focamos a investigação das unidades fraseológicas, mais especificamente das colocações frequentemente empregadas pela comunidade homossexual. Tais colocações foram extraídas de um *corpus* paralelo formado pelas transcrições dos episódios das cinco temporadas do seriado *Queer as Folk* e de um *corpus* comparável, composto de um *subcorpus* em inglês e um *subcorpus* português, a fim de atestar as palavras e as colocações mais frequentemente empregadas pela comunidade homossexual. Para realizar o levantamento dessas colocações, utilizamos o programa *WordSmith Tools* (SCOTT, 2012), versão 4.0. A partir do levantamento e da análise das colocações mais frequentemente empregadas pela comunidade homossexual em inglês e suas respectivas colocações em português, elaboramos uma proposta de um glossário de colocações de *gay language* baseado em *corpus*. Para sua elaboração, adotamos a metodologia proposta por Orenha-Ottaiano (2004, 2016), que trata da compilação de obras fraseográficas, mais especificamente de colocações, baseadas em *corpus*. A fim de verificar a frequência de todas as colocações de nosso estudo, utilizamos a ferramenta *Sketch Engine* (KILGARRIFF et al, 2004). A compilação de um glossário de colocações da comunidade homossexual justifica-se pelo fato de não haver uma obra que aborde tais combinatórias na referida área. Além disso, tradutores aprendizes e profissionais terão uma obra para consulta no processo tradutório de tais colocações.

Palavras-chave: Linguística de *Corpus*. Fraseologia. Colocações. *Gay Language*.

ABSTRACT

*This research is supported by the theoretical-methodological framework of Corpus Linguistics (SINCLAIR, 1991, 1996, 2003; TOGNINI-BONELLI, 1996, 2001, 2004, 2010; HUNSTON, 2002; MEYER, 2004; MCENERY, HARDIE, 2012), as well as in the literature on phraseological, (HAUSMANN, 1984, 1985, 1990; COWIE, 1998, 1999; MOON, 1998; WRAY, 2002; e ORENHA-OTTALIANO, 2004, 2009), area of Linguistics focused on recurrent lexical combinations, the investigation of the phraseological units, more specifically of the collocations often employed by the homosexual community. Such collocations were extracted from a parallel corpus formed by the transcriptions of the episodes of the five seasons of the series *Queer as Folk* and a comparable corpus, composed of an English subcorpus and a Portuguese subcorpus to check the most frequent words used by the homosexual community. To collect the data, we used the program *WordSmith Tools* (SCOTT, 2012), version 4.0 Based on the data collected and analysis of the most frequently used collocations by the homosexual community in English and their correspondent Portuguese collocations, we elaborated a proposal for a corpus-based glossary of gay language collocations. For its compilation, we adopted the methodology proposed by Orenha-Ottaiano (2004, 2016), that deals with the compilation of corpus-based phraseological works, more specifically of collocations. In order to verify the frequency of all the collocations of our study, we used the *Sketch Engine* tool (KILGARRIFF et al, 2004). The compilation of a glossary of collocations of the homosexual community is justified by the fact that there is no work that focuses on such combinations in that area. In addition, learner and professional translators will have a work for consultation in the translation process of such collocations.*

Keywords: Corpus linguistics. Phraseology. Collocations. Gay Language.

LISTA DE FIGURAS

Figura 1 – Classificação de Caseres	34
Figura 2 – Classificação de Zuluaga	35
Figura 3 – Classificação de Corpas Pastor	36
Figura 4 – Classificação de Ruiz Gurillo	37
Figura 5 – <i>Lexico-grammar cline</i>	41
Figura 6 – <i>System of cohesion in English</i>	43
Figura 7 – Convencionalidade e Fraseologia	44
Figura 8 – Nosso objeto de estudo	45
Figura 9 – Combinações de palavras de Cowie	47
Figura 10 – <i>Hausmann’s classification of word-combinations</i>	49
Figura 11 – <i>Cowie’s distribution of word-combinations</i>	51
Figura 12 – Interface do <i>WordSmith Tools</i>	64
Figura 13 – Lista de palavras elaborada pela <i>Wordlist</i> , em ordem de frequência	65
Figura 14 – Lista de palavras elaborada pela <i>Wordlist</i> , em ordem alfabética	66
Figura 15 – Tela do <i>Wordlist</i> com as estatísticas do <i>corpus</i>	67
Figura 16 – Aba de concordâncias da ferramenta <i>Concord</i>	68
Figura 17 – Aba de concordâncias da ferramenta <i>Concord</i> na aba <i>collocates</i>	69
Figura 18 – Aba de concordâncias da ferramenta <i>Concord</i> na aba <i>cluster</i>	70
Figura 19 – Aba da ferramenta <i>Keywords</i>	71
Figura 20 – Aba de compilação do <i>corpus</i>	73
Figura 21 – A ferramenta realizando <i>download</i> dos dados para a compilação do <i>corpus</i> .	74
Figura 22 – Aba com as páginas selecionadas na <i>web</i> para compilação do <i>corpus</i>	75
Figura 23 – Frequência de <i>gay rights</i> no <i>subcorpus</i> comparável em inglês compilado na ferramenta <i>Sketch Engine</i>	80
Figura 24 – Frequência de “direito dos gays” no <i>subcorpus</i> comparável em português	80
Figura 25 – Frequência de <i>gay marriage</i> do <i>subcorpus</i> comparável em inglês	82
Figura 26 – Frequência de “casamento gay” do <i>subcorpus</i> comparável em português	82
Figura 27 – Frequência de <i>fag hag</i> do <i>subcorpus</i> comparável em inglês	86
Figura 28 – Frequência de “líder gay” do <i>subcorpus</i> comparável em português	87
Figura 29 - Frequência de <i>little faggot</i> do <i>subcorpus</i> comparável em inglês	90

Figura 30 - Frequência de “bichinha” do <i>subcorpus</i> comparável em português	91
Figura 31 - Frequência de <i>butt fucking</i> do <i>subcorpus</i> comparável em inglês	94
Figura 32 - Frequência de “sexo anal” do <i>subcorpus</i> comparável em português	94
Figura 33 - Frequência de <i>tight ass</i> no <i>subcorpus</i> comparável em inglês	97
Figura 34 – Público-alvo do Glossário	100
Figura 35 – Guia visual do Glossário	103

LISTA DE ORGANOGRAMAS

Organograma 1 – Colocações formadas pela base <i>gay</i>	83
Organograma 2 – Colocações formadas pela base <i>fag</i>	88
Organograma 3 – Colocações formadas pela base <i>faggot</i>	92
Organograma 4 – Colocações formadas pela base <i>butt</i>	95
Organograma 5 – Colocações formadas pela base <i>ass</i>	98

LISTA DE QUADROS

Quadro 1 – Trajetória da Linguagem <i>Gay</i>	19
Quadro 2 - <i>Tactics of Intersubjectivity</i>	23
Quadro 3 - Quadro de autores.....	38
Quadro 4 – Tipologia do <i>corpus</i> de estudo e características de nosso <i>corpus</i> de estudo ...	61
Quadro 5 – Colocação substantiva a partir da base <i>gay</i>	77
Quadro 6 – Colocações adjetivas a partir do colocado <i>gay</i>	78
Quadro 7 – Colocações substantivas a partir do colocado <i>fag</i>	84
Quadro 8 – Colocações adjetivas a partir do colocado <i>fag</i>	84
Quadro 9 – Colocações adjetivas a partir de <i>fag</i>	85
Quadro 10 – Colocações adjetivas a partir da base <i>faggot</i>	89
Quadro 11 – Colocações substantivas a partir do colocado <i>butt</i>	92
Quadro 12 – Colocações adjetivas a partir da base <i>ass</i>	96

LISTA DE TABELAS

Tabela 1 – Estatística da colocação <i>gay rights</i>	79
Tabela 2 – Estatística da colocação <i>gay marriage</i>	81
Tabela 3 – Estatística da colocação <i>fag hag</i>	85
Tabela 4 – Estatística da colocação <i>little faggot</i>	89
Tabela 5 – Estatística da colocação <i>butt fucking</i>	93
Tabela 6 – Estatística da colocação <i>tight ass</i>	96

SUMÁRIO

INTRODUÇÃO	13
1. GAY TWIST: UM TOQUE GAY PARA COMEÇAR	16
1.1 Linguagem e Sexualidade.....	16
1.2 A maneira como os <i>gays</i> falam – Existe <i>Gay Language</i> ?	25
1.3 Contextualizando a Série <i>Queer as Folk</i>	28
2. FUNDAMENTAÇÃO TEÓRICA	32
2.1 Trajetória dos Estudos Fraseológicos	32
2.2 As colocações da língua geral	41
2.3 Linguística de <i>Corpus</i> : definição e tipologia de <i>corpus</i>	52
3. METODOLOGIA DE PESQUISA	63
4. EXTRAÇÃO E ANÁLISE DAS COLOCAÇÕES DA COMUNIDADE HOMOSSEXUAL	76
5. GLOSSÁRIO DE COLOCAÇÕES DA COMUNIDADE HOMOSSEXUAL	99
5.1 Tipologia do Glossário	99
5.1.1 Macro e Microestrutura	100
5.2 Guia Visual	102
5.3 Amostra do Glossário	103
CONSIDERAÇÕES FINAIS	116
REFERÊNCIAS BIBLIOGRÁFICAS	119

INTRODUÇÃO

A realização de um estudo das unidades fraseológicas (doravante UFs), mais especificamente de colocações presentes no seriado *Queer as Folk* e comumente empregadas pela comunidade homossexual, contribui para as áreas da Fraseologia, da Fraseografia e da Linguística de *Corpus*, uma vez que ainda há poucas pesquisas que envolvem o léxico fraseológico da comunidade homossexual.

Barrett (1997, p. 202) lembra que, na análise da língua, pouca atenção é dada ao estudo do léxico específico de uma determinada comunidade quando na construção da competência lexical de um determinado falante. Livia e Hall (1997, p. 7) evidenciam o lugar da linguagem na vida dos seres humanos e sua integração dentro de um grupo social menos demográfico, tradicionalmente tratado por meio de estudos da variação. Com base nesse estudo, as autoras mostram que a integração/aquisição do léxico empregado por todos os membros do grupo Lésbicas, Gays, Bissexuais e Travestis (LGBT) faz com que o sentimento de “pertencer a” surja, como consequência da possibilidade de se identificar os pares por meio da linguagem, isto é, aqueles que apresentam características comuns.

A primeira coleção de artigos sobre questões da linguagem *gay* e *lésbica*, *Gayspeak: Gay Male and Lesbian Communication*, de Chesebro, foi publicada em 1981, seguida de *Queer words, Queer images: communication and construction of homosexuality*, de Leap, em 1994. Livia e Hall (1997), por exemplo, mostram que um dos primeiros glossários *gays*, *The language homosexuality: An American Glossary*, de Legman, lançado em 1941, é composto de 329 vocábulos, ou seja, contém apenas itens lexicais e não combinatórias de palavras ou fraseologismos.

Podemos também citar a obra *A lexicon of homosexual slang*, de Cory e LeRoy, publicada em 1963; a obra de Strait e Associates, *The lavender lexicon: Dictionary of gay words and phrases*, em 1964; *The queens' vernacular*, de Rodgers, em 1972, e *The Argot of homosexual subculture*, de Farrel, também em 1972. Conforme podemos notar, as obras acima não tratam especificamente de colocações, mais uma razão para enfatizarmos a relevância deste trabalho.

Desse modo, este estudo acerca da extração e análise comparativa de colocações da comunidade homossexual, bem como a compilação de um glossário que contenha tais fraseologismos, contribui para melhorar a compreensão da linguagem desse grupo, visto de

forma tão estigmatizada. Ademais, possibilita o acesso a uma obra fraseográfica rica em colocações nos dois idiomas enfocados, inglês e português do Brasil.

Ainda em relação à contribuição deste estudo, há a disseminação de tais fraseologismos em língua inglesa, por meio de uma amostra de um glossário de colocações da comunidade homossexual, levantadas e analisadas a partir do *corpus* paralelo, composto pelo seriado *Queer as Folk*, e de um *corpus* comparável, composto de um *subcorpus* em inglês e um *subcorpus* em português para atestar as palavras mais frequentemente empregadas pela comunidade homossexual, compilado pela ferramenta *Sketch Engine* (KILGARRIFF et al, 2004).

Ao enfatizar e adotar a perspectiva que privilegie o modo de falar natural no estudo das combinações lexicais, em especial as colocações, é necessário contar com o arcabouço teórico-metodológico da Fraseologia e da Linguística de *Corpus*.

Desse modo, para o presente trabalho foi proposto, como objetivo geral, o levantamento e a análise das colocações a partir do *corpus* paralelo formado pelas transcrições em inglês e em português do Brasil da série *Queer as Folk*, retiradas do *site* de *download* de legendas (*tvsubtitles.net*), com o propósito de observar o léxico fraseológico da comunidade homossexual de maior frequência, em relação ao léxico fraseológico da comunidade homossexual nas traduções das transcrições para o português do Brasil.

Como objetivos específicos, este trabalho procurou seguir as etapas metodológicas seguintes:

1. Compilar um *corpus* paralelo formado pelas transcrições em inglês e em português do Brasil do seriado *Queer as Folk*;
2. Compilar um *corpus* comparável, composto de um *subcorpus* em inglês e um *subcorpus* português, com o propósito de atestar as palavras e as colocações mais frequentemente empregadas pela comunidade homossexual;
3. Comparar o léxico fraseológico da comunidade homossexual de maior frequência no *corpus* paralelo em inglês com o respectivo léxico fraseológico da comunidade homossexual presente no *corpus* paralelo em português do Brasil;
4. Extrair e analisar as colocações mais frequentemente empregadas no *subcorpus* em inglês e suas respectivas traduções; e
5. Elaborar uma amostra de glossário bilíngue de colocações, na direção inglês→português, a partir das palavras de maior chavidade que possam retratar as combinatórias de palavras mais frequentemente empregadas pela comunidade homossexual com base no *corpus* de estudo paralelo compilado.

Dessa maneira, no Capítulo 1, realizamos uma contextualização sobre as motivações para o desenvolvimento desta pesquisa, tendo em vista que estudos da comunidade homossexual já vêm sendo desenvolvidos em outros países. Salientamos que tais estudos não abordam os fraseologismos, especificamente as colocações, existentes na comunidade homossexual e, mediante essa lacuna, optamos por desenvolver uma investigação sobre as colocações empregadas nesse contexto.

No Capítulo 2, abordamos a Fraseologia enquanto ciência, partindo do arcabouço teórico-metodológico com base em diversos estudos realizados, nos quais se discutem denominações terminológicas, classificações e definições dos limites com relação aos fraseologismos.

No Capítulo 3, apresentamos a metodologia utilizada em nossa pesquisa, detalhando cada etapa para a compilação de nossos *corpora*, bem como o levantamento e a extração das colocações da comunidade homossexual.

No Capítulo 4, tratamos da extração e análise das colocações da comunidade homossexual.

No Capítulo 5, apresentamos a Proposta de *Glossário Bilingue de Colocações da Comunidade Homossexual*, assim como seus contextos.

Por fim, nas considerações finais sobre este estudo interdisciplinar fundamentado na Fraseologia e na Linguística de *Corpus*.

Após as considerações finais, trazemos referências bibliográficas utilizadas nesta pesquisa.

1 *GAY TWIST*: UM TOQUE *GAY* PARA COMEÇAR

Apresentamos, no presente capítulo, uma contextualização sobre as motivações para o desenvolvimento desta pesquisa. Embora estudos da comunidade homossexual já venham sendo desenvolvidos em outros países, evidenciamos que tais estudos não abordam os fraseologismos, especificamente as colocações, existentes na comunidade homossexual e, mediante essa lacuna, optamos por desenvolver uma investigação sobre as colocações empregadas neste contexto.

Discorreremos, no capítulo 1, acerca da Linguagem e Sexualidade, posteriormente, da maneira que os *gays* falam e de uma reflexão sobre a existência ou não de uma *Gay Language*, finalizando o capítulo com a contextualização da série *Queer as Folk*.

1.1 Linguagem e Sexualidade

Em *Language and Sexuality*, Cameron e Kulick (2003) mostram-nos que a pesquisa sobre linguagem e homossexualidade, até aquele momento, passou por quatro fases principais. Desde a década de 1920 até a década de 1940, a homossexualidade era considerada patologia, uma perversão e uma ofensa criminal. A pesquisa precoce foi realizada por médicos e focada em confissões de pacientes pervertidos e como eles articularam suas perversões. Pensava-se que os pervertidos falavam um idioma particular, utilizavam as palavras de uma maneira também particular e usavam características linguísticas que podiam enquadrá-los como homossexuais.

Nas décadas de 1950 e 1960, na época em que a luta pelos direitos dos homossexuais surgiu, a pesquisa sobre linguagem *gay* foi realizada por estudiosos *gays* e estudiosas lésbicas. Havia uma luta para criar uma nova identidade homossexual que diferisse da anterior, da homossexualidade sendo uma patologia e perversão, e, portanto, havia uma divisão entre os “homossexuais antiquados e equivocados, que usavam a linguagem e os homossexuais politicamente progressivos e as lésbicas”¹ (CAMERON; KULICK 2003, p. 77).

Durante as décadas de 1970 e 1980, os homossexuais eram vistos como uma minoria oprimida, semelhante a outras minorias raciais e étnicas. Inspirado na pesquisa sobre estudiosos

¹ *Old-fashioned and misguided homosexuals, who used the language and the politically progressive gays and lesbians, who avoided it* (CAMERON; KULICK, 2003, p. 77). [Todas as traduções neste trabalho são de nossa responsabilidade].

da língua inglesa negra e da língua feminina, mais uma vez estudiosos da linguagem argumentaram sobre a existência de uma linguagem homossexual, mas não uma linguagem que refletia uma patologia, e sim que homossexuais, como outras minorias, tinham marcadores que eram definíveis (CAMERON; KULICK, 2003, p. 77).

Em 1981, Hayes escreveu um ensaio sobre *Gayspeak*, cujo conteúdo trata das funções sociais da linguagem na subcultura *gay*. O autor divide *Gayspeak* em três contextos sociais: a secreta, a social e a radical ativista. A secreta significa que o homossexual evitaria a terminologia *gay* em um ambiente inseguro e se referiria a seu amante como seu amigo ou evitaria referências de gênero como, por exemplo, não contar com quem ele saiu ou esteve de férias. No cenário social, o orgulho de ser *gay* é mostrado por meio de um léxico *gay* e, no ambiente ativista radical, os falantes politizam a vida social e reivindicam, por exemplo, palavras como *faggot* e *dick* (*viado* e *rola*, respectivamente.), fazendo delas símbolos de desafio (CAMERON; KULICK, 2006, p. 68-77).

Durante esta fase, Gaudio realizou uma análise da acústica da fala masculina registrada e estudou a percepção dos ouvintes de oito homens homossexuais e heterossexuais. O autor descobriu que os ouvintes eram geralmente capazes de identificar a orientação sexual do falante com base nas estereotipadas ideias de voz e assim por diante. Contudo, as diferenças não foram estatisticamente significantes e não ocorreram em todos os contextos de fala. Em outras palavras, nem todos os homens *gays* têm a voz diferenciada, e nem todos os que têm a voz diferenciada, são homossexuais (CAMERON; KULICK, 2003, p. 90).

Segundo Leap (1996), os homens *gays* aprendem a falar utilizando itens lexicais homossexuais com os outros integrantes da mesma comunidade, ir à biblioteca e assistir séries de televisão como *Queer as Folk*, *Will & Grace* e *Six Feet Under* ajuda-os nessa descoberta da linguagem *Gay*. De acordo com os homossexuais, esta é uma socialização que se faz por conta própria (BETSCH, 2008).

Na década de 1990, a pesquisa deixa de ver como a identidade *gay* é refletida por meio do idioma e passa a investigar como essa identidade é criada por meio da linguagem (CAMERON; KULICK, 2003, p.76).

Leap (1996) dá um exemplo de um estudante heterossexual que fala fluentemente a *Lavender Language* (*Lavender* é o termo usado para questões relacionadas à pronúncia, ao vocabulário e ao significado da língua da comunidade *gay*). De acordo com o aluno hétero, os homens *gays* foram atraídos por ele. Ele cresceu com sua mãe, que era artista, e trabalhou com homens homossexuais, e, por esse motivo, o conhecimento e o uso do discurso *gay* faziam com que os homens *gays* se sentissem atraídos por ele (BETCH, 2008).

Vários recursos foram propostos como marcadores de identidade masculina *gay*, como vocabulário, voz de alto nível, expressões femininas e comutação de código, ou, mais precisamente, identidade *gay* americana de classe média branca, já que esse é o grupo que tem sido o tema da maioria das pesquisas.

Dito isso, Barret (1998) realizou pesquisas sobre as *African American drag queens* e descobriu que elas usavam o estilo de mulher branca estereotipada, coocorrendo com referências sexuais e passavam para o estereotipado *African-American Vernacular English*. De acordo com Barret (1995), a alegria das *drag queens* não foi sinalizada por meio de códigos linguísticos discretos e movimentos discursivos. Em vez disso, as *drag queens* tornaram-se *gays* pela manipulação de identidades contraditórias (CAMERON; KULICK 2003, p. 98-99).

Baker (2002) publicou o livro *Polari - The Lost Language of Gay Men*, um exemplo britânico de linguagem *gay* que deriva de uma variedade de fontes, como animadores itinerantes nos séculos XVIII e XIX, uma linguagem de código entre criminosos. Não está claro o quão generalizado o *Polari* realmente estava entre os homossexuais, no entanto Baker (2002) mostra que era bem conhecido em 1940 e 1950 em cidades britânicas maiores. Com o aumento da exposição por meio do programa de rádio da BBC *Round the Horne*, perdeu seu código especificamente de linguagem *gay*. Para muitas pessoas que conheciam o *Polari*, não era uma linguagem *per se*, mas um léxico de substantivos, verbos e adjetivos que poderiam ser utilizados em lugar de palavras inglesas comuns do inglês padrão.

Baker (2002), entretanto, sugere que alguns falantes usaram o *Polari* de maneira a sugerir uma linguagem com sistema gramatical. Inicialmente, apresentava-se como uma arma e um escudo contra uma atmosfera repressiva na década de 1950, porém, na década de 1970, foi vista como banal ou limitante. O seu eventual desaparecimento foi, em parte, o resultado da ambivalência da subcultura *gay* em relação ao *Polari* (BAKER, 2002).

Destinamos a página seguinte à apresentação do quadro 1, no qual mostramos um resumo da trajetória da linguagem *gay*, percorrida no presente estudo.

PERÍODO	ESTUDIOSO	ACONTECIMENTO	RESULTADO
1910 - 1920	BRONTSEMA (2004)	A palavra <i>queer</i> coexistia com <i>fairy</i> , ambos em referência aos homossexuais. Porém com conotações diferentes. <i>Fairy</i> refere-se aos afeminados, extravagantes, enquanto <i>queer</i> , refere-se aos homens homossexuais masculinos.	Os heterossexuais não fazem qualquer distinção entre <i>queer</i> e <i>fairy</i> .
1920	BRONTSEMA (2004)	Os homens homossexuais afeminados começam a se referir a si próprios como <i>gays</i> .	A palavra <i>gay</i> é empregada com duplo sentido, quando utilizada por homossexuais. Apenas os familiarizados com o uso específico homossexual o entenderia.
1920 - 1940	MÉDICOS	Os homossexuais falavam em idioma particular, utilizavam as palavras de maneira particular.	A homossexualidade era considerada uma patologia, uma perversão e ofensa criminal.
1950 - 1960	ESTUDIOSOS GAYS ESTUDIOSAS LESBICAS	Início da luta pelos direitos homossexuais.	Criação de uma nova identidade homossexual, que difere da anterior, da homossexualidade como patologia e perversão.
1970 - 1980	ESTUDIOSOS DA LÍNGUA INGLESA NEGRA E DA LÍNGUA FEMININA	Os homossexuais eram vistos como uma minoria oprimida.	Argumentam sobre a existência de uma linguagem homossexual, mas não em uma linguagem que reflita uma patologia.
1981	JOSEPH HAYES	Escreve um ensaio sobre <i>Gayspeak</i> , que trata das funções sociais da linguagem na subcultura <i>gay</i> .	Divide <i>Gayspeak</i> em três contextos sociais: a secreta, a social e a radical ativista.
1990	CAMERON; KULICK (2003)	A pesquisa deixa de ver como a identidade <i>gay</i> é refletida por meio do idioma.	A pesquisa passa a investigar como essa identidade <i>gay</i> é criada por meio da linguagem.
1990	HOMOSSEXUAIS	Discussão do item lexical <i>queer</i> por meio do <i>Queer Nation</i> , formado em Nova York como grupo de discussão. Depois, como uma coalizão contra homofobia.	<i>Queer</i> tornou-se um item lexical que inclui <i>gay</i> e lésbica, contra a homofobia. Posteriormente incluiu bissexuais e transexuais.
1996	LEAP (1996)	Os homens <i>gays</i> aprendem a falar utilizando itens lexicais homossexuais com outro integrante da mesma comunidade.	A socialização contribui para a utilização desses itens lexicais.

Quadro 1: Trajetória da Linguagem *Gay*

Fonte: Elaborado pelo autor.

Ainda em relação à linguagem e à sexualidade, ou seja, a relação entre a prática linguística e a orientação sexual como categoria social e como um domínio de desejo, teve início e trajetória semelhante ao campo da linguagem e do gênero.

Para Lakoff (1975, p. 50), aqueles que trabalham com linguagem e sexualidade começaram a procurar correlações entre agrupamentos sociais (neste caso, *gays* e lésbicas) e maneiras pelas quais eles falavam. A visão de Jacobs da literatura descobriu que os estudiosos consideravam a variação fonológica, a variação gramatical, a padronização do discurso, as particularidades lexicais, as formas discursivas e a paralinguagem no exame de linguagem *gay* e lésbica.

Por exemplo, Leap (1993, p. 57) reconhece atributos únicos da linguagem dos homens *gays*, chamada por ele de *linguagem de risco*: os *gays* usam, rotineiramente, estratégias de linguagem que lhes permitam maximizar os ganhos (confirmação da identidade *gay*, linguagem *gay* bem sucedida) e minimizar as perdas (divulgação injustificada de interesses de gênero, contração heterossexista e homofobia). Esses primeiros trabalhos concentraram-se no estabelecimento do idioma e da sexualidade como um campo legítimo, ao demonstrar que as pessoas *gays* e lésbicas falavam de maneiras peculiares.

Kulick (2000) criou uma crítica importante de que pesquisadores de linguagem e sexualidade fundamentaram seu trabalho na ideia essencialista de que lesbianas e *gays* têm uma maneira diferente de falar que está enraizada em suas identidades como lésbicas ou homossexuais; em outras palavras, “linguagem *gay* e lésbica tornou-se um traço essencial constituindo identidades *gays* e lésbicas² (GAL; IRVINE, 1995). Ele escreve que “é importante não confundir recursos simbólicos de que qualquer um pode se apropriar para invocar imagens estereotipadas da homossexualidade com as práticas reais de linguagem, e muito menos as identidades, dos *gays* e lésbicas individuais”³ (KULICK, 2000, p. 257).

Kulick (2000, p. 264) aponta para tentativas de definir o objeto de estudo (ou seja, linguagem *gay*) e argumenta que são logicamente circulares. Ele faz uma objeção especial aos métodos e pressupostos de Leap (1996): “O que torna [o homem inglês] *gay*? O fato de que homens *gays* falam. Por que os homossexuais falam assim? Porque eles são homens *gays*. E assim por diante, sucessivamente”⁴.

² *Gay and lesbian language has become an essential trait constituting gay and lesbian identities* (GAL; IRVINE, 1995).

³ *[It] is important not to confuse symbolic resources that anyone can appropriate to invoke stereotypical images of homosexuality with the actual language practices, much less the identities, of individual gays and lesbians* (KULICK, 2000, p. 257)

⁴ *What makes [the English] gay? The fact that gay men speak it. Why do gay men speak it? Because they are gay men. And so on, round and round* (KULICK, 2000, p. 264)

Kulick (2000) descreve um avanço na teoria do campo quando Barrett (1995) e Queen (1997) desenvolvem a teoria de Butler (1993) sobre a *performance* do gênero para a linguagem e a sexualidade. Kulick (2000, p. 267) afirma que “o foco da pesquisa na linguagem *gay* deve ser deslocado de categorias de identidade para práticas significantes”⁵. Ainda assim, ele julga que a noção de desempenho traz consigo questões de intencionalidade: o orador pretende que seu discurso constitua ou signifique uma certa identidade (ou, pelo menos, o estereótipo de uma).

Existe a possibilidade de que qualquer um, independentemente de ser homossexual ou lésbica, possa utilizar a linguagem de *gay* ou de lésbicas. O autor considera que a teoria da *performance* é pouco mais que uma *proxy* (*representante*) para as categorias de identidade sexual que esperava suplantá-las. Kulick (2000, p. 70) justifica a identidade sexual pelo medo das identidades introduzirem uma limitação estruturalista na compreensão das pessoas que concebem sua sexualidade. O autor escreve,

porque o gênero tem uma forte tendência a ser analisado em termos de categorias de identidade mutuamente exclusivas (ou seja, *homem e mulher*), o risco parece muito grande de que uma análise da sexualidade também será enquadrada em termos de categorias exclusivas de identidade⁶ (KULICK, 2000, p. 70)

Em vez de considerar o papel da linguagem na construção da identidade sexual, Kulick (2000, p. 271) requisita um exame acadêmico do fenômeno, como a fantasia, o desejo, a repressão, o prazer, o medo e o inconsciente; no entanto, finalmente explica que muitos sentidos formam a sexualidade. Esta abordagem é descrita como uma abordagem de *linguagem e desejo*, ocasionando uma ruptura entre os pesquisadores que estudam linguagem e sexualidade: de um lado, aqueles que continuam interessados na constituição da identidade sexual; e, do outro, aqueles que favorecem uma definição de sexualidade concentrada na articulação e na atualização do desejo sexual.

Bucholtz e Hall (2004, p. 471) comprovam a retenção de identidade em estudos de sexualidade. Eles reconhecem a falta de coerência teórica no trabalho anterior, particularmente na visão de identidade e em como as identidades se tornam instanciadas na linguagem e na fala,

⁵ *The focus of research on queer language should be displaced from identity categories to signifying practices* (KULICK, 2000, p. 267)

⁶ *Because gender has a strong tendency to be analyzed in terms of mutually exclusive identity categories (namely, man and woman), the risk looms large that an analysis of sexuality will also be framed in terms of mutually exclusive identity categories.* (KULICK, 2000, p. 70).

mas veem o potencial na capacidade de linguagem e sexualidade para nos permitir conversar sobre ideologias, práticas sexuais, identidades como questões interligadas e sem perder de vista as relações de poder.

Bucholtz e Hall (2004, p. 475) afirmaram que o principal fundamento da oposição ao estudo da identidade com o idioma é a falácia de que as formas linguísticas devem ser atribuídas de maneira exclusiva a identidades particulares para serem socialmente significativas (GAL; IVINE, 1995). Eles empregam o termo *linguística* para descrever uma área de interesse em como as estruturas e práticas linguísticas constituem e apontam para identidades sexuais, direta ou indiretamente, seja intencional ou não, e para uma variedade de propósitos. Mais importante ainda, Bucholtz e Hall (2004, p. 477) escrevem que a noção de identidades essenciais não pode ser descartada porque os próprios atores sociais a utilizam para organizar e entender identidades.

Em outras palavras, as identidades essenciais estão disponíveis para os falantes como ideologias pelas quais eles podem conceber e organizar seus seres sexuais. Além disso, as formas em que os falantes podem ser descobertos por meio da análise das formas linguísticas ou, mais amplamente, também na prática social. Em suma, os autores escrevem que “os usuários [da Linguagem] utilizam e criam associações convencionais entre a forma linguística e o significado social para construir suas próprias identidades”⁷ (BUCHOLTZ; HALL, 2004, p. 478).

Bucholtz e Hall (2004, p. 486) também alertam sobre a adoção de Kulick (2000) a respeito da abordagem de linguagem e desejo por sua incompatibilidade com os métodos estabelecidos de etnografia linguística e de seu movimento em direção à psicanálise. Eles acreditam que a ruptura da teoria feminista com a sexualidade e as estruturas de poder já não enquadram mais uma visão limitada do desejo e da prática sexual.

Os autores argumentam que “o desejo não pode ser separado do poder e da instância”⁸ e enfatizam a importância das ideologias, práticas e identidades que produzem significados sociais da sexualidade (BUCHOLTZ; HALL, 2004, p. 486). Uma vez que a antropologia linguística enfatiza que a linguagem é o nível mediador entre as estruturas de poder e a instância humana, examinar a constituição de práticas e identidades sexuais na fala situada torna-se uma ferramenta potencialmente valiosa para entender como os indivíduos reproduzem sistemas existentes, bem como os alteram criativamente.

⁷ [Language] users both draw on and create conventionalized associations between linguistic form and social meaning to construct their own and others' identities (BUCHOLTZ; HALL, 2004, p. 478).

⁸ Desire cannot be separated from power and agency (BUCHOLTZ; HALL, 2004, p. 486).

Bucholtz e Hall (2004, p. 493) enfatizam que a identidade sexual é um resultado de práticas e ideologias negociadas intersubjetivamente, em vez de uma característica inerente de indivíduos intencionados; isso torna a identidade precisamente útil para cientistas sociais que buscam compreender as formas como a dualidade estrutural afeta o significado social. A natureza negociada da identidade torna-se especialmente susceptível à análise antropológica linguística.

Bucholtz e Hall (2004, p. 494) elaboraram um esquema teórico para a organização de práticas linguísticas em torno da negociação da sexualidade, o que chamam de *estrutura de intersubjetividade*. Essas táticas consistem em três eixos: semelhança/diferença, autenticidade/artifício e reconhecimento/marginalização, denominados pelos autores de *adequação/distinção*, *autenticação/desnaturalização* e *autorização/ilegitimação* respectivamente.

Eles afirmam que essas táticas abrangem a gama de atos linguísticos (tanto em formação como em interpretação) realizados pelos indivíduos no desempenho e na negociação de suas identidades. Como agentes empregam nessas táticas múltiplas formas interativas, pode surgir uma série complexa de formações identitárias, conforme pode ser observado no quadro abaixo:

	TÁTICAS POSITIVAS	TÁTICAS NEGATIVAS
SEMELHANÇA DIFERENÇA	ADEQUAÇÃO - PRÁTICAS QUE ESTABELECEM UMA SIMILARIDADE SUFICIENTE ENTRE UM INDIVÍDUO E OUTROS COM UMA IDENTIDADE PARTICULAR.	DISTINÇÃO - PRÁTICAS QUE AFIRMAM DIFERENÇAS ENTRE UM INDIVÍDUO E OUTROS COM PARTICULAR UMA IDENTIDADE PARTICULAR.
AUTENTICIDADE ARTIFÍCIO	AUTENTICAÇÃO - PRÁTICAS QUE CONSTROEM UMA IDENTIDADE COMO ALGO GENUÍNO, ESSENCIAL E/OU VERDADEIRO.	DESNATURALIZAÇÃO - PRÁTICAS QUE CONSTROEM UMA IDENTIDADE COMO ALGO PRETENSIOSO, NÃO ESSENCIAL OU MENTIROSO.
RECONHECIMENTO MARGINALIZAÇÃO	AUTORIZAÇÃO - PRÁTICAS DE PODER QUE LEGITIMAM CERTAS IDENTIDADES COMO "CULTURALMENTE INTELIGÍVEIS", ACEITÁVEIS.	ILLEGITIMAÇÃO - PRÁTICAS DE PODER QUE IMPEDEM A VALIDAÇÃO E ACEITAÇÃO SOCIAL DE CERTAS IDENTIDADES.

Quadro 2 – *Tactics of Intersubjectivity* (adaptado de BUCHOLTZ; HALL, 2004)

No quadro 2, podemos ver o termo *adequação* sendo considerado como as diferenças potencialmente salientes reservadas em favor de semelhanças percebidas ou afirmadas que são consideradas mais relevantes para a situação. Já *distinção* é o mecanismo pelo qual a diferença saliente é produzida. Distinção é, portanto, o inverso da adequação, na medida em que a diferença de relação é sublinhada e não apagada.

O termo *autenticação* refere-se a como os falantes ativam essas leituras essencialistas na articulação da identidade, vinculado pelo essencialismo por meio da noção de que algumas identidades são mais “reais” do que outras. *Desnaturalização* é a construção de uma identidade aceitável ou genuína e a produção de uma identidade.

Autorização pode envolver a invocação de linguagem de maneiras reconhecidas pelo estado. Enquanto *ilegitimação* envolve tentar legitimar uma identidade por meio de uma autoridade institucional ou outra, ou inversamente, o esforço para reter ou retirar esse poder estrutural.

No meio do debate em curso sobre o lugar da identidade na pesquisa de linguagem e sexualidade, mantemo-nos com aqueles que veem a identidade desempenhando um papel crucial na forma como os indivíduos pensam e falam sobre seu ser sexual. Kulick (2000), juntamente com Cameron (2006), continuam a insistir para que os pesquisadores façam do desejo sexual o principal foco do estudo da linguagem e sexualidade, mas os métodos de antropologia linguística (a saber, a etnografia do discurso e a análise do discurso interacional), bem como os da antropologia em geral, não se prestam ao exame do desejo interior.

Bucholtz e Hall (2004) enfatizam que essa abordagem se aproxima da psicanálise, ao dizer que os antropólogos não são qualificados para abordar tais assuntos, nem eticamente devem presumir estar fazendo com os métodos disponíveis para eles. Além disso, vimos que as identidades sexuais, embora talvez originalmente concebidas para facilitar a categorização de assuntos, podem ser consideradas de forma produtiva como recursos ideológicos, disponíveis para os indivíduos construírem discursivamente sua identidade sexual, de acordo com as táticas de intersubjetividade descritas por Bucholtz e Hall (2004).

Assim, entendendo o papel da identidade nas experiências dos sujeitos e observando as práticas linguísticas que constituem identidade, podemos esperar encontrar o dualismo vivo e instanciado do poder social e da agência individual na construção das identidades sexuais (ECKERT; MCCONNELL-GINET, 1992). Das fontes mencionadas, interessa-nos, especialmente, a maneira como os *gays* falam – Existe *Gay Language*? Tópico que discutiremos a seguir, no subitem 1.2.

1.2 A maneira como os *gays* falam – Existe *Gay Language*?

Algumas palavras e expressões podem ser empregadas para sinalizar e marcar identidade pessoal ou grupal (CHAMBERS, 2009, p. 39) e, durante séculos, homossexuais utilizaram certas palavras e expressões para protegê-los na busca por parceiros, assim como para ocultar sua opção sexual em um ambiente que não os aprova (BAKER, 2002).

Há uma grande variedade de epítetos referente a homossexuais, como *queer*, *fag*, *faggot* (todos esses itens lexicais podem ser traduzidos como: *viado*) e muitos outros itens lexicais, como *top*, *bottom*, *versatile* (*ativo*, *passivo*, *versátil*, respectivamente), que se concentram “em torno da avaliação e conquista sexual”⁹ (BAKER, 2002, p.157). Alguns rótulos ou epítetos são usados positivamente e alguns de forma negativa, muitas vezes dependendo de o falante ser homossexual ou não.

De acordo com Brontsema (2004), a palavra *queer* foi usada pela primeira vez para se referir a objetos, lugares e pessoas estranhas. *Queer*, mais tarde, adquiriu mais uma acepção ao se referir a homossexuais, geralmente um homem, e depois se tornou associado à sexualidade não-normativa e, com esse significado, persistiu até o presente. Nas décadas de 1910 e 1920, *queer* coexistiu com *fairy* (*fada*) para se referir a homossexuais, mas os termos foram usados em conotações de grupos muito diferentes. *Fairy* referia-se a homossexuais afeminados e extravagantes, enquanto *queer*, a homens homossexuais mais masculinos. Os heterossexuais, no entanto, não fizeram qualquer distinção entre os *queers* e *fairies*, independentemente da feminilidade/masculinidade ou auto identificação.

Na década de 1920, os homens homossexuais afeminados começaram a se referir a si próprios como *gays*.

Uma palavra inofensiva, como *gay* originalmente denotou uma brincadeira amigável, apesar de receber duplo significado quando usada por homossexuais. Somente aqueles que estivessem familiarizados com esse uso específico de homossexual, o entenderiam e, portanto, havia um risco pequeno ao usá-la com homens cuja sexualidade era desconhecida¹⁰ (BRONTSEMA, 2004, p. 3).

Homens homossexuais, especialmente jovens, achavam isso depreciativo e, mais tarde, os *queers* (homossexuais masculinos) começaram a usar *gay*.

⁹ *Around sexual evaluation and conquest* (BAKER,2002, p. 157).

¹⁰ *A safe word, gay originally denoted lighthearted pleasantness, yet was given a double meaning when used by homosexuals. Only those familiar with this specific use of gay would understand it, and therefore, there was initially very little risk in using it with men whose sexuality was unknown* (BRONTSEMA, 2004, p. 3)

A palavra *gay* reuniu todos os homens envolvidos sexualmente com outros homens em um grupo homogêneo. Como tal, os homossexuais ignoraram diferenças importantes entre esses homens, da mesma forma que ocorreu com o uso de *queer*, fora do grupo, apenas algumas décadas antes (BRONTSEMA, 2004, p. 4).

Na década de 1990, os homossexuais começaram a discutir o item lexical *queer* por meio da *Queer Nation*, que foi formada em Nova York como um grupo de discussão e, depois, como uma coalizão comprometida com a luta contra a homofobia. *Queer* tornou-se um termo que incluiu *gays* e lésbicas para lutar contra a homofobia e, posteriormente, incluiu, também, bissexuais e transexuais.

Há pessoas que se opõem à recuperação e outras que apoiam a recuperação da palavra, uma vez que é muito depreciativa. Os oponentes pensam que só podem ser autodegradantes e desrespeitosos e uma repetição de intolerância e ódio (BRONTSEMA, 2004, p. 1-7).

A pesquisa sociolinguística sobre os estilos de expressão dos homens homossexuais centrou-se na identificação das características que constituem uma variedade de fala monolítica, muitas vezes referida como *Gay Speech* ou *Gay Accent*. Os estudiosos argumentaram que o discurso de homens *gays* ou, alternativamente, de fala *gay*, diferenciava-se de outros estilos de fala nos níveis lexical (RODGERS, 1972), fonético (CRIST, 1997) e discursivo (LEAP, 1996). Apesar disso, não questionamos que alguns segmentos da comunidade masculina *gay* podem usar os recursos discutidos nesta pesquisa.

Em primeiro lugar, o pressuposto de que existe uma forma de falar *gay* singular homogênea a diversidade dentro da comunidade *gay*, apagando ou, pelo menos, considerando sem importância o inquérito sociolinguístico e as muitas subculturas que compõem a comunidade. A cultura *gay* engloba categorias reificadas, tais como *leather daddies*, *drag queens*, *gay prostitutes* e ativistas tanto convencionais como radicais, bem como comunidades locais de prática que podem nem sequer ter nomes.

A participação em uma das subculturas geralmente tem precedência sobre uma afiliação mais geral com a comunidade *gay* e as atividades sociais e, portanto, as oportunidades para trocas linguísticas costumam ser organizadas em torno de membros, não na comunidade masculina *gay* em geral, mas em suas subculturas. A distinção entre as subculturas é construída de forma estilística, por meio do modo de vestir, do uso e escolha de drogas, das preferências musicais e dos recursos linguísticos.

Em segundo lugar, rotular as características linguísticas como *gay* é muito geral e, também, corre-se o risco de não ser suficientemente geral ao simplesmente atribuir significados homossexuais a características linguísticas como *gay* que são compartilhadas em toda a

sociedade. Por exemplo, Leap (1996) identifica o discurso cooperativo como um marcador do *Gay Men's English*; todavia, Cameron (1998) afirma que o discurso cooperativo também ocorre entre os jovens homens heterossexuais.

Ao rotular o discurso cooperativo como uma característica especificamente *gay*, ignora-se seu uso por mulheres e homens heterossexuais. O que falta é uma análise que permita que o discurso cooperativo contribua para a heterossexualidade, em algumas situações, e para a construção de identidades homossexuais, em outras. Propomos o seguinte questionamento: existe *Gay Language*?

A linguagem *gay* e sua própria existência estão intimamente ligadas às ideias das comunidades de fala. A tentativa de definir o que marca a linguagem *gay* é frequente, mas alguns autores alegam que não há linguagem *gay*. Wolfe (2006) é contra a noção de uma comunidade *gay* e, conseqüentemente, da linguagem *gay* comum. No entanto, Stanley (1970), que também é contra, afirma que existem gírias homossexuais que não são conhecidas por todos os homossexuais. Estas gírias variam de acordo com o gênero e conforme o fato de o falante viver em um centro urbano ou em uma cidade afastada (KULICK, 2000).

Deve-se observar que quase todas as pesquisas preliminares sobre linguagem e sexualidade concentraram-se principalmente na linguagem de homens *gays* em áreas urbanas (CAMERON; KULICK, 2006).

Eckert (2006) sugere que as gírias homossexuais consistem em um conjunto vocabular básico que também é conhecido por muitas pessoas heterossexuais. Em *Gayspeak* (HAYES, 1981, p. 68) o autor diz que o discurso homossexual possui três funções específicas ou dimensões: primeiro, é um código secreto desenvolvido para proteção contra a exposição; em segundo lugar, é um código que permite ao usuário expressar uma ampla gama de funções dentro da subcultura *gay*; e, em terceiro lugar, é um recurso que pode ser usado por ativistas radicais como meio de politizar a vida social.

Kulick (2000, p. 257) afirma que as pesquisas sobre linguagem *gay* ampliaram o conhecimento sobre a subcultura homossexual, porém não conseguiram apresentar quaisquer características estruturais, morfológicas ou fonológicas únicas para homens *gays*. Este ponto de vista é semelhante à crítica de Darsey (1981).

Morrish e Sauntson (2007) defendem a existência da linguagem *gay*, referindo-se a autores como Leap (1996) e Moonwomon (1995). Mencionam especialmente os dois livros de Leap: *Beyond the Lavender Lexicon* (1995) e *Word's Out* (1996). De acordo com Morrish e Sauntson (2007, p. 7), o trabalho de Leap sobre a linguagem *gay* examina a forma pela qual a identidade *gay* se materializa por meio do discurso. O trabalho de Leap não é uma busca direta

das propriedades linguísticas do discurso *gay*. Em vez disso, ele examina como os homossexuais constroem e sinalizam identidade e diferença por meio de linguagem codificada e explícita. Morrish e Sauntson (2007, p. 8) enfatizam o fato de que o contexto e a contingência fazem sentido: sem um contexto adequado, não há linguagem *gay*.

Os estudiosos mostraram que a linguagem *gay* é uma socialização autogerenciada. Os homens homossexuais, em busca de uma identidade, estudam a língua dos homossexuais lendo livros e assistindo séries de televisão, como *Queer as Folk*. A internet tornou muito mais fácil, hoje em dia, fazer novos amigos em todo o mundo e fazer parte de comunidades e redes sem sequer estar fisicamente presente.

Por meio de um contexto adequado, uma situação de diálogo entre os integrantes da comunidade homossexual, podemos afirmar a existência de uma *Gay Language*, ou seja, linguagem *gay*. Leap (1996) mostra que a linguagem utilizada por essa comunidade em especial difere-se em níveis lexicais, fonéticos e discursivos.

No presente trabalho focamos o estudo no nível lexical, estudando as colocações comumente empregadas pela comunidade homossexual. Consequentemente, levando em consideração a necessidade do indivíduo de pertencer ao grupo, autogerenciamento, concluímos que a *Gay Language* existe.

Com base no exposto neste capítulo, apresentamos uma breve ambientação da comunidade homossexual, objeto de estudo em nossa pesquisa, por meio da série *Queer as Folk*, a partir da qual temos como objetivo prático a elaboração de uma amostra de *Glossário Bílingue de Colocações da Comunidade Homossexual*.

1.3 Contextualizando a Série *Queer as Folk*

Na série *Queer as Folk*, há um diálogo entre os personagens Ted e Emmett, descrito abaixo:

- Ted: Ele [Michael] não é como você, ele não é um homem visivelmente *gay*.
- Emmett: Você está me acusando de ser óbvio? Eu poderia ser um homem real se eu quisesse. Basta diminuir a minha voz, parar de gesticular com as mãos, e cuidar para que meu rosto fique sem expressão.

Nunca, nunca usar palavras como *fabulosa* ou *divina*¹¹ (*Queer as Folk*, 102¹²).

Como podemos observar no fragmento acima, o personagem Emmett faz alusão a dois itens lexicais: *fabulosa* e *divina*, que, segundo ele, remetem ao léxico da comunidade *gay*, ao afirmar que, se usar tais itens lexicais, ele seria rapidamente apontado como *gay*.

Sociolinguistas (KAISER, 2000; BECKER, 2006; CAMERON, KULICK, 2006; CHAMBERS, 2009) discutiram a relevância da Linguística *Queer* e a existência da linguagem *gay*. Alguns afirmam que tal linguagem existe e que, de fato, difere em nível lexical, fonético e discursivo da fala heterossexual. Outros afirmam que é apenas parte da criação de marcadores linguísticos como em qualquer outra rede da sociedade. Por exemplo, pesquisas anteriores mostraram que a linguagem *gay* é autogerenciada. Para Chambers (2009), autogerenciado é algo que as pessoas aprendem porque querem fazer parte de um grupo com o qual elas se associam e se identificam.

Como a linguagem *gay* é autogerenciada, é relevante realizar pesquisas acerca do léxico empregado na série *Queer as Folk*, que apresenta a vida dos homens *gays* de maneira mais autêntica do que muitas outras séries de televisão, filmes e *sit-coms*. O discurso heterossexual é frequentemente equiparado ao discurso apropriado para o gênero e a linguagem *gay* é, muitas vezes, considerada inadequada ao gênero (CAMERON; KULICK, 2006, p. 74).

Os personagens, em vários momentos na série, dizem que são diferentes dos heterossexuais, que precisam mudar seus maneirismos e falas para se adequarem e serem aceitos. Dito isto, localizamos o espaço físico no qual a história decorre e analisamos as colocações utilizadas pela referida comunidade, a fim de elaborarmos um *Glossário de Colocações da Comunidade Homossexual*.

Queer as Folk desenvolve-se em Pittsburg (Pensilvânia) e gira em torno da vida de um grupo de seis homossexuais, um casal de lésbicas, uma mãe, o tio do personagem Michel e um adolescente. Três dos atores do sexo masculino declararam que são homossexuais e uma das atrizes femininas se declara bissexual.

Uma vez que a linguagem *gay*, de acordo com pesquisas anteriores, é autogerenciada, é interessante notar que os personagens são muito diferentes em termos de *backgrounds*,

¹¹ *He [Michael] is not like you, he is not an obviously gay man". Emmett answers: "Are you accusing me of being obvious? I could be a real man if I wanted to. Just lower my voice, stop gesturing with my hands, and make sure my face is expressionless. Never, never use words like fabulous or divine (Queer as Folk 102).*

¹² Referências à série *Queer as Folk* terão o seguinte formato: 102 = temporada 1, episódio 2; 310 = temporada 3, episódio 10 e assim por diante.

educação, profissão e promiscuidade: podemos citar o personagem Brian, *gay* alfa¹³ e publicitário; o afeminado Emmett; o jovem estudante de arte Justin; Ted, o contador; Michael, o filho de Debbie, a garçonete; Vic, o tio HIV positivo; o professor da faculdade, Ben; Melanie, a advogada; Lindsay, que trabalha em uma galeria de arte; e Hunter, o adepto da adolescência. Apesar de todas as diferenças, eles são amigos e familiares íntimos e compartilham as mesmas preocupações e lutas da vida.

A série contém muitas cenas inovadoras e, pela primeira vez na televisão americana, uma série mostrou cenas de sexo explícitas entre dois homens, o homem de 29 anos (Brian) e o menino de 17 anos (Justin). Charles Kaiser escreveu no blog *New York Entertainment*: “a série possui a linguagem mais escandalosa que você já ouviu, [e] alguns dos personagens mais convincentes que você já conheceu”¹⁴ (KAISER, 2000).

Os personagens homossexuais não eram novidade em séries americanas de televisão na época, mas, Segundo Kaiser (2000), “ninguém nunca viu personagens homossexuais como esses, personagens tão reais, matizados e hedonistas como muitos dos seus homólogos da vida real”. Os personagens *gays* e o vocabulário *gay* tornaram-se cada vez mais frequentes durante a década de 1990 com *Ellen* e mais tarde *Will & Grace*, no entanto, esses seriados nunca trouxeram de modo tão explícito na linguagem o sexo *gay* quanto *Queer as Folk*. Até então, os personagens homossexuais na televisão tinham, em sua maior parte, características “fofas” e ambíguas e nunca tinham mostrado dois homens juntos na cama (BECKER, 2006, p.175).

Além dos integrantes da série, realizamos de forma concisa uma breve discussão sobre o título da série. O título *Queer as Folk* deriva do antigo *Yorkshire*, podendo ser traduzido como “não há nada tão estranho quanto popular”, o que significa que “não há nada tão estranho quanto as pessoas” (KAISER, 2000), frase que resume a série. Nas palavras do personagem Michael:

Em muitas maneiras, minha vida não é nada como a sua. Por que deveria ser? Na comunidade *gay*, temos *drag queens*, *leather daddies*¹⁵ e *trannies*¹⁶, casais com crianças, todas as cores do arco-íris. [...]. Mas ser diferente é o que nos torna todos iguais. É o que nos faz ... família¹⁷ (*Queer as Folk*, 513, tradução nossa).

¹³ O homem *gay* alfa sabe exatamente o que quer, quem é, e não tem problema nenhum com a sua sexualidade, ou seja, é bem resolvido, independente, seguro de si (Nossa explicação do item lexical).

¹⁴ *No one has ever seen gay characters quite like these, characters just as real and nuanced and hedonistic as many of their real-life counterparts* (KAISER, 2000).

¹⁵ Homem mais velho envolvido sexualmente em um relacionamento ou apresenta interesse sexual em homens mais jovens (Nossa explicação do item lexical).

¹⁶ São homens que se sentem atraídos por transexuais por que se parecem com mulheres (Nossa explicação do item lexical).

¹⁷ *In many ways, my life is nothing like yours. Why should it be? In the gay community, we have drag queens and leather daddies and trannies and couples with children, every colour of the rainbow. [...] But being different is what makes us all the same. It's what makes us... family* (*Queer as Folk* 513).

Queer as Folk é uma coprodução da série de televisão americana e canadense que foi transmitida entre os anos 2000 e 2005, produzida pela *Showtime* e *Temple Street Productions* e com base na série britânica de mesmo nome, criada por Russell T. Davies.

Os principais autores e criadores da série norte-americana são o casal abertamente *gay* Ron Cowen e Daniel Lippman, que também são produtores executivos juntos com Tony Jonas, ex-presidente da *Warner Brothers Television*. De acordo com Cowen e Lippman, a série foi “projetada para agitar as pessoas, [e] levá-las a falar. Nós tentamos mostrar a verdade, manchas e tudo”¹⁸ (GRANT, 2010). No Capítulo 2, a seguir, tratamos da fundamentação teórica.

¹⁸ *Designed to shake people up, [and] get them talking. We tried to show the truth, blemishes and all* (GRANT, 2010).

2 FUNDAMENTAÇÃO TEÓRICA

Para o levantamento, análise e, posteriormente, elaboração de uma amostra prática de *Glossário Bilíngue de Colocações da Comunidade Homossexual*, faz-se necessário discorrer sobre os pressupostos teóricos que embasam a presente pesquisa. Iniciamos com a Fraseologia enquanto ciência que possui um arcabouço teórico-metodológico com diversos estudos realizados, nos quais se discutem denominações terminológicas, classificações e definições dos limites com relação aos fraseologismos.

Para a condução desta pesquisa, estamos amparados pelo referencial teórico da Fraseologia (ZULUAGA, 1980; HAUSMANN, 1984, 1985, 1990; SINCLAIR, 1991; STUBBS, 1995; CORPAS PASTOR, 1996; ORTIZ ALVAREZ, 1997; COWIE, 1998, 1999; MOON, 1998; TAGNIN, 1999, 2013; WRAY, 2002; ORENHA-OTTAIANO, 2004, 2009), na sua interface com os estudos da Linguística de *Corpus* (SINCLAIR, 1991, 1996, 2003; TOGNINI-BONELLI, 1996, 2001, 2004, 2010; BIBER, CONRAD, REPPEN, 1998; KENNEDY, 1998; MCENERY, WILSON, 2001; HUNSTON, 2002; BERBER SARDINHA, 2004, 2009; MEYER, 2004; MCENERY, HARDIE, 2012), cujos aspectos são discutidos em seguida, por meio do seguinte percurso: *Trajatória dos estudos Fraseológicos, As colocações da língua geral, Linguística de Corpus, Definição e Tipologia de Corpus*.

2.1 Trajetória dos estudos fraseológicos

Podemos dizer que a Fraseologia é uma disciplina recente e que sempre esteve à margem nos estudos linguísticos. Os primeiros registros formais de estudos fraseológicos datam do século XX, com Charles Bally (*Précis de Stylistique*, 1905; *Traité de Stylistique*, 1909; e *Linguistique générale et linguistique française*, 1932), que foi aluno de Ferdinand de Saussure e é considerado por Zuluaga (1980) como o “pai” da Fraseologia, já que foi o primeiro a estudar os fenômenos fraseológicos.

Ortiz Alvarez (1997) mostra-nos que as pesquisas pioneiras sobre as combinações fixas de lexemas foram realizadas por Bally (1909). Entretanto, Silva (2006) diverge de Ortiz Alvarez ao dizer que, embora a constituição da Fraseologia como área de investigação seja contemporânea, Saussure (1916) já teria feito referência às locuções, antes de Bally (1909). O

mérito e a contribuição de Bally não são descartados, todavia, por ter sido o primeiro a apresentar o termo *phraséologie*, de acordo com Silva (2006).

Nessa mesma linha de pensamento que apresenta Saussure como precursor dos estudos fraseológicos, Corpas Pastor (1996) assegura que a Fraseologia teve suas origens na década de 1950, na antiga União Soviética, região que propiciou uma enorme contribuição para os estudos sobre a Fraseologia.

O fato é que tanto a obra de Bally como os muitos estudos dos linguistas soviéticos trouxeram grandes contribuições para os estudos fraseológicos e ajudaram a firmar a Fraseologia enquanto disciplina, embora, ainda, atualmente, haja uma polêmica em definir a Fraseologia como uma disciplina autônoma ou como uma subdisciplina da Lexicologia. Klare (1986, p. 356) afirma que o ponto de vista soviético entende que a Fraseologia deve ser vista como disciplina autônoma, excluindo-a da Lexicologia e igualando-as, uma vez que os fraseologismos, tais como as locuções, também possuem suas próprias particularidades e, portanto, são suficientes para a investigação fraseológica.

Além das pesquisas soviéticas, que constituíram toda uma escola Russa de Fraseologia, os estudos fraseológicos ganharam grande repercussão em línguas europeias, como o espanhol e o francês, com estudos descritivos, contrastivos e históricos recolhidos e sistematizados por Häusermann (1977) e por Carneado e Tristán Pérez (1985), conforme explica Corpas Pastor (1996).

Importantes linguistas e pesquisadores dessas línguas contribuíram e contribuem para o fortalecimento desse campo de pesquisa. Dentre eles, podemos destacar Casares (1950), na Espanha, que foi o primeiro a estudar sistematicamente as unidades fraseológicas, especialmente as locuções e os refrãos da língua espanhola e, por conseguinte, o primeiro a propor uma classificação fraseológica. Os grupos fraseológicos propostos por ele foram: locuções, frases proverbiais, refrãos e modismos, dedicando-se, na maior parte, aos estudos das locuções. Demostramos, na figura 1 a seguir a classificação proposta pelo autor:

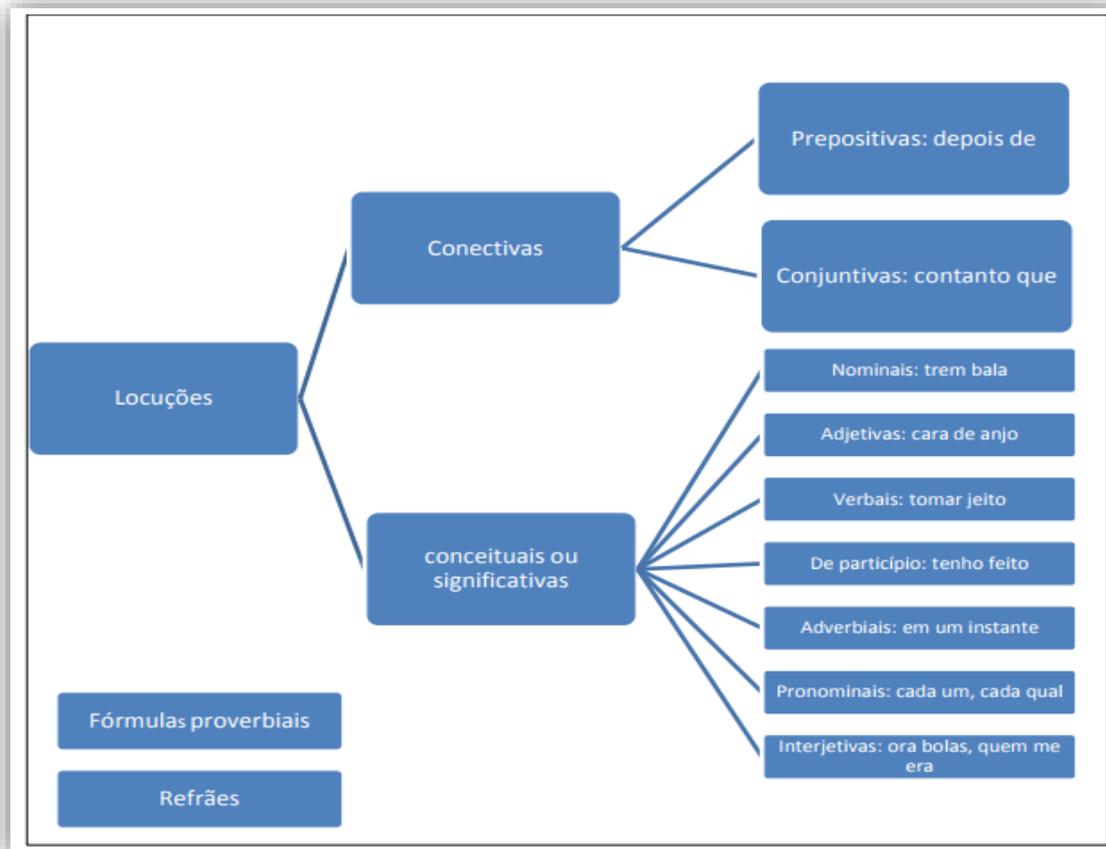


Figura 1 – Classificação de Casares (1950) (CARVALHO, 2016, p. 34)

Zuluaga (1980) define as expressões compostas minimamente por duas palavras e suas características internas e semânticas. Dentre as internas, chama a atenção para as características de *fixação* e *idiomaticidade* que, segundo o autor, variam em graus de acordo com cada expressão. Além de discutir acerca dos diferentes graus dessas características nas expressões em estudo, o autor afirma que as unidades fraseológicas podem se apresentar em nível de palavra, sintagma, oração e texto. Com base nessa afirmação, propõe uma divisão entre locuções e enunciados fraseológicos. A seguir, na figura 2, temos a classificação proposta pelo autor:

Locuções	Instrumentos gramaticais	Prepositivas Conjuntivas Elativas	
	Unidades léxicas	Nominais Adnominais	Cláusulas Circunstanciais Advérbios
		Adverbiais	
Sintagmas	Verbais		
Enunciados	Frases	Clichês Fórmulas Ditos	
	Textos	Refrões	

Figura 2 – Classificação de Zuluaga (1980) (CARVALHO, 2016, p. 36)

Podemos ver, na figura 2, que Zuluaga (1980) elabora dois grupos: no primeiro, encontramos os elementos oracionais, gramaticais ou sintáticos, enquanto, no segundo, o autor insere as frases e os textos.

Coseriu (1981), propõe outra tentativa de classificação. Ele distingue as expressões da língua em *técnica livre do discurso* e *discurso repetido*. No primeiro grupo, estariam as combinações livres de palavras que seguem regras semânticas e morfológicas e, no segundo, as formações fixas, ou seja, pré-fabricadas que se repetem sempre da mesma forma e fogem, muitas vezes, das regras, os fraseologismos. Dentro desse grupo estão as unidades fraseológicas que se subdividem em outros três grupos: as locuções que equivalem à oração, além dos refrões, frases proverbiais, ditos, sentenças e citações.

Tristá (1988) traça um perfil das expressões ou formas expressivas que são comuns em uma comunidade de fala e que, por suas características, podem ser consideradas assemânticas ou ilógicas, fazendo uma análise sintático-semântica. Além disso, ela descreve como ocorrem os processos de *fraseologização* e de *desfraseologização*, considerando o contexto em que estas unidades se inserem. Classifica como fraseologia todas as combinações de duas ou mais palavras que apresentam fixação no uso.

Corpas Pastor (1996) define como unidade fraseológica todas as combinações de palavras que apresentam fixação, alta frequência e estão institucionalizadas na língua, como graus de idiomatidade e variação. Nesse grupo, a autora considera desde as locuções a pequenos textos, como os provérbios e ditos populares. Veremos abaixo, na figura 3, a classificação proposta pela autora:

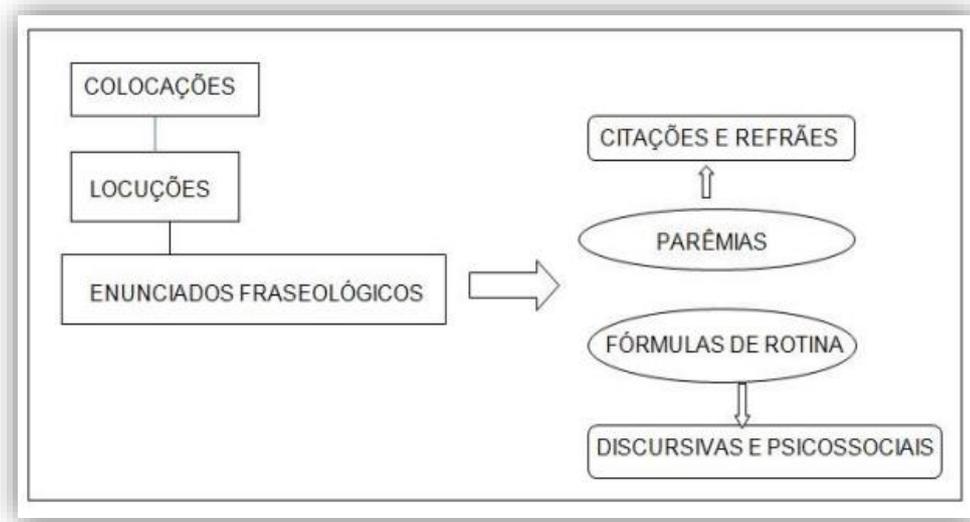


Figura 3 – Classificação de Corpos Pastor (1996) (CARVALHO, 2016, p. 39)

Na figura 3, Corpos Pastor (1996) desmembra os fraseologismos em três grupos: colocações, locuções e enunciados fraseológicos. Para a autora, as colocações são sintagmas livres que apresentam certa fixação, determinadas pelo uso. Observamos a divisão de dois grupos que apresentam elementos oracionais e gramaticais.

Ruiz Gurillo (1997) baseia-se nos estudos fraseológicos anteriores para propor uma classificação não discreta para as unidades fraseológicas. A autora considera a ideia de centro e de periferia, utilizadas pela Escola de Praga, em um *continuum* em que as unidades fraseológicas estariam inseridas em graus.

Ruiz Gurillo chama a atenção para a fixação e a idiomaticidade dessas expressões e as coloca como características essenciais a essas unidades linguísticas. No entanto, a autora destaca também que outras características podem ser encontradas nas expressões, tais como a motivação metafórica e frequência de uso, corroborando, portanto, para sua fixação ou idiomaticidade. Temos na figura 4, a seguir, a classificação proposta pela autora:



Figura 4 – Proposta de Ruiz Gurillo (1997) (CARVALHO, 2016, p. 41)

Ruiz Gurillo (1997), como vemos na figura 4, classifica os fraseologismos em sintagmas nominais, verbais e preposicionais, podendo variar de centro à periferia na medida que são mais ou menos fixos, sendo mais ou menos variáveis.

Burger (1998) aborda as características, as classificações e os problemas envolvendo os fraseologismos e sua inserção em materiais como o dicionário. Ele considera fraseologismo como qualquer expressão com duas ou mais palavras que apresentem como características polilexicalidade, fixação e idiomaticidade.

Para estruturar os estudos acima mencionados de forma breve, apresentaremos o quadro 3, em que estão presentes alguns teóricos da Fraseologia, o ano de sua obra, o termo adotado para se referir aos fraseologismos e uma breve definição desenvolvida por seus autores sobre o que seriam os fraseologismos ou as unidades fraseológicas.

FRASEOLOGIA			
AUTOR	ANO	TERMO	DEFINIÇÃO
Casares	1950	Unidade pluriverbais	Combinação estável de dois ou mais termos que funciona como elemento oracional e cujo sentido unitário consabido não se justifica, a não ser, como uma soma do significado normal dos componentes.
Zuluaga	1980	Unidades (expressões) fixas	Expressões de ao menos duas palavras com características internas, fixação, e semânticas, idiomaticidade.
Coseriu	1981	Discurso repetido	Formações fixas, ou seja, pré-fabricadas que se repetem sempre da mesma forma, fugindo, muitas vezes, das regras, os fraseologismos.
Tristá	1988	Fraseologismos ou Unidades Fraseológicas	Combinação de duas ou mais palavras que apresentam fixação no uso.
Corpas Pastor	1996	Unidades Fraseológicas	Todas as combinações de palavras que apresentam fixação, alta frequência e estão institucionalizadas na língua, com graus de idiomaticidade e variação.
Ruiz Gurillo	1997	Unidades Fraseológicas	Sintagmas que apresentam graus de fixação e idiomaticidade variando em um <i>continuum</i> .
Burger	1998	Expressões fraseológicas	Expressão com duas ou mais palavras que apresente como características polilexicalidade, fixação e idiomaticidade.

Quadro 3 – Quadro de autores (CARVALHO, 2016, p. 42-43)

Após essa breve explanação de alguns teóricos que trabalham na sistematização das unidades fraseológicas, entendemos a necessidade, como em toda área do conhecimento, de delimitação de um objeto de estudo. Desse modo, nos estudos fraseológicos, temos a unidade fraseológica (UF) como elemento central à Fraseologia, como seu objeto estudo.

Esses elementos da língua foram amplamente estudados pelos pesquisadores e linguistas e receberam diferentes nomenclaturas. Dentre as diferentes nomenclaturas utilizadas por diversos autores, adotamos as definições dos fraseologismos propostas por Orenha-Ottaiano (2004), tendo em vista que discutir as divergências de classificação não é o foco deste estudo. Abaixo, apresentamos alguns fraseologismos e suas definições, elencados por Orenha-Ottaiano (2004):

- **As expressões idiomáticas** – formam estruturas sintagmáticas complexas que resultam numa unidade lexical que se refere a uma realidade específica com um sentido particular.
- **As colocações** – são combinações lexicais fixas, recorrentes e arbitrárias.
- **Os marcadores conversacionais** – são expressões que indicam a contribuição e intenção do falante no que se refere a sua participação na conversação (KELLER, 1979).

- **As fórmulas de rotina** – são expressões cuja ocorrência é relacionada a situações sociais específicas e altamente previsíveis em determinados discursos.
- **Os verbos frasais** (*phrasal verbs/two-word verbs/three-word verbs*) – esta classe não existe em português. São formados por um verbo seguido de uma partícula adverbial ou, no caso dos *three-word verbs*, de um verbo seguido de uma partícula adverbial e uma preposição que indica a regência do *phrasal verbs*.
- **Os provérbios** – Segundo Houaiss (2001), o provérbio é uma “frase curta, geralmente de origem popular, frequentemente com ritmo e rima, rica em imagens, que sintetiza um conceito a respeito da realidade ou uma regra social ou moral”.
- **As frases feitas** – De acordo com Tagnin (1989), as frases feitas diferem das fórmulas de rotina porque, se não as utilizarmos em um dado contexto, não implicarão a ruptura dos padrões comportamentais convencionais.
- **Os binômios** – são formados por duas palavras pertencentes à mesma categoria gramatical, ligadas por uma conjunção ou preposição (MALKIEL, 1959).

Pelo exposto, notamos que a Fraseologia ainda necessita de estudos que fortaleçam a teoria para que haja consenso entre os estudiosos. No entanto, algumas características são citadas por todos ou grande parte desses teóricos, o que caracteriza os fraseologismos e permite que essa ciência abarque um grande leque de elementos que podem ser considerados objetos de estudo.

Antes de realizarmos uma análise das colocações da comunidade homossexual, presentes em nosso estudo, destacaremos alguns aspectos básicos em relação às nomenclaturas aqui abordadas, especialmente o conceito de UF e como ela pode ser melhor delimitada.

As marcas externas para reconhecer uma determinada categoria de UF estão relacionadas à forma de grupo, à ordem fixa dos elementos, às possibilidades reduzidas de separação, à impossibilidade de substituir um elemento ou outro. Considerando que as marcas internas estão relacionadas ao fato de que todo o conjunto incorpora um ato de pensamento unitário, equivalente a uma única palavra.

Os termos utilizados para designar uma UF, que receberam mais atenção na literatura linguística, foram frases e expressões idiomáticas. Muitas vezes, sequer existe uma distinção clara entre esses dois termos e seu uso paralelo com o mesmo significado, sendo uma prática comum.

Na literatura que trata da fraseologia, diferentes termos, como *idiom* (COWIE et al, 1983), *phraseme* (MEL'CUK et al, 1995) ou *unidades lexicais múltiplas* (ALVAREZ, 2000) foram frequentemente usados para se referir à mesma categoria. Cada um deles é definido de acordo com diferentes critérios e, por esse motivo, cada termo leva a definições e visualizações

mais amplas (MAKKAI, 1972; LUA, 1998) ou mais estreitas (FERNANDO; FLAVELL, 1981).

No contexto espanhol que lida com a mesma disciplina, as coisas não são dispares, como o emprego de diferentes termos para se referir ao mesmo objeto de estudo, tais como: *locuciones* (CASARES, 1992), *unidades del discurso repetido* (COSERIU 1967), *unidades fraseológicas* (CORPAS PASTOR, 1996).

Alvarez (2000, p. 70) afirma que os fraseologismos “são unidades lexicais múltiplas que apresentam vários graus de transparência semântica, que variam desde uma maior transparência à total opacidade”. De modo geral, os fraseologismos são expressões que possuem como traço distintivo a fixação e, assim, seus componentes caracterizam-se por certa estabilidade sintático-semântica, que geralmente não são comutáveis e não se separam, funcionando como uma única unidade de sentido, motivo pelo qual são denominadas unidades fraseológicas (ZULUAGA, 1980).

Sendo assim, as “expressões fixas se reproduzem em blocos, são construções anteriores ao ato de fala que as cria, mas que as repete [...] são produtos de processos de repetição na diacronia da língua” (ZULUAGA, 1980, p. 16).

Diante do exposto, percebemos que a delimitação dos fraseologismos é uma tarefa complexa para os pesquisadores da área, como podemos ver em Nogueira (2008):

[...] determinar os limites das UFs, classificando-as satisfatoriamente, não tem sido uma tarefa fácil para aqueles que se propuseram a fazê-lo, tampouco se têm obtido resultados que atendam completamente às expectativas dos estudiosos da área. Definir e classificar as UFs constitui-se, portanto, num dos aspectos mais controvertidos da investigação no campo da fraseologia. (NOGUEIRA, 2008, p. 62).

Em suma, a UF ou o fraseologismo constitui o objeto da Fraseologia, marcado por seu caráter expressivo e determinado por combinações de duas ou mais unidades lexicais. É uma ligação fixa ou com certos graus de fixidez estrutural e semântica, cristalizada e, assim, intimamente relacionada à norma linguística, estabelecendo-se como um ato de língua referente à diacronia, imposto pela comunidade de fala e submetido ao contexto.

No entanto, apesar da heterogeneidade de termos, parece haver um acordo geral na medida em que uma UF é denominada como uma combinação de palavras fixas cujas características principais estão resumidas no *Manual de fraseología española*, de Corpas Pastor (1996). Resumimos as principais características das UFs de acordo com a autora:

- é uma expressão feita de várias palavras;
- é institucionalizada (institucionalização);
- apresenta algum tipo de especificidade semântica ou sintática (idiomatização);
- tem diferentes graus de estabilidade (gradualidade e estabilidade);
- é possível uma certa variação de seus componentes (variação);
- geralmente é caracterizada por uma alta frequência de uso.

Na sequência, discorreremos acerca das colocações, objeto principal de nosso estudo, buscando, assim, uma delimitação desse tema.

2.2 As colocações da língua geral

O termo *colocação* foi usada pela primeira vez em um sentido técnico pelo linguista britânico Firth (1957), ao propor que o significado de uma palavra seja parcialmente determinado pelo seu ambiente contextual (FIRTH, 1957, p. 194). No entanto, segundo aponta Orenha-Ottaiano (2016), o termo foi realmente utilizado por Jespersen, em 1917, e, posteriormente na década de 1950, por Palmer. A visão principal desta abordagem questiona a ideia de classes de palavras separáveis (GLEDHILL, 2000; SINCLAIR, 1991) e, em vez disso, concebe o idioma como uma gradação de palavras gramaticais de classe fechada, em uma extremidade, para léxico de classe aberta, no outro (HALLIDAY, 2004), como mostrado na figura 5:

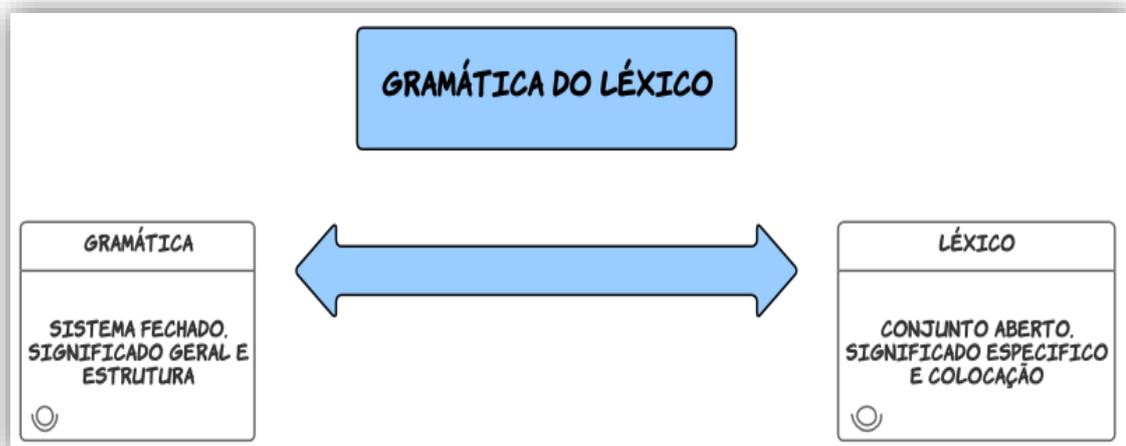


Figura 5 – *Lexico-grammar cline* (adaptado de HALLIDAY, 2004, p. 43)

Ao contrário da abordagem formalista de um sistema baseado em regras internas, idioma, aqui, é visto como um recurso de escolhas para expressar significado (BEAUGRANDE, 1996; MARTIN, 1992) e a colocação neste sistema está no extremo léxico da gradação. A noção de colocação também indica que, ao invés de serem limitadas apenas pelas forças sintáticas, existem influências paradigmáticas e textuais sobre a escolha léxica, o que explica o fato de que, por exemplo, *strong tea* e *powerful car* são aceitáveis, mas **powerful tea* ou **strong car* não são¹⁹ (HALLIDAY, 2005).

Este conceito levou Sinclair (1991, p. 110) a propor uma gradação lexical semelhante, do *idiom principle* (*princípio idiomático*), pela qual o usuário do idioma está disponível a um grande número de frases pré-construídas que constituem escolhas únicas para o *open choice principle* (*princípio da livre-escolha*), em que uma palavra constitui uma *separate choice* (*escolha separada*) (SINCLAIR, 1991, p. 175).

Todavia, embora a colocação e a sua influência na escolha linguística possam ser prontamente observadas, o seu preciso papel dentro do texto não é claro. Um texto pode ser definido como uma unidade de língua em uso (HALLIDAY, HASAN, 1976, p. 1) e se distingue do não-texto pelo duplo conceito de unidade: *unity of structure* e *unity of texture* (HALLIDAY; HASAN, 1985), também, geralmente, denominadas *coerência* e *coesão* (CARTER, 1998).

A coesão versa sobre a forma como o texto se liga internamente em termos de suas relações coesivas e é formada quando um elemento de um texto é dependente de sua interpretação em outro (HALLIDAY, HASAN, 1976; LECKIE-TARRY, 1995). Sem ela, as características de superfície de um texto podem não se relacionar entre si (CARTER, 1998, p. 103) e é, portanto, central para a forma como o texto é produzido e compreendido (MAHLBERG, 2006). Isso, geralmente, é dividido em coesão gramatical e lexical, delineadas na figura 6.

¹⁹ “Dizemos preferencialmente *strong tea* (relacionado à cor do chá) e *powerful car* (um carro potente), enquanto *powerful tea* e *strong car* não são frequentes. A combinação *powerful tea* até poderia ser empregada no sentido de se tratar de um chá poderoso, com poderes talvez curativos. Já o adjetivo *strong* combinado com *car* poderia se referir a uma característica física do *carro* (ORENHA-OTTAIANO, 2004, p. 25)

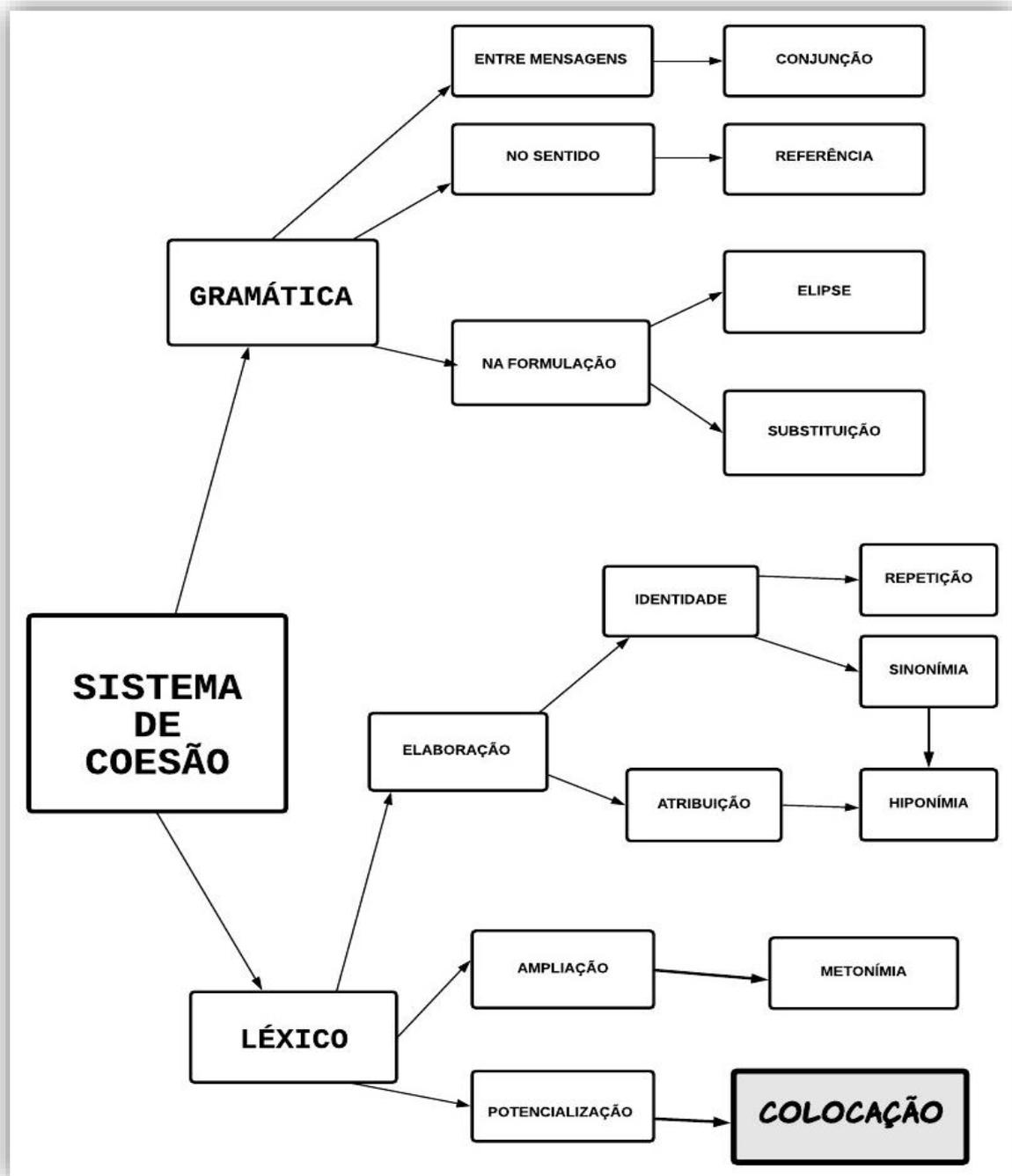


Figura 6 – *System of cohesion in English* (adaptado de HALLIDAY; HASAN, 1976; HALLIDAY, 2004)

Na figura 6, Halliday (2004) mostra-nos que, no item lexical do sistema de coesão, encontramos as colocações, intensificando a localização da colocação nesse sistema.

A colocação é, assim, vista como parte deste sistema de coesão para realçar as relações semânticas (HALLIDAY, 2004). Como salienta o autor, os mesmos itens podem se parecer em diferentes colocações de acordo com a variedade de texto. No entanto, uma definição precisa de *colocação* ainda permanece difícil de ser concebida (BARTSCH, 2004, p. 65). Salientamos

que tanto a figura 5 quanto a figura 6 revelam que as colocações estão no campo do léxico; por esse motivo, utilizamos tais imagens.

Definido a localização das colocações dentro do sistema de comunicação, partimos para a apreciação da caracterização de nosso objeto de estudo: as colocações da comunidade homossexual.

Para Houaiss (2001, p. 189), convenção é “qualquer dos usos ou costumes sociais estabelecidos, geralmente de tácita aceitação pelos indivíduos de uma comunidade, que incluem regras de boa educação, de boa conduta etc”. A partir da ideia de convencionalidade, Orenha-Ottaiano (2004), situa, dessa maneira, os fraseologismos dentro da convencionalidade da língua, conforme apresenta a figura 7 abaixo:

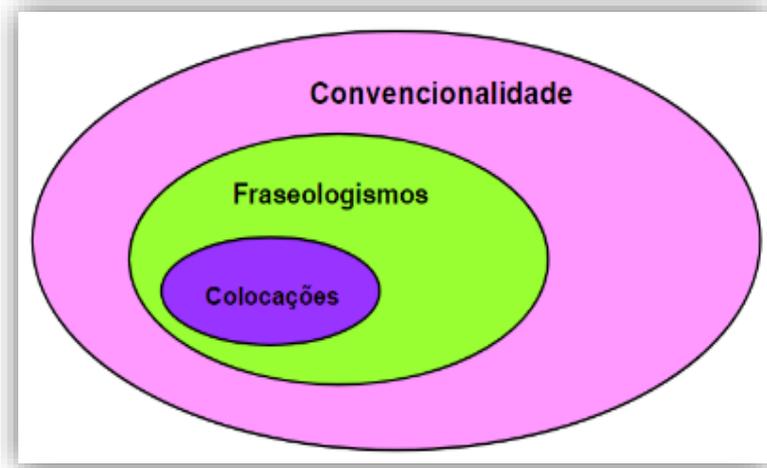


Figura 7 – Convencionalidade, Fraseologismos e Colocações (ORENHA-OTTAIANO, 2016)

A partir da proposta de Orenha-Ottaiano (2004; 2016), elaboramos uma imagem que demonstra de forma clara o nosso estudo, situando as colocações da comunidade *gay*, objeto de nossa pesquisa, nesta esfera:

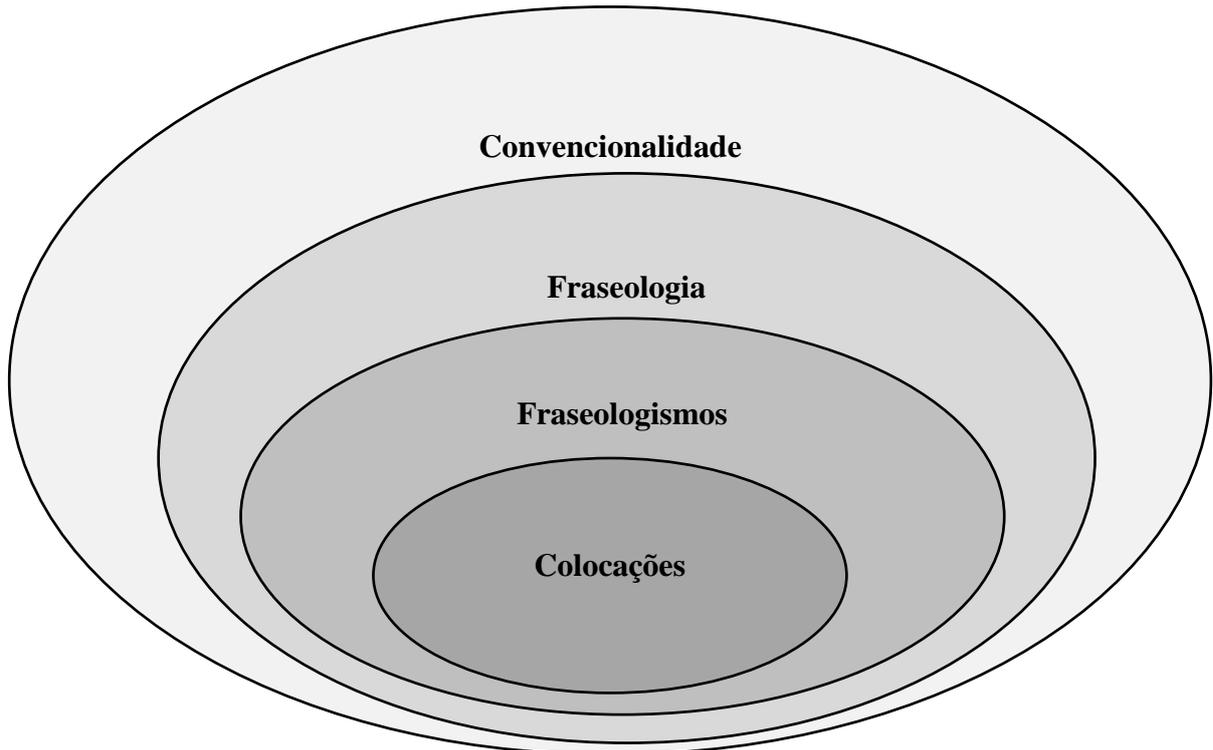


Figura 8 – Nosso objeto de estudo

Fonte: Elaborado pelo autor.

A figura 8, acima, demonstra nosso objeto de estudo de forma mais bem delimitada. Entendemos *Convencionalidade* como o “conjunto de elementos linguísticos, cuja coocorrência não é explicada sintática ou semanticamente, mas sim pelo uso” (FILLMORE, 1979, p. 56).

Fraseologia é entendida como uma área de pesquisa situada dentro da Linguística, voltada à investigação de combinações lexicais recorrentes, enquanto os *fraseologismos* são as “combinações não livres” (SAUSSURE, 1973, p. 77). E as *colocações* são as “combinações lexicais fixas, recorrentes e arbitrárias” (ORENHA-OTTAIANO, 2004, p. 20). Acreditamos que, desta forma, deixamos de forma explícita o foco desta pesquisa.

Há muitos anos, o termo *colocação* foi introduzido pelo linguista J. R. Firth (1957) para definir o fato de algumas palavras co-ocorrerem com certa frequência. Essa coocorrência é arbitrária, uma vez que não existe regra sintática ou justificativa semântica para tal uso. O autor define as colocações como “as palavras que mantêm a companhia, em suas relações com outras palavras” (1957). Outra definição apresentada pelo autor é “a forma como as palavras se combinam de forma previsível” (1957).

McIntosh et al (2009) definem uma colocação como “a maneira como as palavras se combinam em um idioma para produzir discurso e escrita natural²⁰”. No entanto, existem várias maneiras de defini-la e parece que linguistas ainda não concordaram com uma definição comum.

Como um meio de esclarecer o conceito, uma distinção geral que precisa ser feita é, nos termos de Sinclair (1991, p. 109), o princípio de livre escolha e o princípio idiomático. O autor explica que o princípio de livre escolha é a maneira usual de ver e descrever a linguagem, geralmente chamado de modelo *slot-and-filler*, a partir do qual quase qualquer palavra pode preencher os *slots*. Todavia, esse princípio não explica os numerosos constrangimentos na escolha do idioma. Por exemplo, não há razão para que não digamos *to put something on fire*, mas os falantes nativos ingleses concordariam que a colocação apropriada é *set something on fire*.

De acordo com Sinclair (1991), o princípio idiomático refere-se ao fato de um usuário de uma língua ter disponível para ele um grande número de frases semi-préconstruídas que constituem escolhas únicas. Este modelo de interpretação é ilustrado não apenas por colocações, mas também por outros tipos de expressões fixas, como expressões idiomáticas, provérbios, clichês, termos técnicos, expressões de jargão e verbos frasais (SINCLAIR, 1991, p. 111). A ideia desses princípios é que eles coexistem, mas não podem ser empregados simultaneamente.

Embora o termo colocação seja usado em diferentes sentidos por vários autores (HALLIDAY, 1966; SINCLAIR, 1991; STUBBS, 1995; MOON, 1998), dois pontos principais do conceito podem ser identificados: a *phraseological approach* (abordagem fraseológica) e a *frequency-based approach* (abordagem baseada na frequência) (NESSELHAUF, 2005, p. 12). De acordo com a abordagem baseada em frequência, uma colocação é citada como ocorrência de palavras a uma frequência maior do que o esperado se as palavras fossem combinadas arbitrariamente em uma linguagem. Segundo a abordagem fraseológica, as colocações são associações de dois ou mais lexemas (ou raízes) reconhecidas e definidas por sua ocorrência em uma gama específica de construções gramaticais (COWIE, 1994). Esta tradição enfatiza a relação entre padrões lexicais e sintáticos nas colocações e tem suas raízes no trabalho de Palmer (1933) e Hornby (1942). Essa visão tem sido fortemente influenciada pela fraseologia russa e alguns de seus apoiadores são Hausmann (1989), Cowie (1998) e Mel’čuk (1998).

²⁰ *The way words combine in a language to produce natural-sounding speech and writing* (MCINTOSH, 2009)

As colocações são consideradas pela abordagem fraseológica baseada em frequência como combinações de um ou mais *lexemes* ou *lemmas*. De acordo com a definição de Carter (1987, p. 6-7), os *lexemes* são as unidades básicas e contrastantes do vocabulário em uma língua e os *lemmas* são unidades lexicais pertencentes à mesma classe, diferindo na ortografia.

Ainda dentro da abordagem fraseológica, Cowie organiza combinações de palavras em quatro grupos concebidos em termos de transparência e comutabilidade, como podemos ver a seguir:

Combinações livres: a restrição de substituição pode ser especificada por motivos semânticos - todos os elementos da combinação de palavras são usados em um sentido literal.

Colocações restritas: algumas substituições são possíveis, mas existem limitações arbitrárias na substituição - pelo menos um elemento tem um significado não literal, e pelo menos um elemento é usado em seu sentido literal; Toda a combinação é transparente.

Modalidades figurativas: a substituição dos elementos raramente é possível - a combinação tem um significado figurativo, mas preserva uma interpretação literal atua.

Modalidades puras: a substituição dos elementos é impossível - a combinação tem um significado figurativo e não preserva uma interpretação literal atual.

Figura 9 – Combinações de palavras de Cowie (adaptado de COWIE, 1981, p. 224)

Os autores que adotam uma abordagem fraseológica costumam usar o termo colocação para se referir apenas a “colocações restritas” e outros termos para rotular os outros tipos de fraseologias (BENSON, 1997). Todavia, segundo Woolard (2000, p. 29), as colocações são “palavras que são estatisticamente muito mais propensas a aparecer juntas do que o acaso sugere”. Dessa maneira, aprender mais vocabulário não é apenas aprender novas palavras, mas também novas combinações de palavras. Para tanto, o estudo aprofundado das colocações em língua inglesa é necessário.

No caso desta pesquisa, o levantamento e a análise das colocações extraídas de nosso corpus paralelo de estudo, com o propósito de examinar o léxico fraseológico da comunidade homossexual de maior frequência, em relação ao léxico fraseológico da comunidade

homossexual nas traduções das transcrições para o português do Brasil, com o propósito de elaboração de um *Glossário Bilíngue de Colocações da Comunidade Homossexual*.

As colocações são encontradas em blocos, formados pela estrutura base + colocado(s). Caso um de seus componentes seja alterado, no sentido de escolhermos outra base ou colocado, poderá ocorrer interrupção na comunicação se o colocado selecionado não for aquele usualmente utilizado com a base. Ou seja, a escolha passa a não mais ser a convencional, aceita pela comunidade, como observa Tagnin (2005).

Lewis (2000, p. 132) classifica as colocações como “a maneira em que as palavras coocorrem no texto natural em formas estatisticamente significativas”. Nesta definição, destacamos: a combinação previsível é sobre a forma como as palavras co-ocorrem naturalmente, que, a nosso ver, é a língua tal qual é usada.

Segundo Tagnin (2005), não existe uma explicação para o fato de certas palavras combinarem-se com tal naturalidade. Só o uso pode consagrar a maneira que são utilizadas na língua. De acordo com Lewis (2000), alguns tipos de combinações são parte da língua materna do falante nativo que os usam de forma fácil e inconsciente.

Wray (2002, p. 9) define as colocações como segue:

[...] sequência, contínua ou descontínua, de palavras ou outros elementos, que é, ou parece ser, pré-fabricados: isto é, armazenados e recuperados inteiros da memória no momento da utilização, em vez de serem submetidos à geração ou análise pela gramática da língua (WRAY, 2002, p. 9)²¹.

A ideia principal na explicação de Wray (2002) é que a língua estereotipada consiste em unidades de língua ou expressões que são armazenadas e recuperadas como conjuntos, em vez de geradas pelas regras da gramática. Enquanto ela dá importância à totalidade das palavras concorrendo em termos de armazenamento e recuperação, Sinclair (1991, p. 170) as define como “a ocorrência de duas ou mais palavras, dentro de um curto espaço de uns aos outros em um texto”.

Wray afirma que “algumas palavras parecem ter sido feitas uma para a outra numa expressão, enquanto outras, que deveriam ser igualmente adequadas, soam estranhas” (2002).

Identificá-las, assim como qualquer outro fraseologismo, requer o reconhecimento de alguns traços. Nesse sentido, tratando especialmente das colocações, Tagnin (1999) elencou alguns traços que auxiliam na qualificação de uma colocação: (i) *recorrência*: atestada pela

²¹[...] *sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar* (WRAY, 2002, p. 9).

frequência da combinação; (ii) *não idiomaticidade*: o significado da colocação é composicional; (iii) *coesão*: há uma forte atração entre os constituintes da colocação; (iv) *restrição contextual*: há um contexto específico que favorece a ocorrência da combinação; e (v) *coocorrência arbitrária entre os elementos*: não há explicação semântica que justifique a colocação.

Tagnin (2013, p. 151-152) distingue os termos *colocação* e *coligação*. Para a autora, as colocações são combinações lexicais consagradas de duas ou mais palavras de conteúdo, como *repolho roxo*, *mundos e fundos*, *aviar uma receita*, por exemplo. Por outro lado, as coligações são combinações consagradas de elementos linguísticos em que uma estrutura ou categoria gramatical se combina com a palavra base (*obedecer a*, *cumpridor de*).

Hausmann (1984), elabora a classificação das combinações de palavras de acordo com as características fixas vs. não fixas, como veremos a seguir:

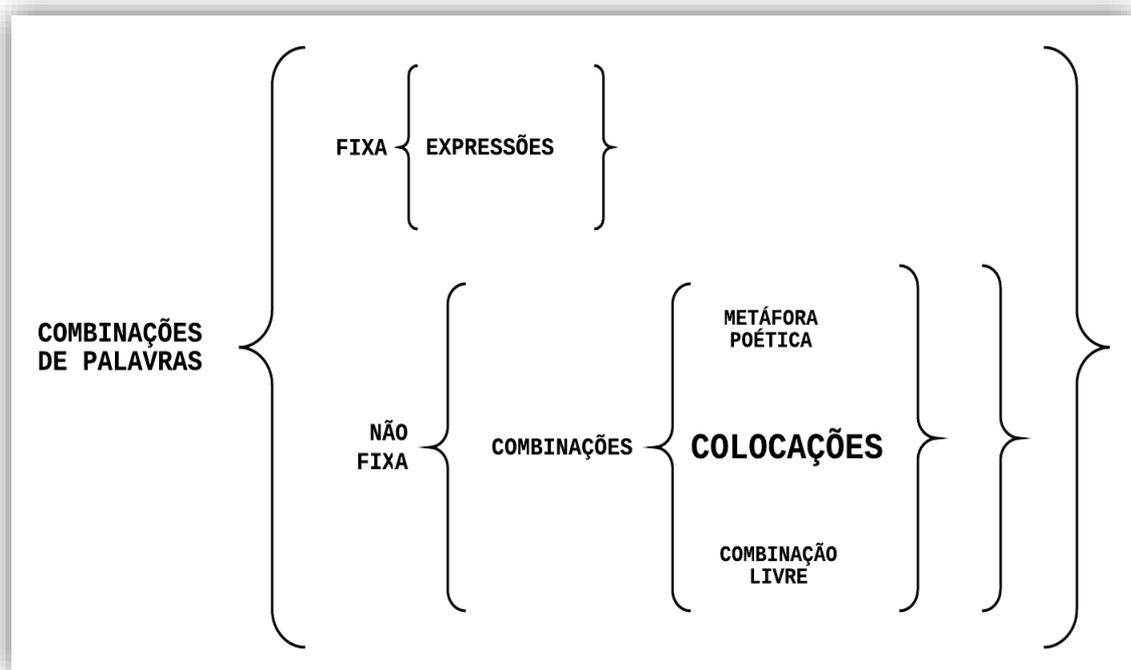


Figura 10 – Hausmann's classification of word-combinations (adaptado de HAUSMANN, 1984, p. 399)

A tipologia de Hausmann (1984) depende essencialmente da distinção entre expressões fixas e não fixas. O primeiro item compreende as expressões idiomáticas; no segundo, temos três tipos de combinações não fixas subdivididas em três itens, um deles as colocações, consideradas combinações lexicais.

As combinações de palavras binárias consistem em palavras com capacidade combinatória limitada, são produtos semiacabados da linguagem, combinações afins de

habitabilidade marcante. Para o autor, as combinações de palavras consistem em palavras com capacidade combinatória limitada, sendo assim, produtos semiacabados da linguagem (HAUSMANN, 1989).

Segundo Hausmann (1997, p. 65), as colocações são formadas pela contribuição de uma palavra autosemântica (base) com uma sinsemântica (colocada), uma vez que a escolha da palavra sinsemântica está restringida pela palavra autosemântica. Para a classificação das colocações encontradas em nossos *corpora*, utilizaremos a tipologia apresentada por Orenha-Ottaiano (2004, 2009) à luz da tipologia proposta por Hausmann (1985). Veremos, a seguir, os tipos de colocação propostos:

Verbais – com quatro formas básicas:

- Verbo (colocado) + Substantivo (base)
- Substantivo (base) + Verbo (colocado)
- Verbo (colocado) + Preposição + Substantivo (base)
- Verbo (colocado) + Partícula Adverbial + Substantivo (base)
- Verbo (colocado) + Adjetivo (base)

Nominais - com duas formas básicas:

- Substantivo (base) + Substantivo (colocado)
- Substantivo (colocado) + Preposição + Substantivo (base)

Adjetivas – com uma forma:

- Adjetivo (colocado) + Substantivo (base)

Adverbiais – com três formas básicas:

- Advérbio (colocado) + Adjetivo (base)
- Verbo (base) + Advérbio (colocado)
- Advérbio (colocado) + Verbo (base)

Os componentes de uma colocação são ordenados hierarquicamente (HAUSMANN, 1984):

- a base é semanticamente autônoma;
- a colocação precisa da base para obter o seu significado completo.

Cowie (2001) distribui as combinações de palavras em duas categorias: *combinações semânticas* e *combinações pragmáticas* ou *fórmulas*. As combinações semânticas são ainda distribuídas em *combinações livres*, *colocações* e *expressões idiomáticas*, enquanto as combinações pragmáticas são divididas em *provérbios* e *fórmulas de rotina*.

De acordo com essa classificação de combinações semânticas, as combinações livres não são pré-construídas, elas são usadas em um sentido literal e sem restrições em combinações, como *drink tea* (*beba chá*), por exemplo. A substituição de ambos os elementos da combinação livre por outro elemento pode ser feita com base na restrição semântica, como veremos, na figura 11, segundo a proposta de Cowie (2001):

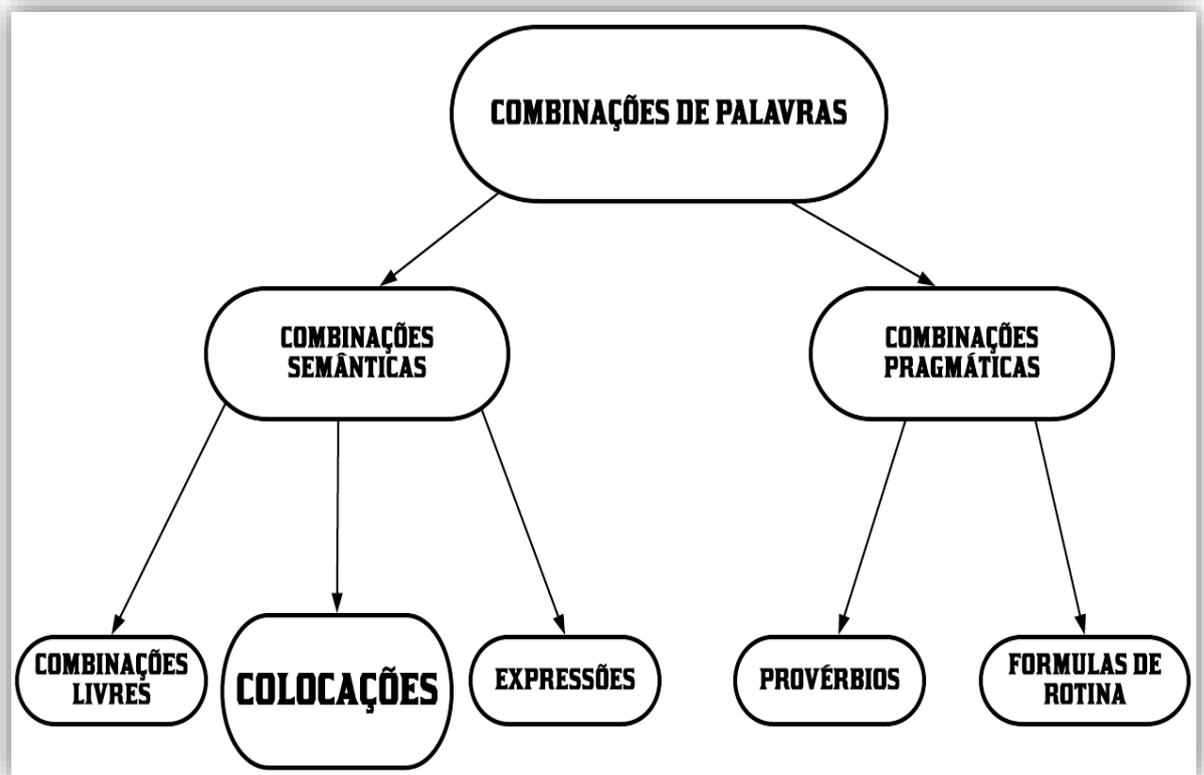


Figura 11 – Cowie's distribution of word combinations (adaptado de COWIE, 2001)

As colocações são consideradas pré-construídas; pelo menos um elemento da colocação é semanticamente literal, com alguma possibilidade de substituição de elementos, mas com base em limitações arbitrárias e não semânticas. As expressões idiomáticas são combinações de palavras pré-construídas com significado não literal, ou seja, o significado não pode ser deduzido do significado individual das palavras e com pouca ou nenhuma possibilidade de substituição de elementos (NESSELHAUF, 2005).

Pariforme, Benson, Benson e Ilson (1986) elucidam as colocações como combinações vagamente fixas e semanticamente transparentes, em algum lugar entre expressões idiomáticas e combinações livres.

Para Leśniewska (2006), algumas combinações de palavras são processadas como partes integrantes. Segundo o autor, as combinações fixas de múltiplas palavras envolvem vínculos muito fortes entre as palavras, enquanto as colocações utilizam elos de força variável, ou seja, as ligações entre certas palavras são mais difíceis e sutis do que entre outras.

Os falantes nativos percebem as colocações como lexemas individuais, enquanto os falantes estrangeiros devem reconhecê-los como sintagmas vinculados, de modo a não traduzirem cada elemento individualmente. Devido a isso, as colocações são mais propensas a influências do que outras combinações de palavras, embora haja relativamente poucos dados sobre sua extensão (LEŚNIEWSKA, 2006).

Mediante a exposição das definições de alguns autores e das principais características das colocações, podemos reafirmar que elas são empregadas pelos falantes nas mais diversas situações da vida cotidiana, lançando mão do léxico geral compartilhado entre si. Na sequência, apresentaremos o item intitulado *Linguística de Corpus: definição e tipologia de corpus*.

2.3 Linguística de *Corpus*: definição e tipologia de *corpus*

Ao considerar que a Linguística é a área do saber voltada ao estudo científico da linguagem humana, podemos inferir que a Linguística de *Corpus* (LC) trabalha na produção de conhecimento sobre a língua mediante o uso de investigação de *corpora* analisados por meio de métodos de pesquisa inerentes. Em pesquisas no campo da linguagem, a utilização da palavra *corpus* é comum, por se tratar de um conjunto de dados examinados pelo pesquisador. Entretanto, não consideramos qualquer tipo de análise de dados linguísticos como sendo um trabalho da área de LC, portanto precisamos delimitar tal abrangência.

Tratando-se da análise de uma grande quantidade de dados, é indispensável destacarmos que se torna de grande eficácia a utilização de um aparato teórico-metodológico que permita a extração de unidades que são objeto desta pesquisa, uma vez que o manuseio de dados tornar-se-ia inviável sem a utilização de ferramentas para extração. Nesse cenário, nossa pesquisa está inserida na área denominada de LC.

Ao contrário de disciplinas linguísticas como morfologia e sintaxe, por exemplo, que investigam morfemas e construções sintáticas, ou seja, objetos de estudo específicos, a LC não opera em um nível linguístico específico. McEnery e Hardie (2012, p. 1) consideram a LC como

uma área que concentra um conjunto de procedimentos, ou métodos, para o estudo da língua. No entanto, reconhecem que existem pesquisadores que não consideram a área como uma simples metodologia.

Referem-se a um novo olhar a respeito dos estudos da língua, de uma base metodológica que permite a pesquisa linguística mediante uma variedade de procedimentos e métodos capazes de redefinir e reorganizar outras teorias linguísticas (MCENERY; HARDIE, 2012; ROCHA, 2017). Os que julgam que a LC é um mero método de pesquisa, enganam-se; a LC transcende essa visão restrita, por abranger os usos que um *corpus* pode oferecer para o entendimento dos fenômenos da linguagem (KENNEDY, 1998). Além disso, trata-se de uma abordagem que pode ser utilizada para investigações empíricas em qualquer área de análise linguística (BIBER; CONRAD; REPPEN, 1998; ROCHA, 2017).

Independentemente da posição adotada, a LC faz objeção a teorias centrais no campo da Linguística. Inserida em uma abordagem fundamentalmente empírica dos estudos da linguagem, a LC vai na contramão da perspectiva racionalista proposta pelo Gerativismo (BERBER SARDINHA, 2004, p. 30). Na perspectiva empírica, a língua é vista como um sistema probabilístico em desvantagem de um sistema “possibilístico”. De outro modo, a probabilidade de ocorrência de certo dado linguístico é mais significativa do que a possibilidade, já que nem tudo o que é possível dizer na língua é usual de fato.

Em meio à dicotomia empirismo *versus* racionalismo, Halliday (1991) chama-nos à atenção para a probabilidade de ocorrer determinados fenômenos linguísticos em contextos específicos. Para o autor, pensar a língua como sistema probabilístico significa reconhecer a regularidade e a padronização de suas estruturas. Sendo assim, a verificação empírica dos traços linguísticos torna-se inevitável, já que, embora possíveis por meio de introspecção, os fatos da língua nem sempre ocorrem com a mesma frequência e não são aleatórios.

Esse é um aspecto fundamental que confronta metodologias racionalistas e empíricas nos estudos da linguagem. No lado racionalista, encontramos a introspecção para confirmar a possibilidade de um enunciado linguístico ocorrer, ou seja, o próprio falante julga a adequação de um fenômeno. Já no lado empírico, é necessário que os dados comprovem eventuais afirmações do pesquisador.

Sinclair (1996) discorre a respeito da padronização das estruturas linguísticas:

O estudo de corpus permite-nos observar os usos que as pessoas fazem da língua, reuni-los, e olhá-los cientificamente. Quando se faz isso, começa-se a notar que há uma grande quantidade de coocorrência de palavras que não são aleatórias, mas claramente regulares e frequentes. De fato, a ideia de qualquer

tipo de aleatoriedade, de qualquer tipo de ocorrência determinada pelo acaso é bastante remota na língua. Tudo é altamente determinado, altamente planejado e altamente organizado (SINCLAIR, 1996, p. 99)²².

Este equilíbrio na organização da língua leva Sinclair a afirmar que ela é “uma grande coordenação de escolhas, cuja frequência e regularidade de algumas faz com que adquiram um significado único” (SINCLAIR, 1996, p. 99).

Os dois tipos de abordagem, racionalista e empírica, são válidas e podem produzir resultados pertinentes, entretanto, a tarefa é mais difícil quando o pesquisador estuda uma língua que não domina. Nesse sentido, o recurso a um *corpus* pode ser mais oportuno do que a simples introspecção.

Todavia, com o progresso da LC, várias definições para o termo *corpus* têm sido propostas. Segundo Tagnin (2013, p. 29), um *corpus* é “uma coletânea de textos, necessariamente em formato eletrônico, compilados e organizados segundo critérios ditados pelo objetivo de pesquisa a que se destina”. A respeito de alguns elementos empregados para definir o termo, algumas diferenças são observadas nas propostas de autores distintos.

Sinclair (1991, p. 171) define *corpus* como:

Um *corpus* é uma coletânea de textos que ocorrem naturalmente, escolhidos para caracterizar um estado ou uma variedade de uma língua. Na linguística computacional moderna, um *corpus* tipicamente contém muitos milhões de palavras: isto é porque se reconhece que a criatividade da linguagem natural leva a uma imensa variedade de expressão que é difícil isolar padrões recorrentes que são indícios da estrutura lexical da língua (SINCLAIR, 1991, p. 171)²³.

Nessa perspectiva, Sinclair (1991) leva-nos à discussão dos conceitos de *coletânea* e *caracterização*. A ambiguidade dos documentos é substituída por textos e a função do *corpus* como representativo de uma língua é acrescentada. Além disso, ainda que mencione a questão do tamanho, o autor opta por não a incluir na definição de *corpus*, mas na sua caracterização específica.

Para Tognini-Bonelli (2001), um *corpus* pode ser delimitado como:

²²The study of corpus texts allow us to stand back from people' usage, gather it together, and look at it in a scientific fashion. When you do this, you begin to notice that there is a great deal of co-occurrence of words that are not all random but, quite clearly, regular and frequent. In fact, the idea of any kind of randomness, any kind of chance occurrence in Language is very remote indeed. It is all highly determined, highly planned, highly organized (SINCLAIR, 1996, p.99). [Todas as traduções são de nossa responsabilidade.]

²³A collection of naturally-occurring language text, chosen to characterize a state or variety of a language. In modern computational linguistics, a corpus typically contains many millions of words: this is because it is recognized that the creativity of natural language leads to such immense variety of expression that it is difficult to isolate the recurrent patterns that are the clues to the lexical structure of the Language (SINCLAIR, 1991, p.171).

Uma coleção de textos que se assume ser representativa de uma dada língua, compilada de forma que possa ser usada para análise linguística. Usualmente a pressuposição é que a língua armazenada em um *corpus* ocorre de forma natural, que ela é colhida de acordo com critérios de planejamento explícitos, com um propósito específico em mente, e com uma reivindicação de representar padrões maiores da língua selecionados de acordo com uma tipologia específica. Nem todo mundo, evidentemente, aceita estas pressuposições, mas em geral há um consenso que um *corpus* contempla língua natural, autêntica (TOGNINI-BONELLI, 2001, p. 2)²⁴.

Notamos na definição acima a necessidade de haver um propósito de pesquisa bem preciso para a compilação do *corpus*. Este propósito deve embasar a compilação e deve iluminar os critérios de seleção acerca do que incluir no *corpus*. Dependendo do propósito da pesquisa, vários tipos de *corpora* podem ser empregados.

Hunston (2002, p. 2), diferentemente das definições anteriormente revisadas, detalha melhor o conceito:

Um *corpus* é definido em termos tanto de sua forma como de seu propósito. Linguistas têm sempre usado a palavra *corpus* para descrever uma coleção de exemplos naturais da linguagem, consistindo de qualquer coisa desde algumas frases até um conjunto de textos escritos ou gravações de fitas que foram coletadas para um estudo linguístico. Mais recentemente, a palavra tem sido reservada para coleções de textos (ou partes de texto) que são armazenadas e acessadas eletronicamente. Porque os computadores podem armazenar e processar grandes quantidades de informação, os *corpora* eletrônicos são usualmente maiores do que as coleções pequenas e baseadas em papel previamente utilizadas para estudar aspectos da linguagem. (HUNSTON, 2002, p. 2)²⁵.

Uma nova definição de *corpus* é elaborada por Sinclair (2005), após 14 anos da publicação de *Corpus, concordance and collocation* (SINCLAIR, 1991): “um *corpus* é uma coleção de porções de textos em formato eletrônico, selecionados de acordo com critérios externos para representar, até onde for possível, uma língua ou uma variedade da língua como

²⁴A collection of texts assumed to be representative of a given Language put together so it can be used for linguistic analysis. Usually that assumption is that the Language stored in a corpus is naturally-occurring, that it is gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology. Not everybody, of course, goes along with these assumptions but in general there is consensus that a corpus deals with natural, authentic language. (TOGNINI-BONELLI, 2001, p. 02).

²⁵A corpus is defined in terms of both its form and its purpose. Linguists have always used the word corpus to describe a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study. More recently, the word has been reserved for collections of texts (or parts of text) that are stored and accessed electronically. Because computers can hold and process large amounts of information, electronic corpora are usually larger than the small, paper-based collections previously used to study aspects of language. (HUNSTON, 2002, p.2)

uma fonte de dados para a análise linguística²⁶”. Quando essa definição é comparada à de 1991, várias diferenças podem ser notadas.

O *corpus* passa a ser, para Sinclair (2005), eletrônico, compilado com base em critérios externos, construído com o objetivo de pesquisa linguística e podendo ser representativo de uma língua. Enquanto em 1991, apesar de incluir a ideia de porções de textos na definição de *corpus*, Sinclair argumenta que “as amostras de linguagem para um *corpus* devem, sempre que possível, consistir de documentos inteiros ou transcrições de eventos de fala completos, ou devem chegar o mais próximo possível deste alvo” (SINCLAIR, 2005, p. 04).

Acrescentamos, também, a visão de Hunston (2002, p. 2):

Um *corpus* é planejado apesar de o acaso poder desempenhar um papel na coleção de textos, e ele é projetado para algum propósito linguístico. O propósito específico do planejamento determina a seleção de textos, e o objetivo é outro que não a preservação de textos propriamente ditos por causa de seus valores intrínsecos. Isto diferencia um *corpus* de uma biblioteca ou arquivo eletrônico. O *corpus* é armazenado de tal forma que pode ser estudado de forma não-linear, e tanto quantitativa como qualitativamente. O propósito não é simplesmente acessar os textos de forma a lê-los, o que novamente distingue o *corpus* de uma biblioteca ou arquivo (HUNSTON, 2002, p. 2)²⁷.

Como podemos observar no fragmento acima, é extremamente importante ter definido o propósito do *corpus* a ser compilado, selecionando os textos de forma criteriosa, a fim de garantir a análise do *corpus* de forma qualitativa e quantitativa. Para nosso trabalho, compilamos um *corpus* paralelo, utilizando as transcrições da série *Queer as Folk* e, para nosso *corpus* comparável, utilizamos a ferramenta *BootCat* que faz uma varredura na *web*, a fim de compilar o *corpus*.

De acordo com McEnery e Wilson (2001), qualquer coleção de mais de um texto pode ser chamada de *corpus*: o termo *corpus* vem do latim *corpo*; portanto, um *corpus* pode ser definido como qualquer texto. Embora simples, essa definição não é suficientemente abrangente, pois falta um significado adicional que o termo *corpus* carrega na linguística

²⁶ *A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as source of data for linguistic research* (SINCLAIR, 1991).

²⁷ *A corpus is planned, though chance may play a part in the text collection, and it is designed for some linguistic purpose. The specific purpose of the design determines the selection of texts, and the aim is other than to preserve the texts themselves because they have intrinsic value. This differentiates a corpus from a library or an electronic archive. The corpus is stored in such a way that it can be studied non-linearly, and both quantitatively and qualitatively. The purpose is not simply to access the texts in order to read them, which again distinguishes the corpus from the library and the archive* (HUNSTON, 2002, p. 2)

moderna e, mais especificamente, na LC. Por esse motivo, os autores fornecem quatro características adicionais de *corpus* (MCENERY; WILSON, 2001, p. 29):

→ **Amostragem e representatividade:** não é possível coletar todos os textos de uma língua. A população de texto para idiomas como o inglês é enorme e novos enunciados são criados todos os dias. Por esse motivo, os *corpora* baseiam-se na amostragem. Além disso, ao estudar a variedade de uma língua, os *corpora* precisam ser *altamente representativos* dessa variedade, a fim de fornecer uma imagem o mais precisa possível e evitar a distorção.

→ **Tamanho finito:** o *corpus* tende a ter um tamanho finito de palavras. Assim que um *corpus* atinge o objetivo da pesquisa, a coleta para o *corpus* não aumenta em tamanho mais.

→ **Legível por máquina:** hoje em dia, é dar por certo que os *corpora* são “legíveis por máquina”, isto é, eles existem em um formato eletrônico. Esses *corpora* possuem as seguintes vantagens: podem ser pesquisados e manipulados de maneiras que não são possíveis para *corpora* em outros formatos.

→ **Referência padrão:** existe uma compreensão implícita de que um *corpus* funciona como uma referência padrão para a variedade de linguagem que representa.

Observamos, acima, quatro características relevantes durante a compilação de um *corpus* de estudo; tais características foram levadas em consideração durante a compilação dos *corpora* utilizados no presente estudo.

Um *corpus* é sempre projetado com um propósito específico em mente, o qual, por sua vez, caracteriza o próprio *corpus*. O que se segue é uma lista de alguns dos tipos comuns de *corpus* (HUNSTON, 2002, p. 14):

→ **Corpus especializado:** um *corpus* de textos de um tipo específico, como por exemplo, editoriais de jornais, artigos científicos e assim por diante. Ele é usado para investigar um tipo particular de linguagem.

→ **Corpus geral:** um *corpus* de textos de vários tipos. Ele pode incluir material escrito, falado ou ambos e tende a incluir uma variedade de textos o mais amplo possível. Os *corpora* gerais normalmente são muito maiores do que os especializados e são comumente usados como fontes de referência para estudos gerais de linguagem.

→ **Corpus comparável:** compostos por textos originalmente escritos em uma ou mais línguas e compilados de acordo com os mesmos critérios. Eles são usados principalmente para identificar semelhanças ou diferenças entre as línguas ou variedades comparadas.

→ **Corpus paralelo:** compostos por textos originais e suas traduções em uma ou mais línguas. Realçamos que utilizamos um *corpus* paralelo para o desenvolvimento desta pesquisa, composto originalmente em língua inglesa e suas respectivas traduções em português do Brasil.

→ **Corpus de aprendiz:** um *corpus* desse tipo consiste em uma coleção de textos produzidos por alunos de uma língua. O objetivo de tal *corpus* é identificar diferenças entre os próprios alunos e entre alunos e falantes nativos do idioma.

→ **Corpus monitor:** um *corpus* projetado para rastrear mudanças em um idioma. Conforme mencionado anteriormente, os *corpora* monitores aumentam constantemente de tamanho à medida que novos textos são adicionados.

Enfatizamos que trabalhamos com dois tipos de *corpora* nesta pesquisa: um *corpus* paralelo, composto de seus originais e suas respectivas traduções, e um *corpus* comparável, formado por textos coletados pela ferramenta *Sketch Engine* (KILGARRIFF et al, 2004), por meio de uma varredura na *web*, com base nas *seeds* (palavras de busca), para compilar um *corpus* exclusivo para compararmos a utilização e frequência das colocações levantadas.

O desfecho de uma pesquisa depende essencialmente do conteúdo e da tipologia do *corpus* de estudo. A literatura da área distingue vários tipos de *corpora* e, vemos a seguir, a tipologia proposta por Leech (1991) e Berber Sardinha (2004)

Modo:

→ Falado: composto por textos transcritos de fala;

→ Escrito: composto por textos escritos.

Tempo:

→ Sincrônico: compreende um período no tempo;

→ Diacrônico: compreende vários períodos no tempo;

→ Contemporâneo: composto por textos de um período de tempo corrente;

→ Histórico: composto por textos de um período de tempo passado.

Seleção:

- Amostragem: representa uma amostra finita da linguagem; é composto por porções de textos ou variedades textuais;
- Monitor: “a composição é [atualizada] para refletir o estado atual de uma língua. Opõe-se a *corpora* de amostragem” (BERBER SARDINHA, 2004, p. 20);
- Dinâmico e orgânico: caracteriza o *corpus* monitor, sofre acréscimos e decréscimos;
- Estático: caracteriza o *corpus* de amostragem, é oposto ao dinâmico; não sofre acréscimo nem decréscimo;
- Equilibrado: os gêneros textuais, tipos textuais, fontes e número de palavras dos textos são distribuídos uniformemente.

Conteúdo:

- Especializado: composto por textos de tipos específicos;
- Regional ou dialetal: textos de uma ou mais variedades sociolinguísticas específicas;
- Multilíngue: textos de idiomas diferentes.

Tamanho:

Assim como outros aspectos da LC, discussões teóricas sobre o tamanho ideal de um *corpus* tendem a ser controversas. Há quem defenda que quanto maior o *corpus*, melhores serão os resultados. Entretanto, existe uma tendência cada vez maior para a compilação de mini *corpora* especializados para a extração terminológica e produção de glossários (MAIA, 2000).

Esses *corpora* são formados por textos de uma área de especialidade em que o léxico não apresenta muita variação, como as áreas técnicas da cardiologia e da dermatologia.

Leech (1991) diz que “o tamanho não é tão importante²⁸”, sendo assim, podemos afirmar que nossos *corpora* de estudo apresentaram-se suficientes para o propósito deste estudo. Nosso *corpus* paralelo é considerado adequado, possuindo um *subcorpus* em inglês com mais de 5 milhões de palavras, enquanto o *subcorpus* com as traduções possui 4,5 milhões de palavras. E nosso *corpus* comparável foi suficiente para o desenvolvimento desta pesquisa.

²⁸ *Size is not all-important* (LEECH, 1991).

Conteúdo:

O conteúdo de um *corpus* é definido com base nos resultados que uma pesquisa pretende alcançar. Portanto, nenhum *corpus* deve ser considerado ruim ou bom em sua essência. O que se deve levar em consideração é o estudo linguístico que se almeja realizar (MATUDA, 2011).

Outro aspecto a ser considerado durante a compilação do *corpus* é a autenticidade dos textos. Se a LC representa uma visão probabilística da linguagem que tem raízes na linguística hallidayana, teoria que descreve a probabilidade dos sistemas linguísticos, dados os contextos em que os falantes os empregam (BERBER SARDINHA, 2004, p. 30), os textos que fazem parte de um *corpus* devem ser autênticos, ou seja, representações reais da língua que se pretende estudar e não textos produzidos com vistas à análise linguística.

Em nossa pesquisa, trabalhamos com representações reais da linguagem *gay* contidas na Série *Queer as Folk*, que retrata a realidade linguística da comunidade homossexual para a composição do nosso *corpus* paralelo.

Partindo do descrito acima, na tipologia, nosso *corpus* paralelo é composto pelas legendas em inglês e pelas traduções das respectivas legendas em português do Brasil. O tempo classifica-se como contemporâneo, já que realizamos uma análise comparativa, por meio do *corpus* comparável e a seleção se dá por amostragem.

Já o *corpus* comparável é composto por textos originalmente escritos em inglês e em português do Brasil e, dividido em dois *subcorpora*: um *subcorpus* em inglês e um *subcorpus* em português do Brasil, os quais são escritos e contemporâneos, uma vez que foram compilados exclusivamente para o desenvolvimento desta pesquisa.

Com base nas considerações abordadas acima, apresentamos no quadro 4, abaixo, as características dos *corpora* de estudo paralelo e comparável desta pesquisa; fundamentamo-nos na proposta de Leech (1991), Hunston (2002) e Berber Sardinha (2004):

Tipologia do <i>Corpus</i> segundo Leech (1991), Hunston (2002) e Beber Sardinha (2004)		Características de nosso <i>corpus</i> paralelo	Características de nosso <i>corpus</i> comparável
Modo	Textos falados e/ou escritos	Textos escritos	Textos escritos
Tempo	Textos sincrônicos, diacrônicos, contemporâneos e históricos	Textos sincrônicos e contemporâneos	Textos sincrônicos, diacrônicos e contemporâneos
Seleção	<i>Corpus</i> de amostragem, monitor, dinâmico, estático e equilibrado	<i>Corpus</i> de amostragem, estático e equilibrado	<i>Corpus</i> de amostragem, estático e equilibrado
Conteúdo	Textos especializados, regionais ou dialetais e multilíngues	Textos dialetais e multilíngues	Textos dialetais e multilíngues
Autoria	Textos de aprendizes e de língua nativa	Textos de língua nativa	Textos de aprendizes e de língua nativa
Disposição Interna	<i>Corpus</i> paralelo, alinhado e comparável	<i>Corpus</i> paralelo	<i>Corpus</i> comparável
Finalidade	<i>Corpus</i> de estudo, referência e de treinamento	<i>Corpus</i> de estudo	<i>Corpus</i> de estudo

Quadro 4 – Tipologia do *corpus* de estudo e características de nosso *corpus* de estudo

Fonte: Elaborado pelo autor.

Como podemos observar, nosso *corpus* é caracterizado pelo uso de *textos escritos*, uma vez que identificamos e extraímos as colocações da comunidade homossexual a partir da série *Queer as Folk*. No que diz respeito ao *tempo*, utilizamos apenas textos sincrônicos e contemporâneos, pois objetivamos analisar um aspecto específico da língua.

Quanto à *seleção*, nosso *corpus* de estudo apresenta uma amostra finita da linguagem da comunidade homossexual. Devido ao fato de estudarmos colocações da língua geral, fazemos uso de textos de *conteúdo* dialetais e multilíngues. No tocante à *autoria*, notamos que nosso *corpus* é composto por textos de língua nativa, neste caso a língua inglesa e as respectivas traduções em português.

Em relação à disposição interna do *corpus* de estudo, tratamos de um *corpus paralelo*, por meio do qual extraímos as colocações da comunidade homossexual e realizamos a comparação da frequência e do uso, por meio dos *corpora* comparáveis, que foram compilados pela ferramenta *Sketch Engine* (KILGARRIFF et al, 2004), exclusivamente para este estudo.

No que tange à *finalidade*, analisamos nosso *corpus* de estudo, contrastando-o com o *corpus* comparável. A partir disso, analisamos quais itens lexicais possuem chaticidade significativa, a fim de constituírem colocações da comunidade homossexual.

Pudemos observar no quadro 4 que nosso *corpus* comparável é muito parecido com nosso *corpus* paralelo, exceto em relação ao *tempo*, que também é considerado *diacrônico* em relação à *autoria*, uma vez que pode ser de aprendizes, e em relação à *disposição interna*, nossa maior diferença observada, pois trata-se de um *corpus* comparável.

Após a revisão teórica que embasou o presente estudo, discorreremos sobre a metodologia empregada para o desenvolvimento desta pesquisa. Seguiremos com o Capítulo 3, *Metodologia de pesquisa*.

3 METODOLOGIA DE PESQUISA

Para o levantamento e a análise das colocações, compilamos um *corpus* paralelo composto por dois *subcorpora*: um *subcorpus* em inglês e um *subcorpus* em português, ambos constituídos pelas transcrições das legendas em inglês e em português das cinco temporadas da série *Queer as Folk*, que retrata as dificuldades e as conquistas de cinco homens homossexuais.

Além desse *corpus* paralelo, observamos a necessidade da compilação de um *corpus* comparável, composto de dois *subcorpora*: um *subcorpus* em inglês e outro *subcorpus* em português do Brasil, a fim de certificarmos a coocorrência das colocações levantadas de nosso *corpus* paralelo. Tal compilação objetiva a verificar a frequência e o uso das colocações, já que o *corpus* paralelo foi compilado por meio das legendas do seriado *Queer as Folk*, retiradas do site de *download* de legendas (*tvsubtitle.net*). Evidenciamos, também, que essa compilação se fez necessária, a fim de confirmarmos se as colocações extraídas de nosso estudo são frequentemente empregadas pela comunidade homossexual. Desse modo, evitamos questionamentos no sentido de que as colocações extraídas do *corpus* paralelo da série *Queer as Folk* poderiam ser específicas daquele contexto. Por meio do *corpus* comparável, foi-nos possível atestar que as colocações inseridas no Glossário são usuais e frequentemente utilizadas.

Seguindo a tipologia de *corpus* sugerida por Berber Sardinha (2004), para a compilação dos *corpora* de estudo, serão considerados alguns aspectos importantes, como a autenticidade dos textos, o propósito de pesquisa linguística, o manuseio por computador e a representatividade de uma dada variedade.

Conforme mencionamos, para Berber Sardinha (2004), o *corpus* é um objeto criado com fins específicos de pesquisa e de acordo com a sua extensão e o propósito da pesquisa. O *corpus* paralelo desta pesquisa é considerado por nós suficiente para o desenvolvimento deste estudo, pois é composto de 81 episódios, totalizando 5.177.124 palavras para o *subcorpus* em inglês, e 4.556.915 palavras para o *subcorpus* paralelo em português do Brasil. Enquanto o *corpus* comparável é composto por 1.174.067 em seu *subcorpus* comparável em inglês e 1.000.598 em seu *subcorpus* comparável em português. Embora seja menor em relação ao *corpus* paralelo, atendeu aos propósitos desta pesquisa.

Por meio do *software WordSmith Tools* (SCOTT, 2012), versão 4.0, foram salvas as transcrições dos episódios do seriado em formato *txt*, permitindo o manuseio dos *corpora* pelo referido programa e possibilitando, assim, a realização das análises baseadas na frequência, no

uso e na co-ocorrência das palavras em nosso *corpus* paralelo. Dessa forma, realizamos o levantamento das colocações mais frequentes em nosso *corpus* paralelo para, logo após, podermos realizar uma análise comparativa entre a co-ocorrência das colocações levantadas no *corpus* paralelo *Queer as Folk* e aquelas presentes no *corpus* comparável, compilado exclusivamente para o desenvolvimento desta pesquisa.

Frisamos que o *WordSmith Tools*, versão 4.0, é um *software* importante para a análise linguística: cada uma de suas ferramentas possui vários instrumentos de análise. Entretanto, descrevemos, aqui, somente as ferramentas utilizadas em nossa pesquisa. A figura 12 mostra a interface do *software*:

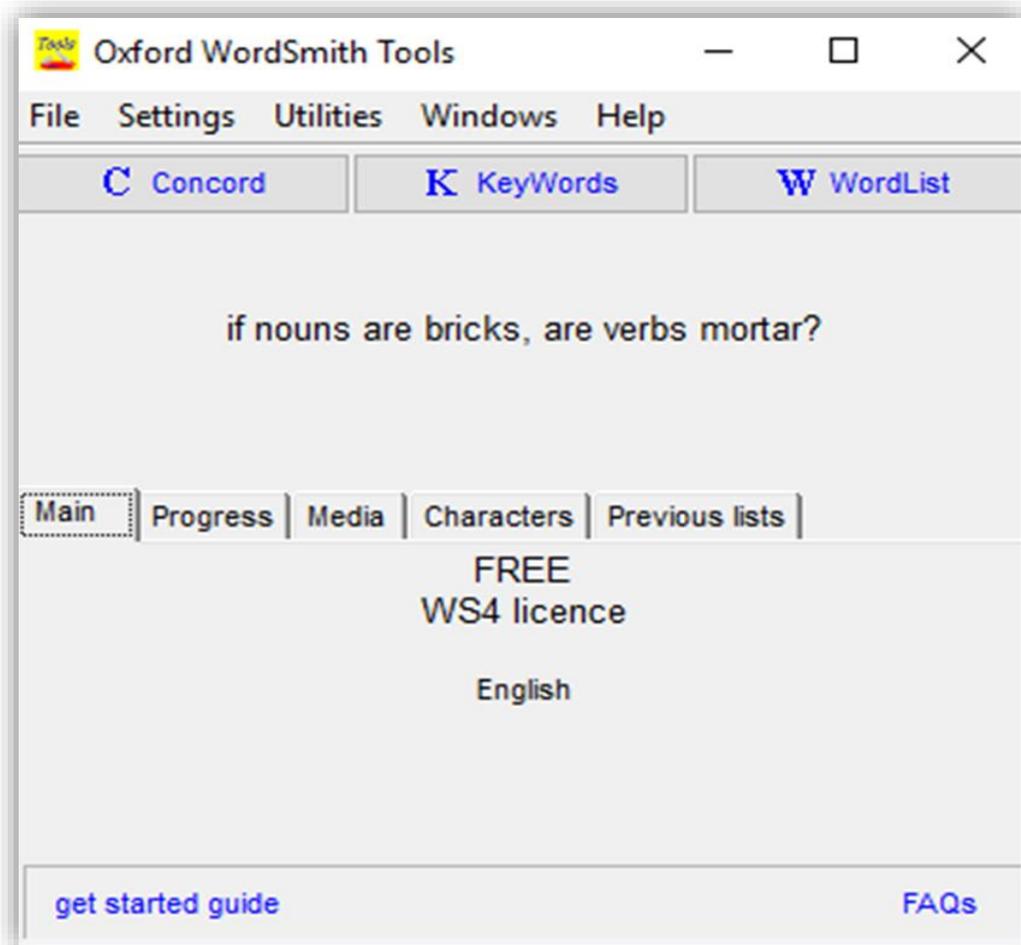


Figura 12 – Interface do *WordSmith Tools*, versão 4.0²⁹.

Fonte: *Printscreen* da tela do *WordSmith Tools*.

A ferramenta apresenta três recursos: *WordList*, *Concord* e *KeyWords*. A *WordList* produz listas de palavras contendo todas as ocorrências do arquivo ou arquivos selecionados,

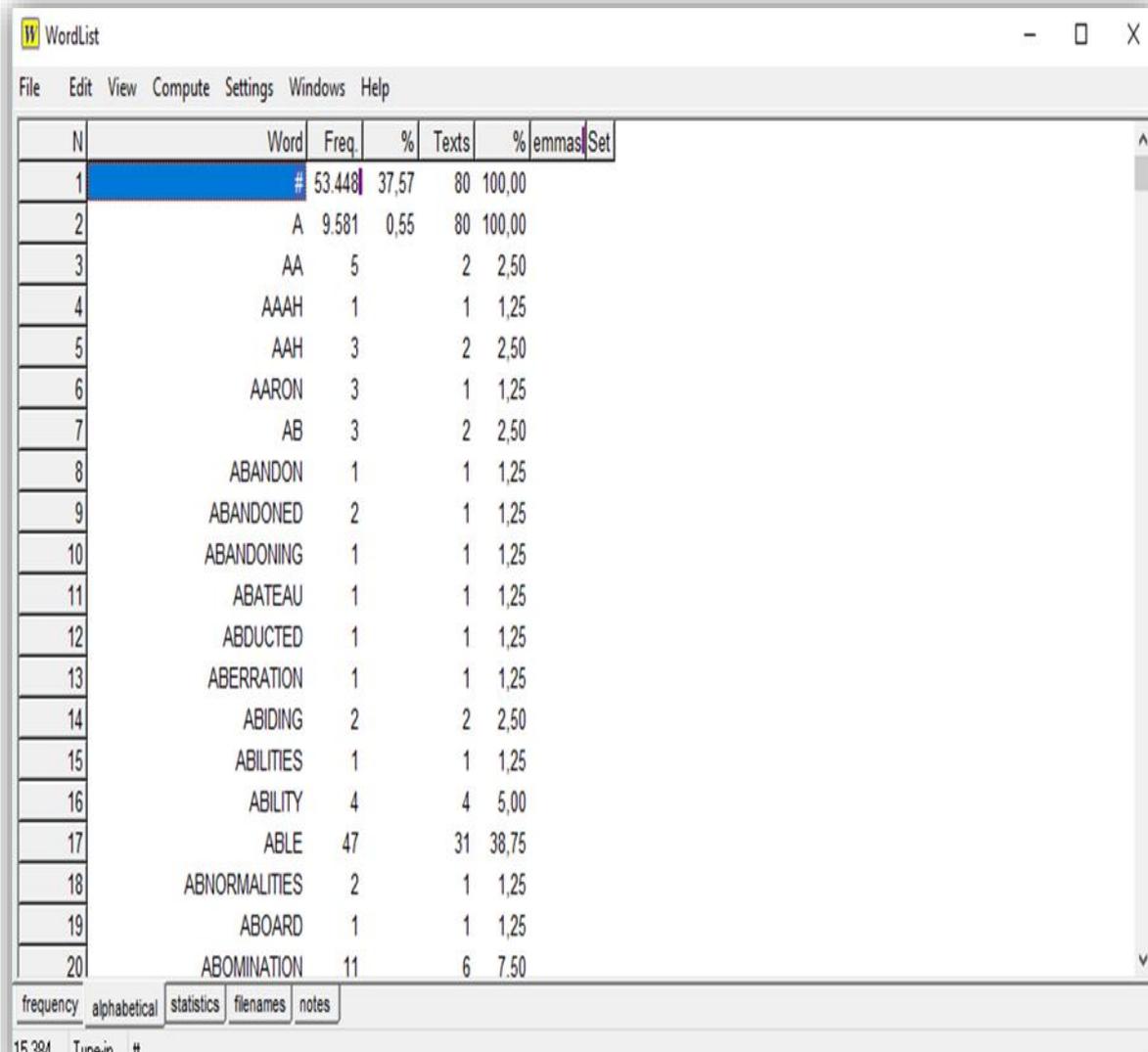
²⁹ Optamos pela utilização da versão 4.0 do programa *WordSmith Tool* em virtude de sua disponibilidade para download gratuito.

elencadas em conjunto com suas frequências absolutas e percentuais. O programa exibe duas linhas de palavras: uma em ordem de frequência, com a palavra mais frequente no início da lista, como demonstrado na figura 13, e outra em ordem alfabética, do A ao Z, como na figura 14. Em companhia das listas de palavras, o programa exibe, em outra janela, estatísticas do *corpus*, conforme figura 15, como o número total de palavras (*tokens*), o número de palavras distintas (*types*), a razão forma/ocorrência ou, em inglês, *type/token ratio*, o número de sentenças, o número de parágrafos, entre outras.

N	Word	Freq.	%	Texts	% emmas	Set
1	#	60.348	37,58	81	100,00	
2	YOU	18.051	1,03	81	100,00	
3	I	14.655	0,83	81	100,00	
4	TO	12.305	0,70	81	100,00	
5	THE	11.263	0,64	81	100,00	
6	A	9.702	0,55	81	100,00	
7	AND	7.120	0,41	81	100,00	
8	IT	5.967	0,34	81	100,00	
9	THAT	5.203	0,30	81	100,00	
10	OF	4.956	0,28	81	100,00	
11	ME	4.230	0,24	81	100,00	
12	WHAT	3.957	0,23	81	100,00	
13	FOR	3.870	0,22	81	100,00	
14	IN	3.792	0,22	81	100,00	
15	YOUR	3.299	0,19	81	100,00	
16	IT'S	3.297	0,19	81	100,00	
17	MY	3.230	0,18	81	100,00	
18	IM	3.214	0,18	81	100,00	
19	HAVE	3.189	0,18	81	100,00	
20	BE	3.153	0,18	81	100,00	
21	IS	3.113	0,18	81	100,00	
22	HE	2.955	0,17	81	100,00	
23	NOT	2.894	0,16	81	100,00	
24	KNOW	2.836	0,16	81	100,00	
25	SO	2.813	0,16	81	100,00	
26	WELL	2.767	0,16	81	100,00	
27	DO	2.740	0,16	81	100,00	
28	DONT	2.652	0,15	81	100,00	
29	WE	2.580	0,15	81	100,00	
30	ON	2.568	0,15	81	100,00	
31	WAS	2.563	0,15	81	100,00	

Figura 13 – Lista de palavras elaborada pela *WordList*, em ordem de frequência.

Fonte: *Printscreen* da tela do *WordSmith Tools*.



N	Word	Freq.	%	Texts	% emmas Set
1	#	53.448	37,57	80	100,00
2	A	9.581	0,55	80	100,00
3	AA	5		2	2,50
4	AAAH	1		1	1,25
5	AAH	3		2	2,50
6	AARON	3		1	1,25
7	AB	3		2	2,50
8	ABANDON	1		1	1,25
9	ABANDONED	2		1	1,25
10	ABANDONING	1		1	1,25
11	ABATEAU	1		1	1,25
12	ABDUCTED	1		1	1,25
13	ABERRATION	1		1	1,25
14	ABIDING	2		2	2,50
15	ABILITIES	1		1	1,25
16	ABILITY	4		4	5,00
17	ABLE	47		31	38,75
18	ABNORMALITIES	2		1	1,25
19	ABOARD	1		1	1,25
20	ABOMINATION	11		6	7,50

frequency alphabetical statistics filenames notes

15.384 Type-in #

Figura 14 – Lista de palavras elaborada pela *WordList*, em ordem alfabética.

Fonte: *Printscreen* da tela do *WordSmith Tools*.

N	Overall	1	2	3	4	5	6	7	8	9	10	11	12	13	14
text file	Overall	v.en.txt													
file size	5.124.356	65.537	56.250	56.766	58.019	50.382	58.078	58.280	54.453	65.145	60.349	54.283	11.181	56.978	54.301
tokens (running words) in text	1.739.151	22.675	19.349	19.926	19.935	17.387	20.242	18.504	18.530	22.333	20.567	18.672	36.225	19.332	19.101
tokens used for word list	1.085.703	14.002	12.013	12.216	12.330	10.770	12.398	12.114	11.602	13.903	12.870	11.578	23.120	11.976	11.731
types (distinct words)	15.384	1.202	1.110	1.016	1.121	1.004	1.228	1.400	1.095	1.251	1.070	1.027	1.248	1.256	941
type/token ratio (TTR)	1	9	9	8	9	9	10	12	9	9	8	9	5	10	10
standardised TTR	20,34	20,70	21,15	20,09	20,92	21,17	21,07	25,63	21,72	21,44	20,91	20,11	14,40	22,10	19,6
standardised TTR std.dev.	79,04	72,11	70,94	71,52	71,16	68,92	70,83	66,88	69,56	71,47	70,99	70,98	79,48	69,46	71,6
standardised TTR basis	1.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00	.000,00
mean word length (in characters)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
word length std.dev.	1,56	1,48	1,52	1,43	1,59	1,49	1,55	1,69	1,53	1,51	1,50	1,50	1,65	1,65	1,4
sentences	73.774,00	.007,00	779,00	886,00	787,00	776,00	866,00	802,00	836,00	950,00	999,00	807,00	481,00	769,00	814,00
mean (in words)	15	14	15	14	16	14	14	15	14	15	13	14	16	16	1
std.dev.	10,20	8,70	8,34	7,55	9,83	9,14	8,83	8,45	8,67	9,53	7,93	8,16	15,90	10,52	7,5
paragraphs	80,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
mean (in words)	13.571	14.002	12.013	12.216	12.330	10.770	12.398	12.114	11.602	13.903	12.870	11.578	23.120	11.976	11.731
std.dev.	3.684,42														
headings															
mean (in words)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figura 15 – Tela do *WordList* com as estatísticas do *corpus*.

Fonte: *Printscreen* da tela do *WordSmith Tools*.

A ferramenta *Concord*, exemplificada na figura 16, gera linhas de concordância, ou seja, listagens das ocorrências de um item específico, chamadas de palavras de busca ou base; optamos, em nosso estudo, pela utilização da *base*, por utilizarmos a tipologia de Hausmann (1985, *apud* ORENHA-OTTAIANO, 2004). Esse item pode ser formado por uma ou mais palavras e é apresentado com o contexto ao seu redor.

Para gerar as linhas de concordância, devemos abrir a ferramenta *Concord* na tela inicial e clicar no arquivo (*file*) e em novo (*new*). O software abrirá uma janela com a opção *escolher textos* (*choose texts now*). Clicando na opção, abre-se uma nova janela na qual o usuário pode selecionar a(s) pasta(s) em que os textos estão armazenados. Selecionados os textos, basta escolher uma palavra de busca e clicar no botão da seta verde.

Concord												
File Edit View Compute Settings Windows Help												
N	Concordance	Set	Tag	Word #	t. #	os. #	. #	os. #	. #	os. #	File	%
1	00:39:08,734 -> 00:39:11,367 And it's gay ski week. 785 00:39:11,367 ->			7.381	628	0%	0	3%	0	3%	13.hr hdtv.en.txt	82%
2	-> 00:39:07,068 I'm on the gay ski team now. 784 00:39:08,734 ->			7.371	627	5%	0	2%	0	2%	13.hr hdtv.en.txt	82%
3	00:34:03,667 -> 00:34:04,868 Straight, gay, 695 00:34:04,868 -> 00:34:06,968			6.519	556	7%	0	3%	0	3%	13.hr hdtv.en.txt	73%
4	00:23:16,968 -> 00:23:19,667 In the gay community, we have drag queens			4.602	383	3%	0	1%	0	1%	13.hr hdtv.en.txt	51%
5	-> 00:22:15,801 just because they're gay. 478 00:22:22,867 -> 00:22:25,534 I			4.488	374	0%	0	0%	0	0%	13.hr hdtv.en.txt	50%
6	for all americans, straight and gay. 465 00:21:24,201 -> 00:21:26,634			4.359	367	0%	0	9%	0	9%	13.hr hdtv.en.txt	49%
7	-> 00:09:37,700 with a slightly gay twist. 196 00:09:37,700 ->			1.841	140	7%	0	1%	0	1%	13.hr hdtv.en.txt	20%
8	-> 00:07:42,467 they see at the gay pride parade. 154 00:07:42,467 ->			1.438	110	5%	0	6%	0	6%	13.hr hdtv.en.txt	16%
9	00:07:36,100 -> 00:07:37,601 that the gay community 152 00:07:37,601 ->			1.416	110	4%	0	6%	0	6%	13.hr hdtv.en.txt	16%
10	on whisking me off to mount flame for gay ski week, 71 00:04:30,633 ->			603	58	9%	0	7%	0	7%	13.hr hdtv.en.txt	7%
11	Remember, you were once a gay young thing yourself. 1191			12.354	171	9%	0	8%	0	8%	12.hr hdtv.en.txt	77%
12	Remember, you were once a gay young thing yourself. 1190			12.341	170	9%	0	8%	0	8%	12.hr hdtv.en.txt	77%
13	-> 00:33:56,620 And I have fought for gay rights more than you have or ever			11.493	113	8%	0	2%	0	2%	12.hr hdtv.en.txt	72%
14	-> 00:33:52,658 And I have fought for gay rights more than you have or ever			11.475	112	8%	0	2%	0	2%	12.hr hdtv.en.txt	72%
15	..where they accept and respect gay people, 1079 00:32:43,505 ->			11.146	079	9%	0	0%	0	0%	12.hr hdtv.en.txt	70%
16	..where they accept and respect gay people, 1078 00:32:40,919 ->			11.135	079	9%	0	0%	0	0%	12.hr hdtv.en.txt	70%
17	-> 00:30:10,519 It's his first time in a gay club, 1015 00:30:10,519 ->			10.464	029	7%	0	6%	0	6%	12.hr hdtv.en.txt	66%
18	-> 00:30:08,851 It's his first time in a gay club, 1014 00:30:08,851 ->			10.452	029	4%	0	6%	0	6%	12.hr hdtv.en.txt	65%
19	-> 00:28:46,185 You know, it's the gay pulse that keeps on beating no			10.069	992	6%	0	3%	0	3%	12.hr hdtv.en.txt	63%
20	-> 00:28:40,220 You know, it's the gay pulse that keeps on beating no			10.052	991	6%	0	3%	0	3%	12.hr hdtv.en.txt	63%
21	00:25:51,844 -> 00:25:55,305 If I was gay and I had kids, I'd be worried too.			9.031	900	3%	0	7%	0	7%	12.hr hdtv.en.txt	57%
22	00:25:51,802 -> 00:25:51,844 If I was gay and I had kids, I'd be worried too.			9.015	899	3%	0	7%	0	7%	12.hr hdtv.en.txt	57%
23	00:25:39,164 Yeah, for you. You're not gay. 869 00:25:39,164 -> 00:25:39,206			8.881	890	0%	0	6%	0	6%	12.hr hdtv.en.txt	56%
24	00:25:37,329 Yeah, for you. You're not gay. 868 00:25:37,329 -> 00:25:39,164			8.871	888	0%	0	6%	0	6%	12.hr hdtv.en.txt	56%
25	doctor refused to take it because he's gay. 111 00:03:52,107 -> 00:03:52,149			947	87	0%	0	6%	0	6%	12.hr hdtv.en.txt	6%
26	doctor refused to take it because he's gay. 110 00:03:50,063 -> 00:03:52,107			933	87	3%	0	6%	0	6%	12.hr hdtv.en.txt	6%
27	This is nothing to do with being gay or straight. 380 00:13:31,103 ->			4.105	414	7%	0	5%	0	5%	11.hr hdtv.en.txt	25%
28	This is nothing to do with being gay or straight. 379 00:13:27,808 ->			4.091	413	7%	0	5%	0	5%	11.hr hdtv.en.txt	25%
29	about losing your fans now that you're gay? 1153 00:38:01,506 ->			12.703	372	0%	0	2%	0	2%	10.hr hdtv.en.txt	92%
30	about losing your fans now that you're gay? 1152 00:37:59,046 ->			12.688	371	0%	0	2%	0	2%	10.hr hdtv.en.txt	91%
31	00:37:17,917 -> 00:37:20,588 Are you gay? What's that got to do with			12.357	233	0%	0	0%	0	0%	10.hr hdtv.en.txt	80%

concordance collocates plot patterns clusters filenames source text notes

438 Set GAY

Figura 16 – Aba de concordâncias da ferramenta *Concord*.
Fonte: *Printscreen* da tela do *WordSmith Tools*.

Oferece também listas de *collocates*, como na figura 17, isto é, palavras que ocorreram perto da base. O sucesso da busca no *Concord* depende da especificação correta do termo de busca.

The screenshot shows the Concord software interface with a concordance table for the word 'GAY'. The table has columns for N, Word, With, elation, Total, tal Left, tal Right, L5, L4, L3, L2, L1, Centre, R1, R2, R3, R4, and R5. The word 'GAY' is highlighted in blue in the first row. The table lists various words that co-occur with 'GAY' and their respective counts in different positions relative to the search term.

N	Word	With	elation	Total	tal Left	tal Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	GAY	gay	0,000	440	1	1	0	0	0	1	0	438	0	1	0	0	0
2	OO	gay	0,000	129	56	73	33	23	0	0	0	0	0	17	32	7	17
3	THE	gay	0,000	107	101	6	9	14	7	11	60	0	0	0	4	0	2
4	A	gay	0,000	86	77	9	1	6	9	9	52	0	0	0	5	0	4
5	TO	gay	0,000	66	54	12	4	12	16	13	9	0	0	0	6	4	2
6	AND	gay	0,000	59	20	39	6	6	1	4	3	0	25	8	0	4	2
7	OF	gay	0,000	54	49	5	1	5	9	15	19	0	0	3	1	0	1
8	YOU	gay	0,000	44	35	9	1	14	8	9	3	0	0	5	2	2	0
9	I	gay	0,000	40	36	4	11	6	7	12	0	0	0	3	0	1	0
10	IN	gay	0,000	33	16	17	1	0	1	12	2	0	0	7	5	4	1
11	THAT	gay	0,000	29	26	3	2	4	5	9	6	0	0	2	1	0	0
12	STRAIGHT	gay	0,000	28	4	24	0	1	0	2	1	0	11	10	3	0	0
13	IS	gay	0,000	26	18	8	5	1	4	5	3	0	0	6	0	2	0
14	WAS	gay	0,000	24	22	2	2	1	0	3	16	0	0	1	1	0	0
15	NOT	gay	0,000	24	22	2	1	4	6	2	9	0	0	1	1	0	0
16	YOU'RE	gay	0,000	23	21	2	3	1	1	4	12	0	1	1	0	0	0
17	AS	gay	0,000	22	10	12	1	1	1	2	5	0	9	2	0	0	1
18	IT	gay	0,000	22	22	0	3	3	3	3	10	0	0	0	0	0	0
19	BE	gay	0,000	21	17	4	0	0	5	6	6	0	0	0	1	3	0
20	MEN	gay	0,000	21	2	19	0	0	2	0	0	0	19	0	0	0	0
21	PEOPLE	gay	0,000	20	3	17	0	2	0	1	0	0	17	0	0	0	0
22	AT	gay	0,000	20	12	8	0	0	0	11	1	0	2	2	4	0	0
23	FOR	gay	0,000	20	17	3	2	1	1	3	10	0	1	1	0	0	1
24	HE	gay	0,000	19	16	3	1	3	1	10	1	0	2	0	0	0	1
25	LESBIAN	gay	0,000	19	0	19	0	0	0	0	0	0	0	19	0	0	0
26	OR	gay	0,000	18	5	13	0	0	1	3	1	0	10	2	0	1	0
27	ABOUT	gay	0,000	18	16	2	2	0	3	5	6	0	0	0	1	1	0
28	WITH	gay	0,000	18	13	5	0	2	5	4	2	0	0	2	0	1	2
29	HAVE	gay	0,000	17	7	10	1	0	3	3	0	0	0	4	3	0	3
30	IT'S	gay	0,000	17	17	0	3	5	1	7	1	0	0	0	0	0	0
31	WHEN	gay	0,000	17	16	1	0	1	5	4	5	0	0	0	1	0	0

At the bottom of the window, there are tabs for 'concordance', 'collocates', 'plot', 'patterns', 'clusters', 'filenames', 'source text', and 'notes'. The 'collocates' tab is selected. Below the tabs, it shows '144 Type-in GAY'.

Figura 17 – Aba de concordâncias da ferramenta *Concord* na aba *collocates*.

Fonte: *Printscreen* da tela do *WordSmith Tools*.

Outro dispositivo, conforme figura 18, é o *Clusters*, que pode ser descrito como associações de palavras que co-ocorrem com uma determina frequência dentro de um agrupamento de palavras (2, 3, 4 ou mais), sem que necessariamente faça sentido, ou seja, fraseológicas. Essa ferramenta é bastante utilizada quando o número de linhas de concordância é extensivo, o que dificulta a análise. Para obter uma lista de *clusters*, clicamos na aba *clusters* que aparece no fim da tela da ferramenta *Concord*, como mostra a figura a seguir:

N	Cluster	Freq	Length
1	GAY AND LESBIAN	19	3
2	THE GAY AND	16	3
3	AND LESBIAN CENTER	14	3
4	AT THE GAY	11	3
5	I WAS GAY	9	3
6	GAY OR STRAIGHT	8	3
7	IN A GAY	8	3
8	GAY AS BLAZES	7	3
9	THE GAY COMMUNITY	7	3
10	THE GAY STRAIGHT	6	3
11	OF THE GAY	6	3
12	A GAY BAR	6	3
13	A GAY MAN	6	3
14	HE WAS GAY	6	3
15	TO BE GAY	5	3
16	TO BE A	5	3
17	A GAY SUPERHERO	5	3
18	MY GAY SON	5	3
19	GAY STRAIGHT STUDENT	5	3
20	STRAIGHT STUDENT ALLIANCE	5	3
21	BE A GAY	5	3

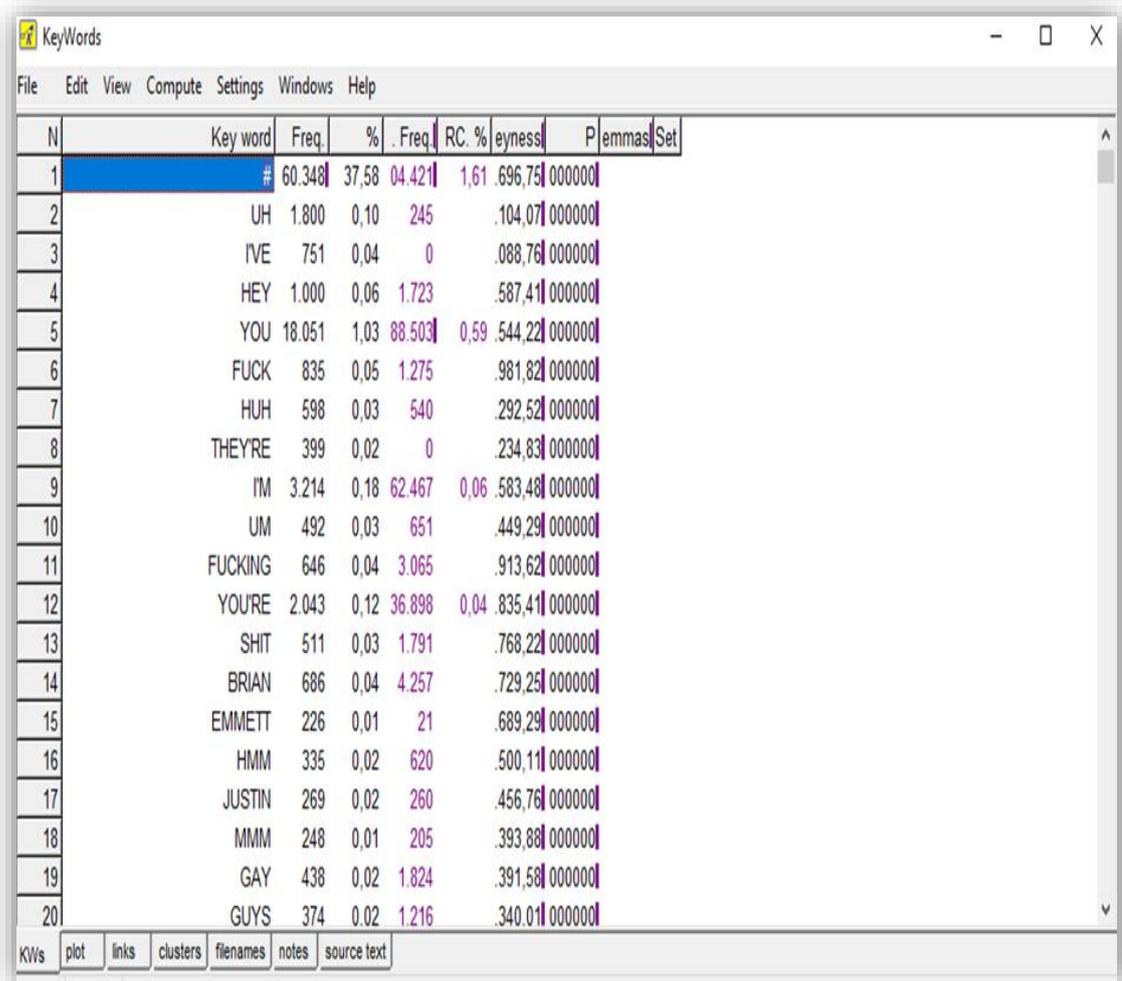
↓

concordance collocates plot patterns **clusters** filenames source text notes

21 Type-in GAY

Figura 18 – Aba de concordâncias da ferramenta *Concord* na aba *clusters*.
Fonte: *Printscreen* da tela do *WordSmith Tools*.

A *KeyWords*, demonstrada na figura 19, extrai palavras de uma lista, cujas frequências são estatisticamente diferentes (maiores ou menores) do que as frequências das mesmas palavras em outro *corpus*. Calcula também palavras-chave, que são chave em vários textos, observando que as palavras-chave não são consideradas o mesmo que palavras “importantes”.



The screenshot shows the KeyWords software window with a menu bar (File, Edit, View, Compute, Settings, Windows, Help) and a table of keyword statistics. The table has columns for N, Key word, Freq., %, .Freq., RC. %, eyness, P, emmas, and Set. The first row is highlighted in blue.

N	Key word	Freq.	%	.Freq.	RC. %	eyness	P	emmas	Set
1	#	60.348	37,58	04.421	1,61	.696,75	000000		
2	UH	1.800	0,10	245		.104,07	000000		
3	IVE	751	0,04	0		.088,76	000000		
4	HEY	1.000	0,06	1.723		.587,41	000000		
5	YOU	18.051	1,03	88.503	0,59	.544,22	000000		
6	FUCK	835	0,05	1.275		.981,82	000000		
7	HUH	598	0,03	540		.292,52	000000		
8	THEYRE	399	0,02	0		.234,83	000000		
9	IM	3.214	0,18	62.467	0,06	.583,48	000000		
10	UM	492	0,03	651		.449,29	000000		
11	FUCKING	646	0,04	3.065		.913,62	000000		
12	YOU'RE	2.043	0,12	36.898	0,04	.835,41	000000		
13	SHIT	511	0,03	1.791		.768,22	000000		
14	BRIAN	686	0,04	4.257		.729,25	000000		
15	EMMETT	226	0,01	21		.689,29	000000		
16	HMM	335	0,02	620		.500,11	000000		
17	JUSTIN	269	0,02	260		.456,76	000000		
18	MMM	248	0,01	205		.393,88	000000		
19	GAY	438	0,02	1.824		.391,58	000000		
20	GUYS	374	0,02	1.216		.340,01	000000		

Figura 19 – Aba da ferramenta *Keywords*.
Fonte: *Printscreen* da tela do *WordSmith Tools*.

Após a seleção das palavras-chave, o passo seguinte foi o levantamento das colocações, com o auxílio da ferramenta *Concord*, por meio das abas *concordance* e *collocates*. Durante nosso estudo, escolhemos a base *gay* (homossexual), da qual foram levantadas colocações com este vocábulo que nos parece relevante para a comunidade homossexual, também levando em consideração a frequência apresentada em nosso *corpus* paralelo e no *corpus* comparável, chamamos a atenção para as colocações com *gay*, já que tal palavra se torna *colocado* em tais combinações.

Outra ferramenta para análise lexical, análise comparativa de dados e interpretação das colocações levantadas é a ferramenta *Sketch Engine* (KILGARRIFF *et al*, 2004), que é utilizada para explorar como a língua funciona. Seus algoritmos analisam textos autênticos de bilhões de palavras para identificar instantaneamente o que é típico na língua e o que é uso raro ou incomum.

No escopo da referida plataforma está inserida a ferramenta *BootCat*, por meio da qual foi possível compilar o *corpus* comparável, composto por um *subcorpus* em inglês e pelo *subcorpus* em português do Brasil, para, assim, podermos aferir com precisão a utilização das colocações destacadas em nosso estudo.

No próprio *Sketch Engine*, podemos clicar no ícone do *BootCat*, conforme figura 20. De forma geral, a ferramenta permite a compilação automática de *corpora* via *web*, partindo da combinação de uma lista de *seeds* (*sementes*; esses itens lexicais foram escolhidos por meio da ferramenta *Keywords*, presente no *WordSmith Tools*, versão 4.0), ou palavras-chave de determinada área de especialidade ou do léxico da língua geral.

Listadas e combinadas as primeiras *seeds* (palavra de busca) para a compilação do *subcorpus* comparável em inglês, utilizamos as *seeds*: *gay advocate*, *gay bashing*, *gay crusader*, *gay eunuch*, *gay issues*, *gay marriage*, *gay mayor*, *gay rag*, *gay rights*, *gay thump-thump*, *gay twist*, *fag hater*, *stepford fag*, *fag hag*, *fag jokes*, *goddamn fag*, *godless fag*, *professional fag*, *brave faggot*, *fucking faggot*, *little faggot*, *silly faggot*. e, para a compilação do *subcorpus* comparável em português, utilizamos as *seeds*: *ladrão gay*, *defensor dos gays*, *preconceito gays*, *guerreiro gay*, *gay estéreo*, *causas gays*, *casamento gay*, *destaque gay*, *direitos dos gays*, *vibração gay*, *toque gay*, *odiar viado*, *bicha pra casar*, *líder gay*, *piadas de bichas*, *bicha maldita*, *viado ateu*, *bicha profissional*, *viado corajoso*, *bicha de merda*, *bichinha*, *bicha idiota*; o programa, por conseguinte, faz uma varredura em *sites* pela *web* à procura dos textos em que elas ocorrem.

Terminada a varredura, as páginas selecionadas são baixadas e os códigos *HTML* são removidos, como na figura 21. Finalizando tais procedimento, tem-se um *corpus* compilado, salvo em um arquivo em formato *txt*, como podemos ver na figura 22:

Sketch Engine

Home

- + Create corpus
- + WebBootCaT
- + Upload TMX or XLS

Parallel corpora

Compare corpora

My jobs

Advanced features

- Corpus templates
- Sketch grammars
- Subcorpus definitions
- GDEX configurations
- User groups
- Subscription overview

Support

- User guide
- Feedback

WebBootCaT: Create corpus ?

[Get seed words from Wikipedia](#)

Corpus name

Language WebBootCaT is available for languages that can be automatically tokenised.

Input type

- Seed words
- URLs
- Website

Select Seed words for finding documents using a search engine. Use URLs to download texts directly from specified locations. Switch to Website in case you need to obtain all documents from a particular web domain.

Seed words

Random tuples will be selected from the seed words to query a search engine. Input 3 to 20 words or multiword expressions. Use space as separator. Enclose multiword expressions in quotes (").

Compile corpus when finished Automatically compile corpus when WebBootCaT processing is finished.

[Show advanced options](#)

Figura 20 – Aba de compilação do *corpus*.
Fonte: *Printscreen* da tela do *Sketch Engine*.

Sketch Engine

Home

- + Create corpus
- + WebBootCaT
- + Upload TMX or XLS

Parallel corpora

Compare corpora

My jobs

Advanced features

Corpus templates

Sketch grammars

Subcorpus definitions

GDEX configurations

User groups

Subscription overview

Manage corpus

- Show corpus files
- Compile corpus
- Configure corpus
- Set sketch grammar
- Set subcorpora
- Download corpus
- Share corpus
- View logs

Search corpus

Corpus is not compiled.

Support

Gay Language: WebBootCaT: Downloading data...

You can close this page and return to check the progress later.

36%

⚠ Harvesting web sites is 2 times faster with the full account.

Successfully processed files	20	Errors	15
Files remaining	62	- unable to retrieve	5
Data downloaded	4207 kB	- invalid content-type	0
Words retrieved	129,439	- file size out of range	0
Words per file (avg)	6,471	- cleaned file size out of range	10
Time elapsed	0:52	- keywords filter applied	0
Estimated time remaining	1:33	- unable to convert to text	0
Average file processing time	1.5 s	- duplicate	0

✖ [Cancel processing](#)

```

- Data read: 256.7 kB
- Detected character encoding: utf_8
- Plain text size: 10321 characters
Processing https://www.filmdoo.com/gay
- Failed to retrieve:
Processing http://corsinet.com/braincandy/jkmen.html
- Content-type: text/html
- File type: html
- Data read: 26.1 kB
- Detected character encoding: utf_8
- Plain text size: 5190 characters
Processing http://www.theonion.com/blogpost/let-us-identify-the-faggots-and-then-inform-
- Content-type: text/html; charset=utf-8
- File type: html
- Data read: 134.5 kB
- Detected character encoding: utf_8
- Plain text size: 9744 characters
Processing http://www.huffingtonpost.com/2013/10/16/anti-gay-companies_n_4110344.html
- Failed to retrieve:

```

Figura 21 – A ferramenta realizando *download* dos dados para a compilação do *corpus*.

Fonte: *Printscreen* da tela do *Sketch Engine*.

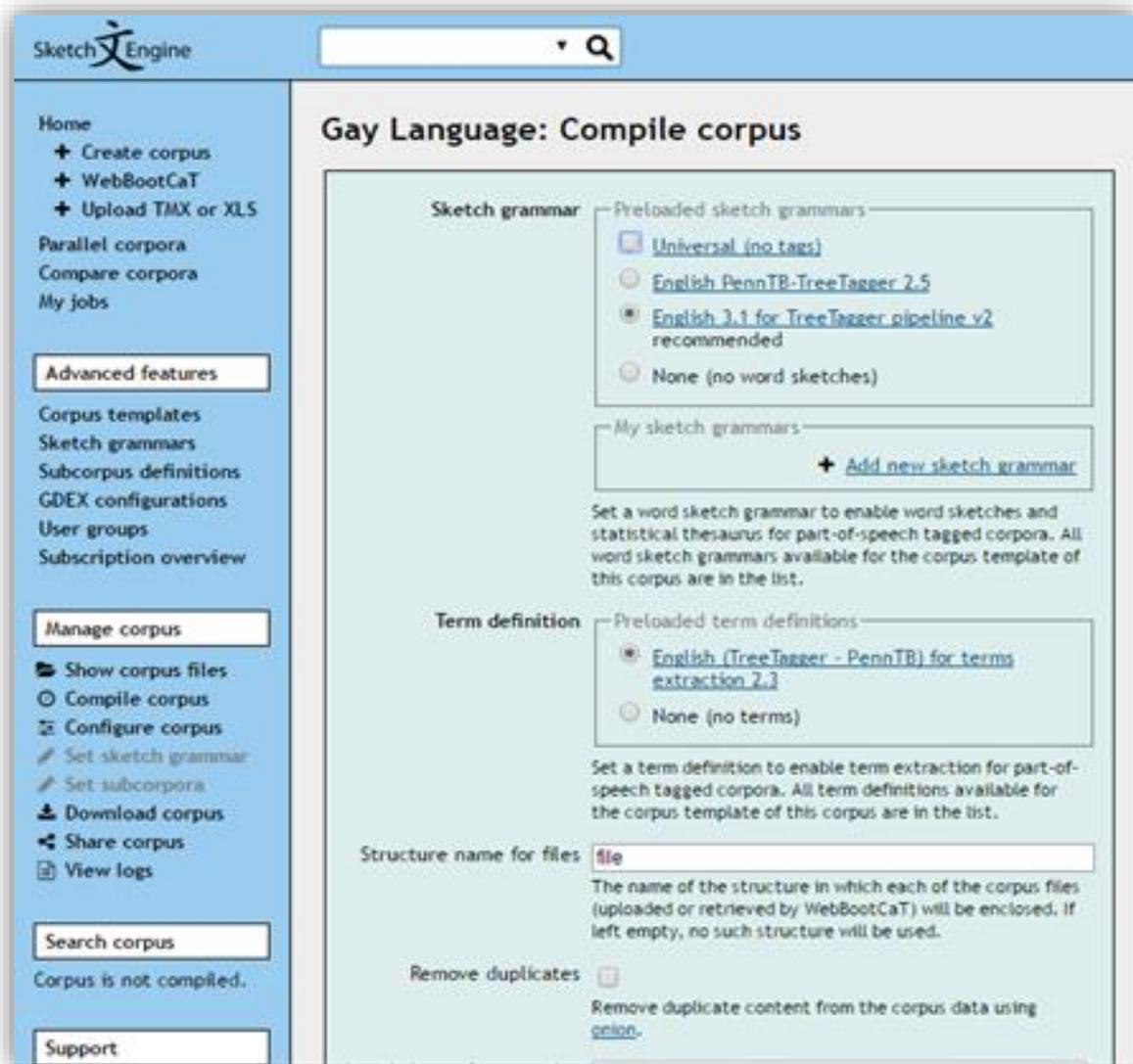


Figura 22 – Aba com as páginas selecionadas na *web* para a compilação do *corpus*.

Fonte: *Printscreen* da tela do *Sketch Engine*.

Os dois *subcorpora* comparáveis compilados foram importantes para comprovar se os item lexicais levantados são frequentemente usuais, já que a utilização de um *corpus* de referência não demonstrou ser de grande valia em nosso estudo, por se tratar de colocações restritas em sua grande maioria, utilizadas preferencialmente pela comunidade homossexual e, desse modo, não ocorreriam de maneira constante neste *corpus*.

Na sequência, vamos analisar e discutir os dados contidos em nosso *corpus* paralelo e no *corpus* comparável, as frequências que correspondem a algumas colocações levantadas. No capítulo a seguir, temos a *Extração e Análise das Colocações da Comunidade Homossexual*.

4 EXTRAÇÃO E ANÁLISE DAS COLOCAÇÕES DA COMUNIDADE HOMOSSEXUAL

Das 500 palavras-chave (*KeyWords*) de nosso *corpus* paralelo, composto pelo *subcorpus* em inglês, disponibilizadas pelo *software WordSmith Tools 4.0* (SCOTT, 2012), foram selecionadas, para este trabalho, cinco itens lexicais principais que remetem ao léxico da comunidade homossexual, comumente empregados pelo referido grupo e que poderiam formar colocações da língua geral, a saber: *ass*, *butt*, *fag*, *faggot* e *gay*. Como se trata de uma delimitação de pesquisa, embora estas palavras possam ocorrer ora como base ora como colocado, decidimos por selecioná-las, dada sua importância neste contexto.

Conforme mencionado na metodologia, as linhas de concordância para a palavra de busca *gay*, geradas pela ferramenta *Concord*, mostram que se trata de uma palavra recorrente em nosso *corpus* paralelo, com a frequência de 438. Por meio da análise preliminar quanto à sua co-ocorrência, notamos que *gay* tinha potencial para formação de colocações no *corpus* paralelo. Dessa forma, utilizamos a aba *collocates* da ferramenta *Concord* para identificar palavras que apareciam à direita e à esquerda da palavra de busca, a partir das quais selecionamos aquelas pertencentes ao léxico da comunidade homossexual e que formavam colocações.

Vale lembrar que as colocações para estas bases estão separadas por sua formação morfológica, segundo classificação apresentada na *Fundamentação Teórica*, em que se destacam, neste levantamento, as colocações substantivas: substantivo + substantivo (*fag hater* – “pessoa que odeia gays”, observamos que em português não gera uma colocação) e as colocações adjetivas: adjetivo + substantivo (*gay twist* – “toque gay”).

Observamos também o equivalente de cada colocação em língua portuguesa, segundo a legenda em português que forma nosso *corpus* paralelo. Nos exemplos em que as traduções sugeridas não nos pareciam adequadas ao contexto analisado e ao emprego daquela colocação junto à comunidade de estudo, optamos por apresentar outras opções tradutórias. Nestes casos, as traduções serão marcadas com um asterisco (*), para que, dessa maneira, fique claro quais foram as nossas sugestões. Ao oferecer outra possibilidade de tradução, não utilizamos os exemplos em português do *corpus* paralelo, *subcorpus* de português, e inserimos exemplos extraídos do *corpus* comparável, *subcorpus* do português. Caso não haja essa marca (asterisco), subentende-se que as traduções e os exemplos em português foram retirados do *corpus* paralelo.

A seguir, conforme metodologia empregada por Orenha-Ottaiano (2004, 2016) para a extração de colocações e compilação de obras fraseográficas, especialmente dicionários e glossários de colocações, apresentamos a análise de colocações a partir de cinco itens lexicais selecionados por sua alta frequência no *corpus* paralelo, segundo mencionamos anteriormente: *ass*, *butt*, *fag*, *faggot* e *gay*. As colocações geradas a partir destes cinco itens lexicais irão compor a amostra de nosso *Glossário Bilíngue de Colocações da Comunidade Homossexual*.

Gay (substantivo) + Substantivo	
Gayon-gay	Ladrão gay
<i>I'm starting to suspect gayon-gay crime.</i>	Estou começando a suspeitar de um ladrão gay .

Quadro 5 – Colocação substantiva a partir da base *gay*

Gay (adjetivo) + Substantivo	
Gay advocate <i>The centre's giving him their outstanding gay advocate award.</i>	Defensor dos gays Dizem aqui que ele receberá o prêmio de Defensor dos Gays .
Gay bashing <i>They don't want to think about it. Or about gay bashing or Aids, or any of the other shit we all have to live with.</i>	Preconceito gay Eles não querem pensar nisso. Nem em preconceito gay ou Aids, ou nem nas coisas com as quais convivemos.
Gay crusader <i>My development people are going to shop it around, the gay crusader is too powerful to be defeated by some asshole who can't see beyond the box office.</i>	Guerreiro gay Meu pessoal de desenvolvimento está tentando vendê-lo, um guerreiro gay é muito poderoso para ser derrotado por alguns idiotas que não podem ver além das bilheterias.
Gay eunuch <i>It's an honour to count among our close personal friends television's latest gay eunuch.</i>	Gay estéril* É uma honra apoiar meu amigo íntimo que se tornou o gay estéreo da televisão.
Gay issues <i>Senator Bexter's very supportive of gay issues.</i>	Causas gays A senadora Bexter apoia causas gays , claro que gostariam.
Gay marriage <i>It's not enough they can say no to gay marriage. Now they're gonna go after everything?</i>	Casamento gay Não foi o bastante o fato de terem dito "não" ao casamento gay e agora eles querem lhe tirar tudo?
Gay mayor <i>Well, last time I heard, they had a gay mayor.</i>	Prefeito gay A última coisa que ouvi é que eles tiveram um prefeito gay .
Gay rag <i>Anyone seen a certain hot little item in our local gay rag?</i>	Jornal gay Já viram o destaque no jornal gay ?
Gay rights <i>And I have fought for gay rights more than you have or ever will.</i>	Direitos dos gays E eu luto pelos direitos dos gays mais do que você luta ou alguma vez lutou.
Gay thump-thump <i>Come on, you. Ah-h! Do you feel that gay thump-thump? God, I missed that!</i>	Vibração gay Venha. Sente a vibração gay ? Puxa, senti falta disso!
Gay twist <i>Welcome to that quaint, heterosexual male tradition known as "the stag party", with a slightly gay twist.</i>	Toque gay Bem-vindo a curiosa tradição heterossexual masculina conhecida como "Despedida de Solteiro". Com um leve toque gay .

Quadro 6 – Colocações adjetivas a partir do colocado *gay*

A fim de verificar o uso e a frequência das colocações levantadas no *corpus* paralelo, em relação ao *corpus* comparável anteriormente mencionado, apresentamos a frequência levantada da colocação *gay rights* no *subcorpus* paralelo em inglês (item 1, da tabela 1 abaixo), no *subcorpus* paralelo em português do Brasil, ou seja, de sua tradução para o português – “direitos dos gays” (item 2, da tabela 1), frequência do *subcorpus* comparável em inglês (item 3, da tabela 1) e frequência do *subcorpus* comparável em português (item 4, da tabela 1). Salientamos que os *corpora* comparáveis foram compilados dentro da plataforma *Sketch Engine* (KILGARRIFF *et al*, 2004), tendo por finalidade conferir se as colocações extraídas em nossa pesquisa realmente são usuais na língua geral pela comunidade homossexual.

Tabela 1 – Estatísticas da colocação *gay rights* e sua tradução para o português

Gay (adjetivo) + Substantivo			
Gay rights - Direitos dos gays			
1 – Subcorpus Paralelo em Inglês	2 – Subcorpus Paralelo em Português	3 – Subcorpus Comparável em Inglês	4- Subcorpus Comparável em Português
04	04	36	01

A colocação adjetiva extraída de nosso *corpus* paralelo, como descrito na tabela acima e exemplificada nas figuras 23 e 24, estruturam-se sintagmaticamente em *adj. + noun*, em inglês. É formada pelo adjetivo em função atributiva (QUIRK *at al*, 1985), isto é, modificando o substantivo *rights*.

Ao observarmos a frequência de tal colocação em nosso *corpus* paralelo, a referida colocação ocorre 04 vezes e, no *subcorpus* comparável em inglês, aparece com 36 ocorrências. No entanto, a opção tradutória, *direitos dos gays* ocorre apenas uma vez no *subcorpus* comparável em português.

A seguir, temos as figuras 23 e 24 que comprovam a frequência de nossa colocação em nosso *corpus* comparável, compilado para nossa pesquisa, a fim de garantir que nossa colocação seja de fato usual na língua em uso.

Sketch Engine gay rights Subcorpus Inglês

Home Search Word list Word sketch Thesaurus Sketch diff Keywords/terms Corpus info Manage corpus My jobs User guide

Save Make subcorpus View options KWIC Sentence Sort Left Right Node References Shuffle Sample Filter Sub-hits 1st hit in doc Frequency

Query **gay, rights** 36 (24.83 per million) ⓘ

Page 1 of 2 Go Next Last

file540633... my life. I have been an active participant in the **gay rights** movement and my family sacrificed more than you
 file540621... 1950s, in an early piece of lesbian studies, the **gay rights** campaigning organisation ONE, Inc. assigned
 file540621...). Before Stonewall: Activists in lesbian and **gay rights** in historical context. New York: Harrington
 file540641... only democrats have abortions, they believe in **gay rights**, that is like saying there is no gay republicans
 file540626... played by actor and entertainer. Some **gay rights** advocates acknowledge that as a spoof it is
 file540626... in the LGBT community; GLAAD and other **gay rights** organizations demanded to know the reason why
 file540626... [37][38] Her comments triggered a campaign by a **gay rights** group and media watchdog to persuade
 file540639... played by actor and entertainer. Some **gay rights** advocates acknowledge that as a spoof it is
 file540639... , Value war: public opinion and the politics of **gay rights**, page 60 </p> <p> ^ abThe American Heritage
 file540638... played by actor and entertainer. Some **gay rights** advocates acknowledge that as a spoof it is
 file540632... , Value war: public opinion and the politics of **gay rights**, page 60 </p> <p> ^ abThe American Heritage
 file540632... diversity/ curriculum" recently clashed with **gay rights** proponents: </p> <p> When up to a thousand
 file540632... more lurid aspects of later tellings: </p> <p> [**Gay rights** activists] interrupted the speeches with
 file540622... , and the assumed Conservative position on **gay rights** generally, and perhaps gay marriage in
 file540622... of Republicans break party lines to fight for **gay rights**, he decided then and there he was going to be a
 file540644... <p> "the imagery of sex should be downplayed and **gay rights** should be reduced to an abstract social
 file540642... work right-wingers did for women's rights, **gay rights**, the civil rights movement, etc! Such freedom
 file540621... the same rights as men </p> <p> Had to explain **Gay Rights** and Equality to someone who likened gay
 file540634... and outspoken critic of feminism, Muslims and **gay rights** (despite being openly gay himself). Although
 file540637... all the work he has done with gay marriage, **gay rights** and transgendered issues you wonder how he has

Figura 23 – Frequência de *gay rights* no *subcorpus* comparável em inglês

Fonte: Printscreen da tela do Sketch Engine.

Sketch Engine direito dos gays Subcorpus Português

Home Search Word list Word sketch Thesaurus Sketch diff Keywords/terms Corpus info Manage corpus My jobs User guide

Hide ads

Query **direito, dos, gays** 1 (0.80 per million) ⓘ

file540659... , que usou sua posição política para lutar pelos **direitos dos gays** nos anos 70. Segundo relata o filme, isso só

Figura 24 – Frequência de “direitos dos gays” no *subcorpus* comparável em português

Fonte: Printscreen da tela do Sketch Engine.

Na figura 23, na primeira linha de concordância, podemos observar que a colocação *gay rights* faz referência a “uma participação ativa no movimento dos direitos dos gays” e, na figura 24, temos “um grupo formado por ativistas dos direitos gays”. Vemos por meio desta colocação o empenho da comunidade para garantir os “direitos dos gays”.

A tabela 2, a seguir, mostra a frequência da colocação *gay marriage* em nosso *subcorpus* paralelo em inglês e a respectiva colocação *casamento gay* em nosso *subcorpus* paralelo em português do Brasil, em relação aos *subcorpora* comparáveis, em inglês e em português. Neste caso, *gay* figura como um *colocado* e foi selecionado para análise dada sua relevância para o presente estudo.

Tabela 2 – Estatísticas da colocação *gay marriage* e sua tradução para o português

Gay (adjetivo) + Substantivo			
Gay marriage - Casamento gay			
1 – Subcorpus Paralelo em Inglês	2 – Subcorpus Paralelo em Português	3 – Subcorpus Comparável em Inglês	4 – Subcorpus Comparável em Português
04	04	42	27

Com base nos dados coletados acima, observamos que a frequência das colocações *gay marriage* em inglês e sua respectiva tradução *casamento gay* em português apresentaram diferenças. No item 1, a referida colocação aparece quatro vezes no *subcorpus* paralelo, já no item 2, no *subcorpus* paralelo em português ela aparece com a frequência igual, o que nos mostra equilíbrio no uso no *corpus* paralelo analisado.

Tendo em vista que se trata de uma colocação frequentemente empregada pela comunidade *gay*, resolvemos realizar a análise utilizando nosso *corpus* comparável. Depreendemos, no item 3, que a colocação é muito empregada em língua inglesa, já que apresenta uma frequência de 42 ocorrências, mas, por outro lado, sua tradução para o português, *casamento gay*, não apresenta uma frequência tão alta em relação à colocação em língua inglesa, com apenas 27 ocorrências, conforme mostra o item 4. Mesmo assim, após tal levantamento, podemos concluir que a colocação em questão apresenta uso frequente na comunidade homossexual.

Exemplificaremos a seguir, por meio das figuras 25 e 26, a frequência da colocação *gay marriage* e “casamento gay” apresentadas em nosso *Corpus Comparável*. Frisamos, ainda, que o *subcorpus* comparável em inglês é composto por 1.174.067 milhão de palavras, enquanto o *subcorpus* comparável em português, é composto por 1.000.598 milhão de palavras. Ilustraremos a seguir nossos levantamentos, a fim de comprovar nossos dados.

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Corpus info
Manage corpus
My jobs
User guide

Save
Make subcorpus
View options
KWIC
Sentence
Sort
Left
Right
Node
References
Shuffle
Sample
Filter
Sub-hits
1st hit in doc
Frequency
Node tags
Node forms
Doc IDs
Collections

Query **Gay, marriage** 42 (72.35 per million)

Page 1 of 3 Go Next Last

file409639... superficialities and hedonistic interests. When **gay marriage** was legalised my husband and I didn't give
file409640... perversion. Just because i dont believe in **gay marriage** does not i 'hate' gays, just as non-believers
file409638... equal protection, Constitution, gay rights, **gay marriage** , gays in the military, DADT, same sex marriage
file409635... rights advances, such as the legalization of **gay marriage** . It says that (gay) people still
file409638... Constitution or the Bible. If you're against **gay marriage** , just be honest, put a scarlet 'H' on your
file409640... Union Foundation. Lord Advocate issues **gay marriage** debate warning The same sex marriage
file409640... prosecutors as the historic bill to legalise **gay marriage** was published in the Scottish Parliament
file409638... equal protection, Constitution, gay rights, **gay marriage** , gays in the military, DADT, same sex marriage
file409636... they pass a US federal amendment to ban **gay marriage** . Oh, then Christian marriage, must be in
file409636... 31;1-3). Ok, so say Caesar's law banning **gay marriage** is passed (fat chance: it's unconstitutional
file409636... due to marriage, so when states which ban **gay marriage** refuse to grant those benefits, the case
file409636... heterosexual marriage 'compete' freely with **gay marriage** , so people won't mistake the prohibition
file409636... the Mosaic Law. By crusading on abortion, **gay marriage** and other religious issues, this separation
file409637... Massachusetts, have passed laws allowing **gay marriage** . Other states allow civil unions for same-sex
file409637... officer trying to harm him. On the issue of **gay marriage** , my state of Arkansas had passed Amendment
file409637... substantially similar to marital status" **gay marriage** . I have nothing against gay marriage, however
file409637... status" gay marriage. I have nothing against **gay marriage** , however the Supreme Court is overstepping
file409637... states should be able to choose if they allow **gay marriages** or not, because not everyone agrees with
file409637... because not everyone agrees with having **gay marriages** . 4/12/2016 new milford ct M
file409637... states should be able to choose if they allow **gay marriages** or not, because not everyone agrees with

Page 1 of 3 Go Next Last

Figura 25 – Frequência de *gay marriage* no *subcorpus* comparável em inglês
Fonte: Printscreen da tela do *Sketch Engine*.

Sketch Engine casamento gay Linguagem Gay 15-01-2017

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Corpus info
Manage corpus
My jobs
User guide

Save
Make subcorpus
View options
KWIC
Sentence
Sort
Left
Right
Node
References
Shuffle
Sample
Filter
Sub-hits
1st hit in doc
Frequency
Node tags
Node forms
Doc IDs
Collections

Query **casamento, gay** 27 (61.83 per million)

Page 1 of 2 Go Next Last

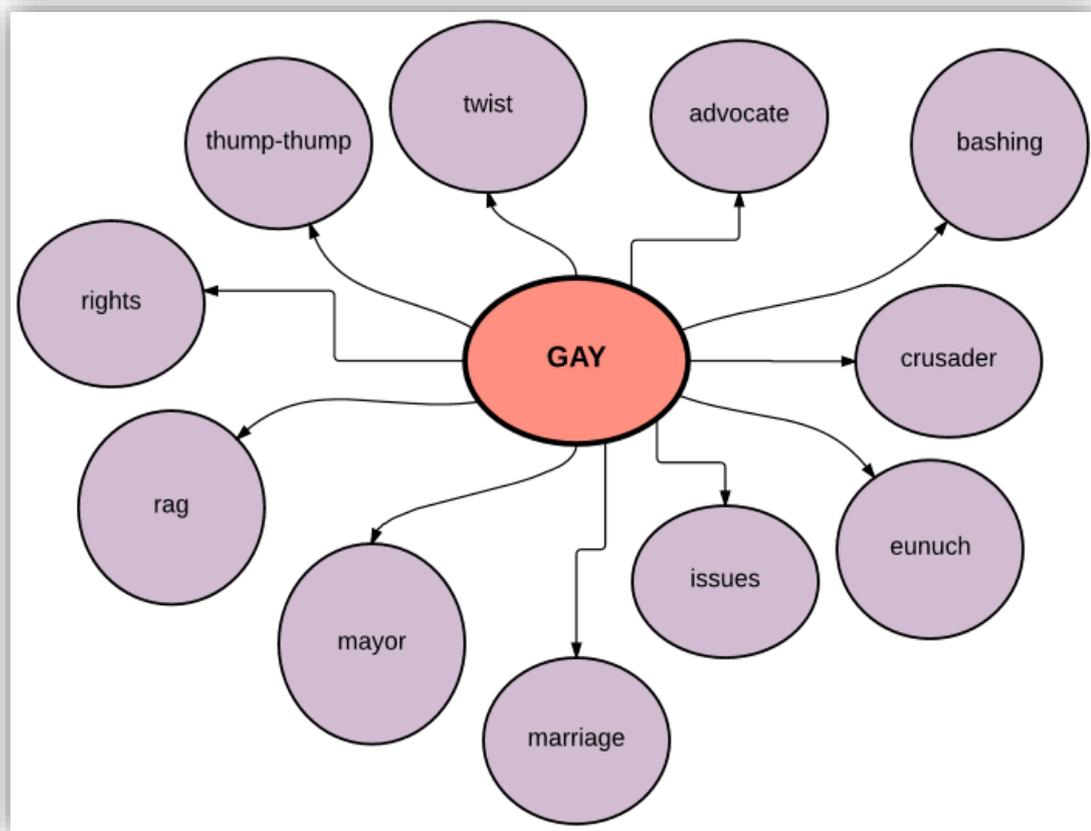
file409642... estamos especialmente mordazes. "E o **casamento gay** ? Também não aprovo." A Bicha Cobarde
file409642... coisas como na TV Guia em 2007: "E o **Casamento gay** ? Também não aprovo. Sou Católico
file409642... Hoje estamos especialmente mordazes. "E o **casamento gay** ? Também não aprovo." A Bicha Cobarde deste
file409646... Ver notícia 23 Outubro - Novo governo tem **casamento gay** na agenda Ver notícia 16 Outubro - Bloco
file409646... Constituição não obriga a reconhecimento de **casamentos gays** em Portugal Ver notícia 29 Julho - Programa
file409643... comente referido como casamento homossexual , **casamento gay** ou casamento homoafetivo) é o casamento
file409643... Argentina [51] Manifestação de apoio ao **casamento gay** em Paris em janeiro de 2013 Em seguida
file409643... parlamento francês votou a legalização do **casamento gay** . [54] No dia 23 de abril de 2013 a lei
file409643... Felicidades..." Quando tirarem o "gay" do **casamento gay** eu acredito que o preconceito tá acabando
file409644... a sua legalização em diversos países, os **casamentos gay** vieram confirmar uniões civis há muito
file409643... Paraíso. Entenda as diferenças entre o **casamento gay** dos EUA e o do Brasil Do UOL, em
file409643... derrubar os vetos de Estados contrários ao **casamento gay** , o que na prática legalizou a união entre
file409643... Procuradores-gerais de alguns Estados que eram contra o **casamento gay** , como Mississippi e Louisiana, já informaram
file409643... juizes da Suprema Corte que votaram contra o **casamento gay** , também disse no texto final que a questão
file409643... uma mulher poderão continuar a se opor ao **casamento gay** , com base na liberdade religiosa garantida
file409644... desculpas. Suprema Corte dos EUA aprova o **casamento gay** em todo o país Estados não poderão
file409644... outros quatro estados. Opositores do **casamento gay** protestam nesta sexta-feira em Washington
file409644... Corte anulou parte da lei federal contra o **casamento gay** , que negava uma série de benefícios governamentais
file409644... (Foto: Eric Gay/AP) Defensores do **casamento gay** comemoram em frente à Suprema Corte em
file409644... Nova York, após o anúncio da liberação do **casamento gay** em todo o país (Foto: Timothy A. Clary/

Page 1 of 2 Go Next Last

Figura 26 – Frequência de “casamento gay” no *subcorpus* comparável em português
Fonte: Printscreen da tela do *Sketch Engine*.

Na figura 25, nas três primeiras linhas de concordâncias da colocação *gay marriage*, distingue-se a questão da legalização do casamento entre pessoas do mesmo sexo, o direito constitucional de igualdade e a presença dos *gays* nos mais variados segmentos da sociedade, enquanto, na figura 26, podemos constatar um discurso mais individual, que reprova as escolhas do outro, a não aceitação do direito de igualdade. É intrigante a forma pela qual o mesmo assunto é discutido: de um lado temos uma discussão pautada no direito de igualdade, de outro, a não aceitação das escolhas do outro.

Outra forma de visualização das palavras que apresentam maior chavidade com bases de nosso estudo, é o organograma 1, que possibilita uma melhor visualização das colocações, como podemos ver logo abaixo:



Organograma 1 – Colocações formadas pela base *gay*.

Fonte: Elaborado pelo autor.

Vemos por meio do organograma 1 os itens lexicais que estão ligados a *gay* e, assim, formam colocações.

Os quadros 7 e 8, abaixo, mostram as colocações extraídas da palavra de busca *fag*, encontrada em nosso *corpus* paralelo, que formam colocações nominais (substantivo + substantivo) e adjetivas (adjetivo + substantivo), em inglês.

<i>Fag</i> (substantivo) + Substantivo	
<i>Fag hag</i> <i>Nice to see you back on the job. As Liberty Avenue's leading fag hag?</i>	Líder gay Bom vê-la de volta ao trabalho. Como a líder gay da Av. Liberty?
<i>Fag hater</i> <i>What more do you need? Someone who's not a homophobe. Because hem ade a couple of cracks? That doesn't mean he's a fag-hater.</i>	Odiar viado De que mais precisa? De alguém que não seja homofóbico. Por que ele disse algumas besteiras? Não significa que ele odeie veados .
<i>Stepford fag</i> <i>I'm a Stepford fag.</i>	Bicha pra casar* Eu sou uma bicha pra casar .

Quadro 7 – Colocações substantivas a partir do colocado *fag*

O item lexical *fag* ora aparece como colocado, ora aparece como base na formação das colocações. Consideramos que, mesmo apresentando duas formas de classificação, tais combinatórias com *fag* apresentam grande riqueza ao nosso estudo. Em *fag hag* e *fag hater*, o item lexical *fag* ocorre como colocado. Já, em relação a *stepford fag*, *fag* ocorre como base.

<i>Fag</i> (Adjetivo) + Substantivo	
<i>Fag jokes</i> <i>You know, just because you found a body on Liberty Avenue, doesn't give you the right to make fag jokes.</i>	Piadas de bichas Só porque você encontrou um corpo na Avenida Liberty não tem o direito de fazer piadas de gays .

Quadro 8 – Colocações adjetivas a partir do colocado *fag*

Em relação à colocação *fag jokes*, destacamos que *fag* trata-se do colocado, enquanto *jokes* é a base.

Adjetivo + <i>Fag</i> (Substantivo)	
<i>Goddamn fag</i> <i>He's a fag like you said, a goddamn fag.</i>	Bicha maldita* Como você disse, ele é uma bicha maldita .
<i>Godless fag</i> <i>Tell them I'm a depraved monster. A godless fag, right, mom?</i>	Viado ateu Diga-lhe que eu sou um monstro depravado. E um viado ateu . Certo, mãe?
<i>Professional fag</i> <i>I told you - I'm a professional football player, not a professional fag!</i>	Bicha profissional Eu te disse, sou um jogador de futebol profissional, não uma bicha profissional!

Quadro 9 – Colocações adjetivas a partir da base *fag*

As colocações, *goddamn fag*, *godless fag* e *professional fag*, presentes no quadro 9, apresentam *fag* como base e os demais itens como colocados, como podemos ver.

Pudemos observar acima as colocações nominais e adjetivas, conforme a tipologia apresentada por Hausmann (1985) e Orenha-Ottaiano (2009, p. 41-42). Neste caso, encontramos colocações formadas por substantivo (colocado) + substantivo (base), como em *fag jokes*, assim como por adjetivo (colocado) + substantivo (base), em inglês, em *Goddamn fag*.

Daremos prosseguimento à nossa análise de dados por meio da tabela 3, na qual discorreremos sobre a colocação adjetiva *fag hag*, extraída do *subcorpus* paralelo em inglês, e de sua tradução para o português *líder gay*, retirada do *subcorpus* paralelo em Português.

Tabela 3 – Estatísticas da colocação *fag hag* e sua tradução para o português
Adjetivo + *Fag* (Substantivo)
Fag hag - Líder gay

1 – <i>Subcorpus</i> Paralelo em Inglês	2 – <i>Subcorpus</i> Paralelo em Português	3 - <i>Subcorpus</i> Comparável em Inglês	4 – <i>Subcorpus</i> Comparável em Português
04	03	80	02

Observando a tabela 3, concluímos que a frequência para a referida colocação no *Corpus* paralelo, *subcorpus* de estudo em inglês (item 1, da tabela 3), e no *corpus* paralelo, *subcorpus* de estudo em português do Brasil (item 2, da tabela 3) é bem próxima, 04 e 03 ocorrências respectivamente.

Já em nosso *corpus* comparável, *subcorpus* comparável em inglês, notamos que a frequência é bem maior (80 ocorrências) em relação à sua tradução para o português (02 ocorrências). Durante a análise das linhas de concordância, observamos que a colocação em inglês é empregada em contextos discursivos da comunidade *gay*.

Veremos, abaixo, a frequência da colocação analisada em nossos *corpora* comparável, nas Figuras 27 e 28.

The screenshot displays the Sketch Engine search results for the query "Fag, hag". The search results are as follows:

File ID	Text Snippet
file409637...	very proud to have her as a hag. And a true fag hag isn't ashamed to admit who she is. </p><p>
file409637...	his hag and she loves her fag. And that a fag hag </p><p> The Urban Dictionary Mug </p><p> The
file409637...	</p><p> The Urban Dictionary T-Shirt </p><p> A fag hag is a woman who prefers the company of gay
file409637...	prefers the company of gay men. </p><p> A fag hag is a woman who prefers the company of gay
file409637...	28, 2007 . Retrieved 2007-02-02 . </p><p> Fag hag </p><p> Fag hag is a gay slang phrase referring
file409637...	Retrieved 2007-02-02 . </p><p> Fag hag </p><p> Fag hag is a gay slang phrase referring to a woman
file409637...	associate with gay men object to being called fag hags while others embrace the term. The male
file409637...	officially a member. [2] </p><p> Contents </p><p> Fag hags are frequently stereotyped as outgoing
file409637...	[4] In fact, many women who identify as fag hags are already in romantic relationships,
file409637...	the straight counterpart. </p><p> American fag hag synonyms include fruit fly , [5] queen
file409637...	Mile and Looney Tunes) has been coined for fag hags who are attractive or have boyfriends,
file409637...	boyfriends, to counteract the stereotype that fag hags are unable to find a suitable straight
file409637...), writer Christopher Gore used the term fag hag in describing the character Doris in her
file409637...	talks in her stand-up routines about being a fag hag . In an episode of the UK TV sitcom Gimme
file409637...	Gimme Gimme Gimme, Tom refers to Linda as a fag hag . [8] In the made-for television series
file409637...	Tolliver (Marcus D'Amico) makes her a " fag hag ." The term is also referenced in Grandmaster
file409641...	doi : 10.1080/1035033042000202915 . </p><p> Fag hag </p><p> A fag hag is a slang term for a woman
file409641...	1035033042000202915 . </p><p> Fag hag </p><p> A fag hag is a slang term for a woman who either
file409641...	associate with gay men object to being called fag hags , [2] while others embrace the term. [3]
file409641...	regardless of their sex. </p><p> Contents </p><p> Fag hags are frequently stereotyped as outgoing

Figura 27 – Frequência de *fag hag* no *subcorpus* comparável em inglês
Fonte: Printscreen da tela do *Sketch Engine*.

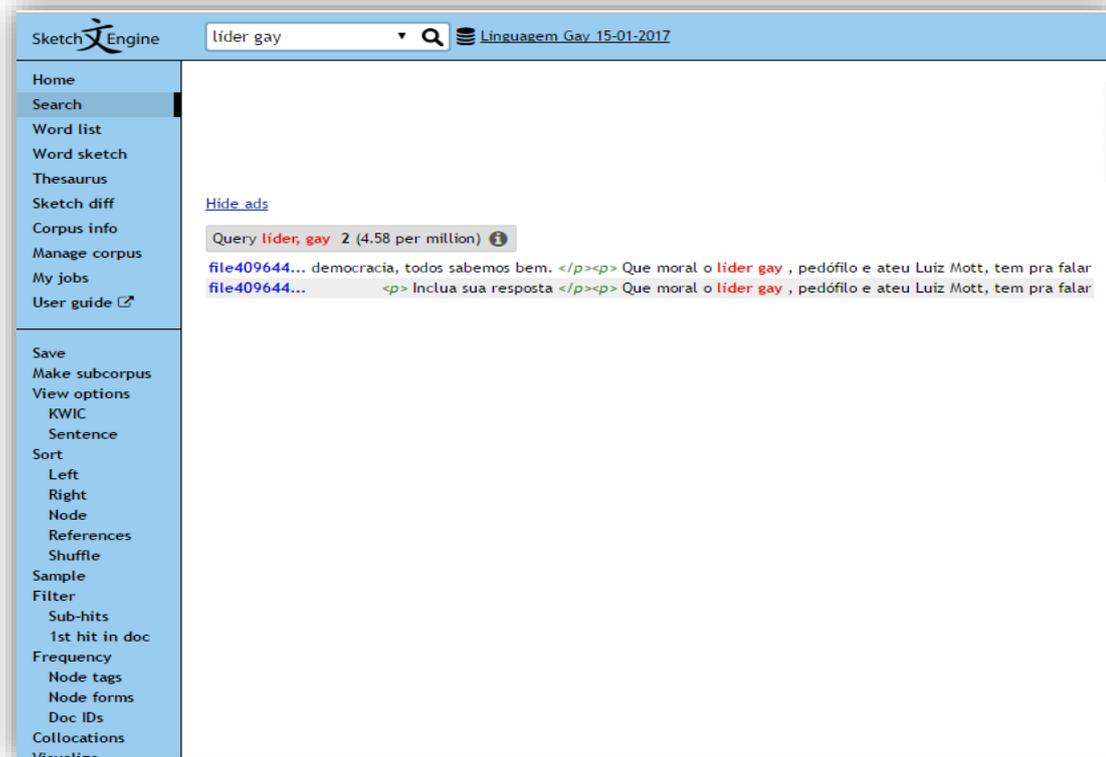
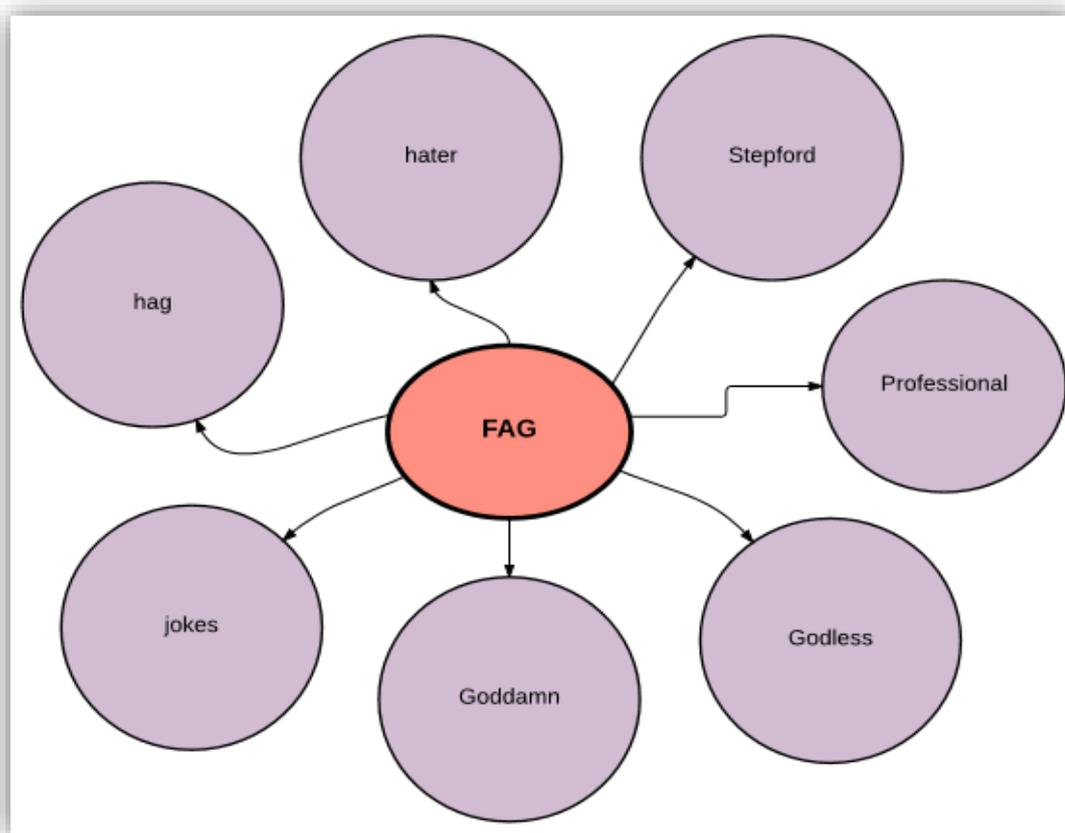


Figura 28 – Frequência de “líder gay” no *subcorpus* comparável em português

Fonte: *Printscreen* da tela do *Sketch Engine*.

Vemos nas figuras 27 e 28 a frequência e as linhas de concordância da colocação *fag hag*, observando que “líder gay” apresenta uma denotação extremamente negativa, associada à “falta de moral”, “pedofilia” e “ateu”, itens que, de modo geral, são vistos como negativos.

Visando à demonstração das palavras que apresentam maior chavicidade com bases de nosso estudo, o organograma 2 apresenta a base *fag* e em seu entorno os colocados. Vejamos a seguir:



Organograma 2 – Colocações formadas pela base *fag*.

Fonte: Elaborado pelo autor.

Encontramos, no organograma 2, os itens lexicais de maior chavicidade em nossa pesquisa que formam colocações com *fag*.

Dando sequência, temos o quadro 9 logo abaixo, no qual encontramos colocações adjetivas formadas pela base *faggot*. No caso desta palavra de busca, verificamos que ocorre somente como base e não como colocado.

Adjetivo + <i>Faggot</i> (Substantivo)	
<i>Brave faggot</i> <i>You stood beside your partner, no matter what anyone said, and that makes you a very loyal, very brave faggot.</i>	Viado corajoso Você apoiou seu companheiro e independente do que os outros digam isso o torna um viado corajoso e muito leal.
<i>Fucking faggot</i> <i>Not before you suck on this, you fucking faggot.</i>	Bicha de merda Não antes de você chupar aqui, sua bicha de merda .
<i>Little faggot</i> <i>Oh, you fucking... dirty little faggot!</i>	Bichinha Oh, você me paga... sua bichinha suja!
<i>Silly faggot</i> <i>Well, what about your apology for calling Melanie a cunt? I'll apologize for calling her a cunt when she apologizes for calling me a silly faggot.</i>	Bicha idiota* Você não tem de se desculpar com Melanie, por tê-la chamado de cretina? Só se ela se desculpar por ter me chamado de bicha idiota .

Quadro 10 – Colocações adjetivas a partir da base *faggot*

Como podemos ver acima, as colocações do quadro 10 são formadas pela base *faggot*. Prosseguindo, temos a tabela 4 juntamente com a discussão dos dados extraídos de nossos *corpora* de estudo.

Tabela 4 – Estatísticas da colocação *little faggot* e sua tradução para o português

Adjetivo + <i>Faggot</i> (Substantivo)			
<i>Little faggot</i> - Bichinha			
1 – <i>Subcorpus</i> Paralelo em Inglês	2 – <i>Subcorpus</i> Paralelo em Português	3 - <i>Subcorpus</i> Comparável em Inglês	4 – <i>Subcorpus</i> Comparável em Português
06	06	15	71

Durante a análise da colocação *little faggot*, observamos que ela possui a mesma frequência no *corpus* paralelo, tanto no *subcorpus* em inglês, quanto no *subcorpus* em português do Brasil (tabela 4, itens 1 e 2). Entendemos, dessa maneira, que essa colocação foi utilizada no *corpus* paralelo, porém resolvemos conferir sua frequência em nosso *corpus* comparável (tabela 4, item 3). Ela apresenta uma frequência 15 no *subcorpus* comparável em inglês, considerada baixa, levando em conta o tamanho do *subcorpus* em inglês, 1.174.067. Acreditamos, que a frequência de *little faggot* no item 3, apresenta-se baixa devido à utilização de outros itens lexicais para designar uma mesma coisa, como, por exemplo, *faggy*, que é utilizado como *little faggot*.

A seguir, encontramos a frequência da colocação *little faggot*, em inglês, e *bichinha*, em português do Brasil, em nosso *corpus* comparável.

Sketch Engine Little faggot Subcorpus Inglês

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Keywords/terms
Corpus info
Manage corpus
My jobs
User guide

Save
Make subcorpus
View options
KWIC
Sentence
Sort
Left
Right
Node
References

Query **Little, faggot** 15 (10.35 per million) **i**

file540642... I am admitting that my greatest turn on is being a **little faggot** , that I love being nude with another faggot and
file540620... 're to cover manly cocks and balls, not </p> <p> **little faggot** clits. </p> <p> "If you ever see this cuntboy
file540639... and slammed his fat foot long black cock down my **little faggot** throat. I started gagging. </p> <p> "That's
file540639... . </p> <p> "That's right bitch choke. Choke you **little faggot** !" He held his dick in my throat and pumped my face
file540639... <p> Tears were running down my face. "Fuck that **little faggot** mouth." He pushed me off his dick and the dread
file540624... had his ass fucked. Six months of dating this **little faggot** and he had just revealed that little tidbit. It
file540624... in his ear, his back against my chest: "Shh, **little faggot** . It'll only hurt for a minute. Relax your ass.
file540624... into me, taking me deeper. "Nice. That's good, **little faggot** . That's what I wanted. Nice." He moaned and
file540624... , crying out, curling again. </p> <p> "Come on, **little faggot** . You can take it." I pulled his wrists over his
file540643... ..." </p> <p> "Wow, my brother being nice to his ' **little faggot** ? That's so unlike you, Billy." </p> <p> "Well,
file540643... plunged her fingers into her wet fuck-hole. The **little faggot** was going to have get a hard pounding; it was how
file540643... . She had to admit that being mean to the **little faggot** eating out her ass was a real turn on. She had
file540643... no, Billy," Becky said. "I'm resting, have your **little faggot** take care of you." She lay down on the bed, giving
file540643... her brother. She wanted her brother to show the **little faggot** who was boss. She knew it would make her pussy
file540643... best to make him cum quickly. </p> <p> "Hey, your **little faggot** is finally getting into it!" Beck exclaimed.

Figura 29 – Frequência de *little faggot* no *subcorpus* comparável em inglês
Fonte: Printscreen da tela do *Sketch Engine*.

Sketch Engine Subcorpus Português

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Keywords/terms
Corpus info
Manage corpus
My jobs
User guide

Save
Make subcorpus
View options
KWIC
Sentence
Sort
Left
Right
Node
References
Shuffle
Sample
Filter
Sub-hits
1st hit in doc
Frequency
Node tags
Node forms

Query **bichinha** 71 (57.04 per million) ⓘ

Page 1 of 4 [Next](#) | [Last](#)

file540654... de carnaval. </p> <p> Bicha Gentil </p> <p> A **bichinha** se apaixonou loucamente por um português
file540654... sua casa. </p> <p> Toda noite era a mesma coisa. A **bichinha** se produzia toda e ficava desfilando pra lá e pra
file540654... . Até que um belo dia o galego deu bobeira e a **bichinha** colou: </p> <p> - Olá gostosão... tá a fim de
file540654... depois, lá estavam os dois no apartamento da **bichinha** . Primeiro veio o jantar, bem requintado. </p>
file540654... . E o Manoel se sentindo prestigiadíssimo. A **bichinha** fazia de tudo pra agradar. </p> <p> Depois veio o
file540654... . </p> <p> Depois veio o banho: primeiro a **bichinha** , depois o galego. Mais um pouco de papo e a
file540654... , depois o galego. Mais um pouco de papo e a **bichinha** achou que estava na hora do "vamos ver": </p> <p>
file540654... ele está? </p> <p> Adivinhe se Puder </p> <p> A **bichinha** chega para a outra com as mãos escondidas atrás
file540654... cremes e mais cremes no rosto, quando a outra **bichinha** , sua amiga, pergunta: - </p> <p> Pra quê tudo
file540666... de banho e de necessidades fisiológicas, ô **bichinha** ruim. </p> <p> Pense num objeto que pra mim não
file540670... tá debaixo do chapéu. </p> <p> A BICHA E O GAUCHOA **bichinha** se apaixonou loucamente por um gaúcho bonito,
file540670... da sua casa. Toda noite era a mesma coisa. A **bichinha** se produzia toda e ficava desfilando pra lá e pra
file540670... . Até que um belo dia o galego deu bobeira e a **bichinha** chegou: - Olá gostosão... tá a fim de conhecer o
file540670... ! Meia hora depois, lá estavam os dois no apê da **bichinha** . Primeiro veio o jantar, bem requintado.
file540670... . E o gaúcho se sentindo prestigiadíssimo. A **bichinha** fazia de tudo pra agradar. Depois veio o banho:
file540670... pra agradar. Depois veio o banho: primeiro a **bichinha** depois o galêgo. Mais um pouco de papo e a
file540670... depois o galêgo. Mais um pouco de papo e a **bichinha** achou que estava na hora do "vamos ver": - Agora
file540666... voltar mas vai voltar como um veado. </p> <p> Uma **bichinha** comprou um fusquinha e ficou toda serelepe,
file540666... ". Logo ela foi chamar a Karlinha - outra **bichinha** - para passear de fusca, e a cada bofe que passava
file540666... Não vou pagar e pronto! Saia da frente! </p> <p> A **bichinha** motorista não se conteve e mandou que Karlinha

Page 1 of 4 [Next](#) | [Last](#)

Figura 30 – Frequência de “bichinha” no *subcorpus* comparável em português

Fonte: *Printscreen* da tela do *Sketch Engine*.

Observamos nas linhas de concordância da figura 29 a colocação *little faggot* e o item lexical *bichinha* na figura 30, uma vez que não forma uma colocação, trata-se de uma colocação de cunho sexual e depreciativa pelos contextos nos quais aparece.

Constatamos, também, que a frequência do item lexical em português do Brasil, *bichinha*, é bem frequente em nosso *subcorpus* comparável em português em relação à frequência da colocação equivalente em inglês, *little faggot*. Atestamos por meio das linhas de frequência seu uso pela comunidade em estudo. A seguir, poderemos encontrar a frequência da colocação *little faggot*, em inglês, e *bichinha*, apenas um item lexical de nosso *corpus* comparável.

Na busca por uma visualização das palavras que apresentam maior chavicidade com bases de nosso estudo, o organograma 3 apresenta a base *faggot* e, em seu entorno, os colocados que apresentam maior chavicidade para a formação das colocações utilizadas pela comunidade homossexual. Vejamos a seguir:



Organograma 3 – Colocações formadas pela base *faggot*.
Fonte: Elaborado pelo autor.

No quadro 11, temos colocações nominais formadas pela base *butt*. Também neste caso, distinguimos que a palavra em questão ocorre tanto como base como colocado. Temos, no quadro 11, as colocações nominais na qual *butt* é base.

<i>Butt</i> (substantivo) + Substantivo	
<p><i>Butt plug</i></p> <p><i>That's my lube, my butt plug, my poppers, latest issue of 'Star'.</i></p>	<p>Vibrador de rabo*</p> <p>Meu lubrificante, meu vibrador de rabo, meu poppers, uma revista.</p>
<p><i>Butt check</i></p> <p><i>I haven't had a chance to do a butt-check. Besides, David and I have better ways of spending our evenings.</i></p>	<p>Check-up de bundas</p> <p>Eu não tive a oportunidade de fazer um check-up de bundas. E por outro lado, eu e o David temos formas melhores de passar as tardes.</p>
<p><i>Butt fucked</i></p> <p><i>As far as the ass business, getting your butt fucked is one of the great pleasures, and privileges of being gay.</i></p>	<p>Dar o rabo</p> <p>Quanto à parte da transa, um dos privilégios e grandes prazeres de ser gay, é dar o rabo.</p>
<p><i>Butt fucking</i></p> <p><i>And what about all this sex? Blow-jobs and butt-fucking on every page.</i></p>	<p>Enrabada*</p> <p>E todo esse sexo? Boquetes e enrabada em cada página.</p>

Quadro 11 – Colocações substantivas a partir do colocado *butt*.

Vimos no quadro 11 que as colocações: *butt plug*, *butt check*, *butt fucked* e *butt fucking* são formadas pela base *butt*. Também chamamos atenção para a colocação *butt fucking*, que não forma uma colocação em português, tratando-se apenas de um item lexical.

Daremos prosseguimento à análise de dados, por meio da tabela 5, na qual discorreremos sobre a colocação substantiva *butt fucking*, extraída do *subcorpus* paralelo em inglês, e de sua tradução para o português do Brasil “enrabada”, retirada do *subcorpus* paralelo em português.

Tabela 5 – Estatísticas da colocação *butt fucking*
Butt (substantivo) + Substantivo
Butt fucking – enrabada

1 – <i>Subcorpus</i> Paralelo em Inglês	2 – <i>Subcorpus</i> Paralelo em Português	3 – <i>Subcorpus</i> Comparável em Inglês	4 – <i>Subcorpus</i> Comparável em Português
03	03	04	02

Observando a tabela acima, concluímos que a frequência para a referida colocação no *corpus* paralelo, *subcorpus* de estudo em inglês (item 1, da tabela 5), e no *corpus* paralelo, *subcorpus* de estudo em português (item 2, da tabela 5), são iguais.

Já em nosso *corpus* comparável, *subcorpus* comparável em português, notamos que a frequência é de 02 ocorrências e, durante a análise das linhas de concordância, observamos que essa colocação é empregada em contextos discursivos da comunidade homossexual. Por outro lado, nosso *subcorpus* comparável em inglês apresenta apenas 04 ocorrências.

Vemos em seguida, a frequência e o contexto de uso da colocação analisada em nossos *corpora* comparáveis, nas figuras 31 e 32.

Sketch Engine Subcorpus Inglês

- Home
- Search
- Word list
- Word sketch
- Thesaurus
- Sketch diff
- Keywords/terms
- Corpus info
- Manage corpus
- My jobs
- User guide

Save
Make subcorpus

Query **butt, fucking** 4 (2.76 per million) ⓘ

file540624... and the hulk like William Ramon go for a raw **butt fucking** romp in 'Raw Latin Fuckers'. With their
 file540620... had never fucked a sissy baby before. Really? **Butt fucking** her while watching her suck cock is a turn on?
 file540620... different type of stimulation. Okay, okay. A **butt fucking** , cock sucking, cum swallowing party with a
 file540620... group sex, movies, xhamster, bitch, latine, **butt fucking** </p> <p> FOR CANDID LOVERS ONLY! </p> <p> 10

Figura 31 – Frequência de *butt fucking* no *subcorpus* comparável em inglês

Fonte: Printscreen da tela do Sketch Engine.

Sketch Engine Subcorpus Português

- Home
- Search
- Word list
- Word sketch
- Thesaurus
- Sketch diff
- Keywords/terms
- Corpus info
- Manage corpus
- My jobs
- User guide

Save
Make subcorpus
View options
KWIC
Sentence

Query **enrabada** 2 (1.61 per million) ⓘ

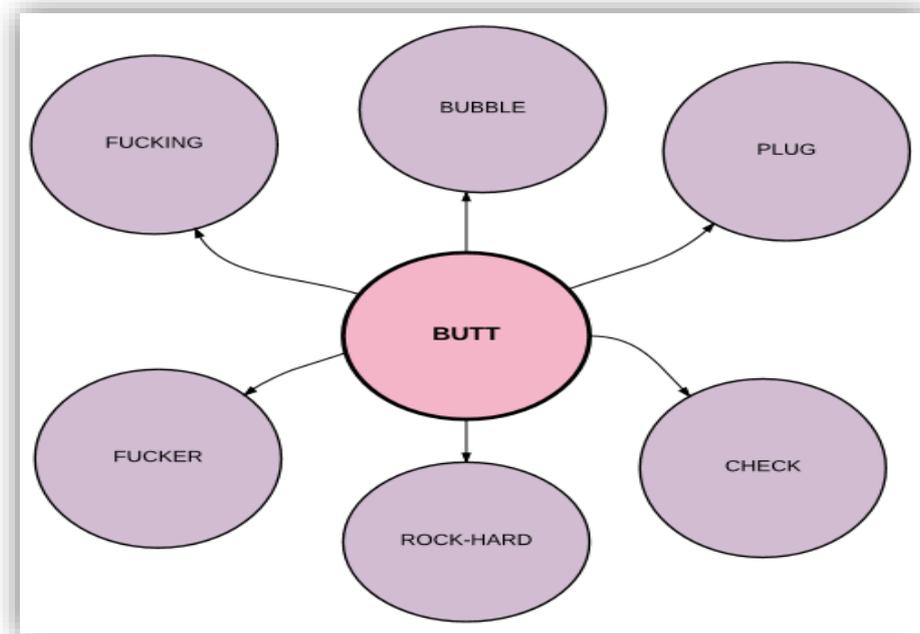
file540655... pra ver como ela ficou ao ver,adoro vê-la sendo **enrabada** por uma jeba assim.,adivinha se também não caio
 file540674... . Eu chupava e me contorcia como podia, sendo **enrabada** por um, e com a boca cheia de pica do outro. Eu

Figura 32 – Frequência de “enrabada” no *subcorpus* comparável em português

Fonte: Printscreen da tela do Sketch Engine.

Tanto nas linhas de concordância da figura 31, quanto nas linhas de concordância da figura 32, atentamos para a colocação *butt fucking* que apresenta conotação sexual e o item lexical “enrabada” também.

Outra forma de observação das palavras que apresentam maior chavidade com bases de nosso estudo é o organograma 4, que garante uma visualização fácil das colocações, como vemos abaixo:



Organograma 4 – Colocações formadas pela base *butt*.
Fonte: Elaborado pelo autor.

O quadro 12, logo em seguida, mostra as colocações extraídas da palavra de busca *ass*, encontrada em nosso *corpus* paralelo, que formam colocações adjetivas (adjetivo + substantivo), em inglês.

Adjetivo + ass (Substantivo)	
Smart ass <i>I would like an honest answer, smart ass.</i>	Espertinho Eu gostaria de uma resposta honesta, espertinho .
Sore ass <i>You get a sore ass?</i>	Bunda dolorida Você ganha uma bunda dolorida ?
Tight ass <i>If it doesn't have a tight ass and at least nine inches, I'd be surprised.</i>	Traseiro apertado Se não tiver um traseiro apertado e pelo menos 22 cm, ficarei surpreso.
Wise ass <i>What's that supposed to be, some wise-ass comment?</i>	Espertalhão* O que isso quer dizer, alguma observação espertalhão ?

Quadro 12 – Colocações adjetivas a partir da base *ass*.

Visando à verificação do uso e da frequência das colocações levantadas no *corpus* paralelo em relação ao *corpus* comparável anteriormente mencionados, apresentamos a frequência levantada da colocação *tight ass* no *subcorpus* paralelo do inglês (item 1, da tabela 6 abaixo), no *subcorpus* paralelo do português, ou seja, de sua tradução – “traseiro apertado” (item 2, da tabela 6), frequência do *subcorpus* comparável em inglês (item 3, da tabela 6) e frequência do *subcorpus* comparável em português (item 4, da tabela 6).

Notamos que a colocação *wise ass* em inglês não forma uma colocação em português, trata-se apenas de um item lexical, “espertalhão”. Além do mais, salientamos que questionamentos sobre as traduções das legendas não foram levantados, já que este não é o objetivo de nossa pesquisa.

Tabela 6 – Estatísticas da colocação *tight ass*.

Tight (adjetivo) + Substantivo			
Tight ass – Traseiro apertado			
1 – Subcorpus Paralelo em Inglês	2 – Subcorpus Paralelo em Português	3 – Subcorpus Comparável em Inglês	4 – Subcorpus Comparável em Português
08	08	05	0

A colocação adjetiva extraída de nosso *corpus* paralelo, como descrito na tabela 6 e exemplificada na figura 33, estrutura-se sintagmaticamente em *adj. + noun*, em inglês. É formada pelo adjetivo em função atributiva (QUIRK *et al.*, 1985), isto é, modificando o substantivo *ass*.

A frequência de tal colocação em nosso *corpus* paralelo do inglês é 08 e, no *subcorpus* comparável em inglês, aparece com 05 ocorrências. No entanto, a opção tradutória, *traseiro apertado*, para a colocação *tight ass*, não ocorre no *subcorpus* comparável em português. Acreditamos que a colocação “traseiro apertado” não apresenta frequência em nosso *corpus* comparável em português, porque em português provavelmente diríamos “rabo apertado” ou “cu apertado”. Lembramos que “traseiro apertado” trata-se da tradução da série.

A seguir, temos a figura 33 que comprova a frequência de nossa colocação em nosso *corpus* comparável, especificamente no *subcorpus* inglês, compilado para nossa pesquisa. Destacamos a ausência da figura com a colocação em português, uma vez que a quantidade de ocorrências é zero.

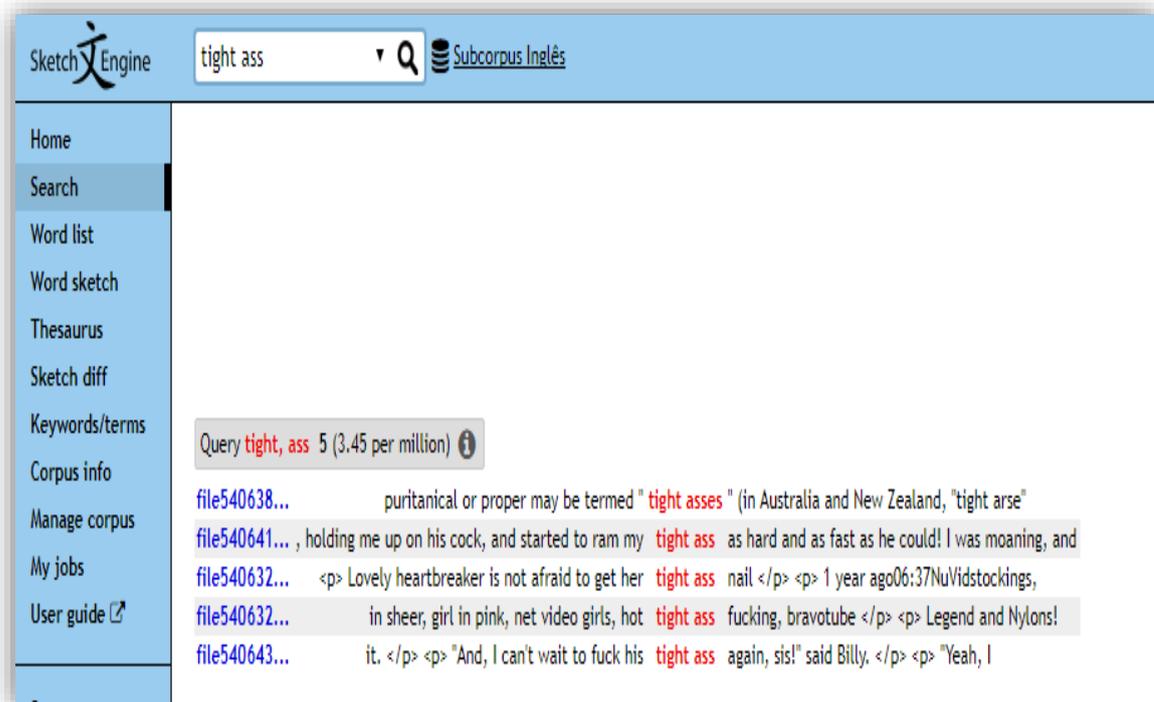
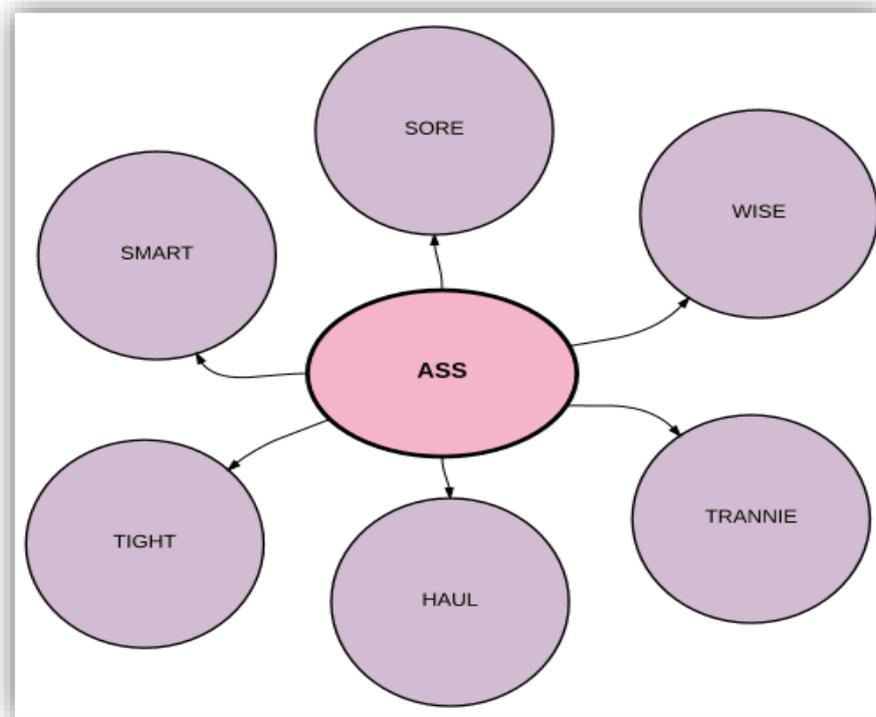


Figura 33 – Frequência de *tight ass* no *subcorpus* comparável em inglês
Fonte: *Printscreen* da tela do *Sketch Engine*.

Vemos nas linhas de concordância, presentes na figura 33, que a colocação *tight ass*, de cunho sexual, está relacionada ao ato sexual, momento da relação.

Na busca por uma visualização das palavras que apresentam maior chavicidade com bases de nosso estudo, o organograma 5 apresenta a base *ass* no centro e, em seu entorno, os colocados que apresentam maior chavicidade para formação das colocações utilizadas pela comunidade homossexual.

Chamamos a atenção para os organogramas, já que nem todas os itens lexicais aparecem durante a análise. Optamos por acrescenta-los nos respectivos organogramas, porque esses itens aparecem em nossa amostra de Glossário. Vejamos a seguir o organograma 5:



Organograma 5 – Colocações formadas pela base *ass*.
Fonte: Elaborado pelo autor.

Para finalizar, seguimos para a amostra do *Glossário Bilíngue de Colocações da Comunidade Homossexual* e, posteriormente, seguimos para as considerações finais com uma discussão sobre os resultados de nossa pesquisa e sua contribuição para as áreas aqui enfocadas.

5 GLOSSÁRIO DE COLOCAÇÕES DA COMUNIDADE HOMOSSEXUAL

Dedicamos este capítulo à apresentação de um resultado prático de nossa pesquisa: uma amostragem do *Glossário Bilingue de Colocações da Comunidade Homossexual*, elaborado com base no *corpus* paralelo formado pelas cinco temporadas da série *Queer as Folk* e analisado por meio de nosso *corpus* comparável, compilado exclusivamente para este estudo, que segue na direção inglês→português.

5.1 Tipologia do Glossário

No que se refere à tipologia, o glossário compilado é bilíngue, na direção inglês→português, sendo destacado que nosso foco de classificação está voltado para as colocações em inglês, já que em alguns casos não há correspondência sintática total quando traduzida para o português do Brasil.

O público-alvo pretendido é bastante abrangente, conforme se pode observar na figura 34:

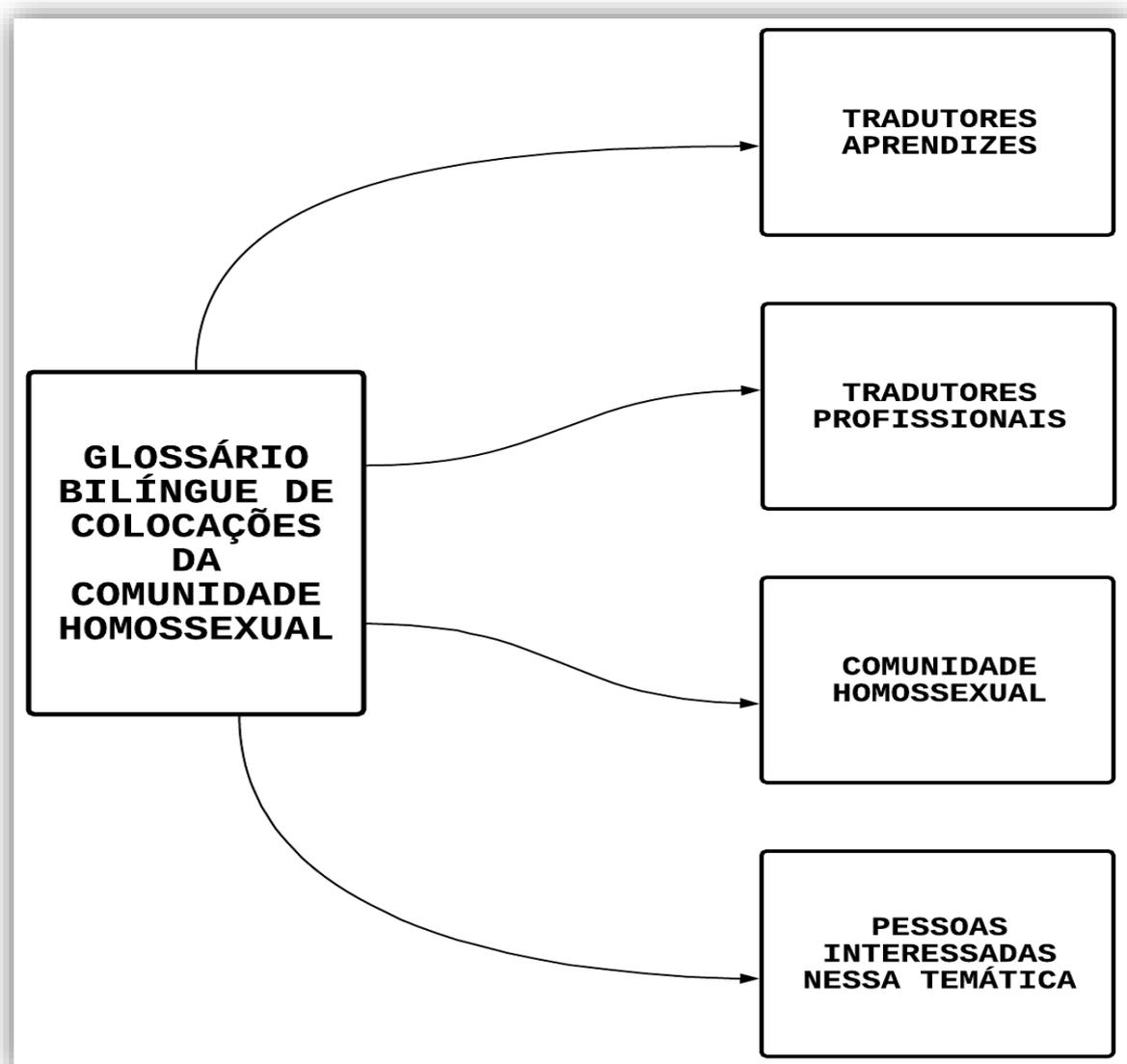


Figura 34 – Público-alvo do Glossário.

Fonte: Elaborado pelo autor.

5.1.1 Macro e Microestrutura

O glossário proposto para esta pesquisa é relativamente amplo, já que pretendemos abarcar o maior número possível de colocações da comunidade homossexual. Entretanto, não é exaustivo, tendo em vista que, no trabalho com linguagem, sabemos que essa pretensão seria incoerente.

Ressaltamos que a escolha das colocações que compõem o glossário não foi realizada de forma aleatória, posto que foram selecionadas pela frequência e/ou pela chavicidade, com o auxílio de programas que podem oferecer maior confiabilidade nos resultados, além do embasamento teórico que envolveu a seleção dos itens lexicais.

Para Orenha-Ottaiano (2004, p. 144), os critérios de seleção dos itens lexicais são “definidos pelo próprio lexicógrafo/terminógrafo, no intuito de atender a seus objetivos e de forma a organizar a obra no sentido de facilitar a busca dos termos por parte de seu consulente”, critérios estes que acreditamos estarem presentes nesta pesquisa.

Para Béjoint (1994, p 11), “a natureza estrutural da macroestrutura é menos clara do que a da microestrutura”, tendo em vista a possibilidade de adição ou mesmo de subtração de uma entrada sem destruir a qualidade da lista de palavras, cuja unidade não pode ser definida de forma clara. Sendo assim, acreditamos que futuramente possamos ampliar nosso corpus de estudo, caso seja necessário, selecionar mais colocações e inseri-las no *Glossário Bilingue de Colocações da Comunidade Homossexual*.

Ainda na esfera da macroestrutura, as colocações estão dispostas em ordem alfabética, para que o consulente tenha acesso a elas de maneira mais fácil e rápida. Evidenciamos que neste glossário foram incluídas apenas colocações, excluindo-se, assim, outros fraseologismos.

No que diz respeito à microestrutura, Orenha-Ottaiano (2004, p. 155) afirma que ela “é o modo como as entradas são organizadas” em uma obra fraseológica, em nosso caso, sendo considerada uma estrutura bastante fixa e rígida. Segundo Rey Debove (1971), a microestrutura é um conjunto de informações ordenadas que seguem a entrada, caracterizado por estrutura constante, correspondente a um programa e a um código aplicável a qualquer entrada, caso contrário, os usuários nunca saberão o que esperar ao procurar uma palavra.

A microestrutura é composta por todas as informações relativas à entrada, como, por exemplo, a classificação gramatical, definição. Por esta razão, acreditamos que a microestrutura representa um conjunto de informações ordenados e estruturados.

Barbosa (1989, p. 571) propõe um modelo de microestrutura, ampliado a partir da estrutura mínima proposta por Vilela (1983, *apud* BARBOSA 1989, p. 571): um paradigma informacional (PI), um paradigma definicional (PD) e um paradigma pragmático (PP) e que esses podem se subdividir dependendo, dentre outras coisas, do objetivo da obra lexicográfica, em nosso caso, fraseográfica.

O PI é constituído pelas informações gramaticais e fônicas, o PD é formado pela definição e o PP é formado pelos exemplos de uso ou abonações, informações culturais, etc. Utilizamos o paradigma pragmático nesta pesquisa. Após essas considerações, apresentamos a seguir uma amostra do glossário.

5.2 Guia Visual

Com o objetivo de apresentar como as entradas estão estruturadas no glossário, apresentamos abaixo um **Guia Visual** da referida obra, seguindo o modelo de guia visual proposto por Orenha-Ottaiano (2004, p. 159):

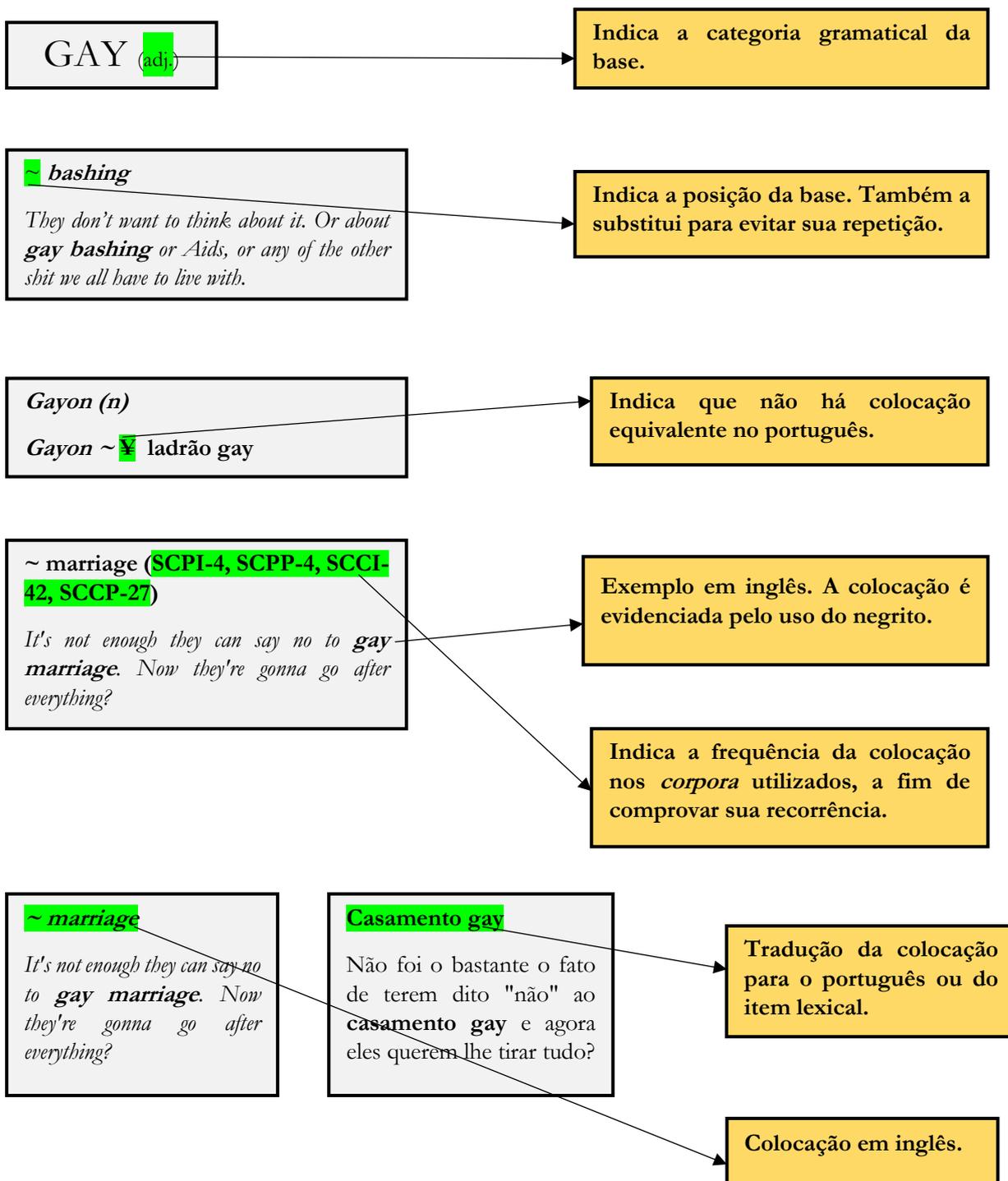


Figura 35 – Guia Visual do Glossário
 Fonte: Adaptado de Orenha-Ottaiano (2004, p. 159)

5.3 Amostra do Glossário

Apresentamos uma amostragem do *Glossário Bilingue de Colocações da Comunidade Homossexual*. As unidades lexicais escolhidas para compor o glossário foram: *ass*, *butt*, *fag*, *faggot* e *gay*, dada sua alta frequência na linguagem da comunidade homossexual.

Seguem as abreviaturas e símbolos utilizados no glossário e, posteriormente, a amostra de *Glossário Bilíngue de Colocações da Comunidade Homossexual*:

- ~ —————→ base/o item pesquisado
- alg** —————→ alguém
- adj** —————→ adjetivo/ adjective
- adv** —————→ advérbio/adverb
- n** —————→ noun (= substantivo)
- poss** —————→ possessivo
- pred** —————→ informa que o adjetivo pode ser usado de forma predicativa
- s** —————→ substantivo
- sb** —————→ somebody (= alguém)
- sth** —————→ something (= algo)
- U** —————→ uncountable (= incontável)
- v** —————→ verbo / verb
- adv** —————→ advérbio / adverb
- ¥** —————→ Indica que não há colocação equivalente em português
- SCPI** —————→ *Subcorpus* paralelo inglês
- SCPP** —————→ *Subcorpus* paralelo português
- *** —————→ Nossa sugestão de tradução

Em seguida, apresentamos a amostra do glossário. Posterior à amostra, temos o capítulo com as considerações finais.

**GLOSSÁRIO BILÍNGUE
DE COLOCAÇÕES DA
COMUNIDADE
HOMOSSEXUAL**



Adjetivo + ass (substantivo)

<p>Smart ~ (SCPI-10) <i>I would like an honest answer, smart ass.</i></p>	<p>Espertinho ¥ (SCPP-3) Eu gostaria de uma resposta honesta, espertinho.</p>
<p>Sore ~ (SCPI-2) <i>You get a sore ass?</i></p>	<p>Bunda dolorida (SCPP-1) Você ganha uma bunda dolorida?</p>
<p>Tight ~ (SCPI-8) <i>If it doesn't have a tight ass and at least nine inches, I'd be surprised.</i></p>	<p>Traseiro apertado (SCPP-2) Se não tiver um traseiro apertado e pelo menos 22 cm, ficarei surpreso.</p>
<p>Wise ~ (SCPI-3) <i>What's that supposed to be, some wise ass comment?</i></p>	<p>Espertalhão* ¥ (SCPP-0) O que isso quer dizer, alguma observação espertalhona?</p>

Butt (substantivo) + Substantivo

<p style="text-align: center;">~ plug (SCPI-3)</p> <p><i>That's my lube, my butt plug, my poppers, latest issue of 'Star'.</i></p>	<p style="text-align: center;">Vibrador de rabo* (SCPP-3)</p> <p>Meu lubrificante, meu vibrador de rabo, meu poppers, uma revista.</p>
<p style="text-align: center;">~ check (SCPI-1)</p> <p><i>I haven't had a chance to do a butt check. Besides, David and I have better ways of spending our evenings.</i></p>	<p style="text-align: center;">Check-up de bunda (SCPP-1)</p> <p>Eu não tive a oportunidade de fazer um check-up de bunda. E por outro lado, eu e o David temos formas melhores de passar as tardes.</p>
<p style="text-align: center;">~ fucked (SCPI-1)</p> <p><i>As far as the ass business, getting your butt fucked is one of the great pleasures, and privileges of being gay.</i></p>	<p style="text-align: center;">Dar o rabo* (SCPP-1)</p> <p>Quanto à parte da transa, um dos privilégios e grandes prazeres de ser gay, é dar o rabo.</p>
<p style="text-align: center;">~ fucking (SCPI-3)</p> <p><i>And what about all this sex? Blow-jobs and butt fucking on every page.</i></p>	<p style="text-align: center;">Enrabada* ¥ (SCPP-2)</p> <p>E todo esse sexo? Boquetes e enrabada em cada página.</p>

***Fag* (Adjetivo) + Substantivo**

Fag ~ (SCPI – 2)

*You know, just because you found a body on
Liberty Avenue, doesn't give you the right to make
fag jokes.*

Piadas de bichas (SCPP – 2)

Só porque você encontrou um corpo na
Avenida Liberty não tem o direito de fazer
piadas de gays.

Fag (substantivo) + Substantivo

<p style="text-align: center;">~ hag (SCPI – 4)</p> <p style="text-align: center;"><i>Nice to see you back on the job. As Liberty Avenue's leading fag hag?</i></p>	<p style="text-align: center;">Líder gay (SCPP – 4)</p> <p style="text-align: center;">Bom vê-la de volta ao trabalho. Como a líder gay da Av. Liberty?</p>
<p style="text-align: center;">~ hater (SCPI-1)</p> <p style="text-align: center;"><i>What more do you need? Someone who's not a homophobe. Because hem ade a couple of cracks? That doesn't mean he's a fag-hater.</i></p>	<p style="text-align: center;">Odiar viado (SCPP-1)</p> <p style="text-align: center;">De que mais precisa? De alguém que não seja homofóbico. Por que ele disse algumas besteiras? Não significa que ele odeie viados.</p>
<p style="text-align: center;">~ fag (SCPI-6)</p> <p style="text-align: center;"><i>I'm a Stepford fag.</i></p>	<p style="text-align: center;">Bicha pra casar* (SCPP-0)</p> <p style="text-align: center;">Eu sou uma bicha pra casar.</p>

Adjetivo + *Fag* (Substantivo)

<p><i>Goddamn</i> ~ (SCPI-2) <i>He's a fag like you said, a goddamn fag.</i></p>	<p>Bicha maldita* (SCPP-0) Como você disse, ele é uma bicha maldita.</p>
<p><i>Godless</i> ~ (SCPI-1) <i>Tell them I'm a depraved monster. A godless fag, right, mom?</i></p>	<p>Viado ateu (SCPP-0) Diga-lhe que eu sou um monstro depravado. E um viado ateu. Certo, mãe?</p>
<p><i>Professional</i> ~ (SCPI-4) <i>I told you - I'm a professional football player, not a professional fag!</i></p>	<p>Bicha profissional (SCPP-1) Eu te disse, sou um jogador de futebol profissional, não uma bicha profissional!</p>

Adjetivo + *Faggot* (Substantivo)

<p style="text-align: center;">Brave ~ (SCPI-1)</p> <p><i>You stood beside your partner, no matter what anyone said, and that makes you a very loyal, very</i></p> <p style="text-align: center;">brave faggot.</p>	<p style="text-align: center;">Viado corajoso (SCPP-0)</p> <p>Você apoiou seu companheiro e independente do que os outros digam isso o torna um viado corajoso e muito leal.</p>
<p style="text-align: center;">Fucking ~ (SCPI-4)</p> <p><i>Not before you suck on this, you</i> fucking</p> <p style="text-align: center;">faggot.</p>	<p style="text-align: center;">Bicha de merda (SCPP-1)</p> <p>Não antes de você chupar aqui, sua bicha de merda.</p>
<p style="text-align: center;">Little ~ (SCPI-4)</p> <p><i>Oh, you fucking... dirty</i> little faggot!</p>	<p style="text-align: center;">Bichinha* ¥ (SCPP-6)</p> <p>Oh, você me paga... sua bichinha suja!</p>
<p style="text-align: center;">Silly ~ (SCPI-4)</p> <p><i>Well, what about your apology for calling Melanie a cunt? I'll apologize for calling her a cunt when she apologizes for calling me a</i> silly faggot.</p>	<p style="text-align: center;">Bicha idiota* (SCPP-0)</p> <p>Você não tem de se desculpar com Melanie, por tê-la chamado de cretina? Só se ela se desculpar por ter me chamado de bicha idiota.</p>

Gay (substantivo) + Substantivo

Gayon-gay (SCPI - 1)*I'm starting to suspect **gayon-gay** crime.***Ladrão gay** (SCPP - 1)Estou começando a suspeitar de um **ladrão gay**.

Gay (adjetivo) + Substantivo

<p style="text-align: center;">Gay ~ (SCPI-2)</p> <p><i>The centre's giving him their outstanding gay advocate award.</i></p>	<p style="text-align: center;">Defensor dos gays (SCPP-3)</p> <p>Dizem aqui que ele receberá o prêmio de Defensor dos Gays.</p>
<p style="text-align: center;">Gay ~ (SCPI-1)</p> <p><i>They don't want to think about it. Or about gay bashing or <i>Aids</i>, or any of the other shit we all have to live with.</i></p>	<p style="text-align: center;">Preconceito gay (SCPP-1)</p> <p>Eles não querem pensar nisso. Nem em preconceito gay ou <i>Aids</i>, ou nem nas coisas com as quais convivemos.</p>
<p style="text-align: center;">Gay ~ (SCPI- 1)</p> <p><i>My development people are going to shop it around, the gay crusader is too powerful to be defeated by some asshole who can't see beyond the box office.</i></p>	<p style="text-align: center;">Guerreiro gay (SCPP- 1)</p> <p>Meu pessoal de desenvolvimento está tentando vendê-lo, um guerreiro gay é muito poderoso para ser derrotado por alguns idiotas que não podem ver além das bilheterias.</p>
<p style="text-align: center;">Gay ~ (SCPI- 2)</p> <p><i>It's an honour to count among our close personal friends television's latest gay eunuch.</i></p>	<p style="text-align: center;">Gay estéril* (SCPP- 0)</p> <p>É uma honra apoiar meu amigo íntimo que se tornou o gay estéril da televisão.</p>
<p style="text-align: center;">Gay ~ (SCPI- 1)</p> <p><i>Senator Bexter's very supportive of gay issues.</i></p>	<p style="text-align: center;">Causas gays (SCPP- 1)</p> <p>A senadora Bexter apoia causas gays, claro que gostariam.</p>
<p style="text-align: center;">Gay ~ (SCPI- 4)</p> <p><i>It's not enough they can say no to gay marriage. Now they're gonna go after everything?</i></p>	<p style="text-align: center;">Casamento gay (SCPP- 4)</p> <p>Não foi o bastante o fato de terem dito "não" ao casamento gay e agora eles querem lhe tirar tudo?</p>
<p style="text-align: center;">Gay ~ (SCPI- 2)</p> <p><i>Well, last time I heard, they had a gay mayor.</i></p>	<p style="text-align: center;">Prefeito gay (SCPP- 2)</p> <p>A última coisa que ouvi é que eles tiveram um prefeito gay.</p>

<p>Gay ~ (SCPI- 1) <i>Anyone seen a certain hot little item in our local gay rag?</i></p>	<p>Jornal gay (SCPP- 1) Já viram o destaque no jornal gay?</p>
<p>Gay ~ (SCPI-4) <i>And I have fought for gay rights more than you have or ever will.</i></p>	<p>Direitos dos gays (SCPP-4) E eu luto pelos direitos dos gays mais do que você luta ou alguma vez lutou.</p>
<p>Gay ~ (SCPI-1) <i>Come on, you. Ab-b! Do you feel that gay thump-thump? God, I missed that!</i></p>	<p>Vibração gay (SCPP-1) Venha. Sente a vibração gay? Puxa, senti falta disso!</p>
<p>Gay ~ (SCPI-1) <i>Welcome to that quaint, heterosexual male tradition known as "the stag party", with a slightly gay twist.</i></p>	<p>Toque gay (SCPP-1) Bem-vindo a curiosa tradição heterossexual masculina conhecida como “Despedida de Solteiro”. Com um leve toque gay.</p>

Nesta amostra do Glossário, apresentamos, então, as 31 colocações de língua geral, que partiram das cinco bases *ass*, *butt*, *fag*, *faggot* e *gay*. A continuidade da pesquisa é, em uma próxima etapa, dar prosseguimento às análises das colocações que partem das demais bases que comporão o glossário, e, posteriormente, adaptá-lo, de modo a inseri-lo em uma plataforma *on-line*, como parte do projeto “A compilação de materiais didáticos e glossários baseados em *corpora* e sua contribuição para uma Pedagogia do Léxico e da Tradução”, desenvolvido e coordenado pela Profa. Dra. Adriane Orenha Ottaiano.

CONSIDERAÇÕES FINAIS

Esta pesquisa teve, por objetivo geral, o levantamento e a análise das colocações a partir do *corpus* paralelo formado pelas transcrições em inglês e em português do Brasil do seriado *Queer as Folk*, com o propósito de observar o léxico fraseológico da comunidade homossexual de maior frequência e/ou chavicidade, em relação ao léxico fraseológico da comunidade homossexual nas traduções das transcrições para o português do Brasil.

Além do referido *corpus* paralelo, fez-se necessária a compilação de um *corpus* comparável, coletado da *web*, automaticamente, com o auxílio da ferramenta *BootCat* inserida no *Sketch Engine* (KILGARRIFF et al, 2004), composto de um *subcorpus* em inglês e um *subcorpus* português para atestar as palavras mais frequentemente empregadas pela comunidade homossexual.

Extraímos e analisamos as colocações mais frequentemente empregadas no *corpus* paralelo em inglês e suas respectivas traduções, para, finalmente, elaborarmos uma amostra de glossário bilíngue de colocações, na direção inglês→português, a partir das palavras de maior chavicidade que retratam as combinações de palavras mais frequentemente empregadas pela comunidade homossexual.

Iniciamos o percurso desta dissertação pela contextualização sobre as motivações para o desenvolvimento desta pesquisa, haja vista que estudos da comunidade homossexual já vêm sendo desenvolvidos em outros países. Tratamos da Fraseologia enquanto ciência, partindo do arcabouço teórico-metodológico com base em diversos estudos realizados, nos quais se discutem denominações terminológicas, classificações e definições dos limites com relação aos fraseologismos.

Apresentamos a metodologia utilizada em nossa pesquisa, detalhando cada etapa para a compilação de nossos *corpora*, bem como o levantamento, extração e análise das colocações da comunidade homossexual.

Destacamos que, das 500 palavras-chave (*KeyWords*) de nosso *corpus* paralelo, foram selecionadas, para este trabalho, cinco itens lexicais principais que remetem ao léxico da comunidade homossexual, comumente empregados pelo referido grupo e que poderiam formar colocações da língua geral, a saber: *ass*, *butt*, *fag*, *faggot* e *gay*.

Durante a análise e discussão dos dados, observamos, no *corpus* paralelo, a frequência das colocações, assim como o contexto de uso, a fim de checar a situação de tais colocações.

Também utilizamos o *corpus* comparável para uma análise de frequência e de contexto, por meio das linhas de concordância. A título de exemplificação, podemos mencionar a colocação *gay rights*, que faz referência a “uma participação ativa no movimento dos direitos dos gays”, no *subcorpus* comparável em inglês e, no *subcorpus* comparável em português, refere-se a “um grupo formado por ativistas dos direitos gays”. Vemos, por meio dessa colocação, o empenho da comunidade para garantir os “direitos dos gays”.

Existem casos nos quais as colocações em inglês não formam colocações ao serem traduzidas para o português do Brasil, tornando-se apenas um item lexical, ou seja, não foi possível identificar uma colocação equivalente na língua de chegada como, por exemplo, *little faggot*, traduzido por “bichinha”; observamos no *corpus* paralelo e no *corpus* comparável que se trata de uma colocação de cunho sexual e depreciativa pelos contextos em que aparece. Chamou-nos também a atenção o fato de que a frequência da colocação em português, “bichinha” (71 ocorrências), é bem frequente em nosso *subcorpus* comparável em português, em relação à frequência de sua colocação equivalente em inglês, *little faggot* (15 ocorrências). Acreditamos, que a frequência de *little faggot* apresentou-se baixa devido à utilização de outros itens lexicais para designar uma mesma coisa como, por exemplo, *faggy*, que é utilizado como *little faggot*.

Sendo assim, acreditamos que este estudo acerca da extração e análise comparativa de colocações da comunidade homossexual, bem como a compilação de um glossário que contenha tais fraseologismos contribuiu para melhorar a compreensão da linguagem desse grupo, visto de forma tão estigmatizada. Ademais, possibilitou o acesso a uma obra fraseográfica rica em colocações nos dois idiomas enfocados, inglês e português do Brasil.

Por conseguinte, o estudo desenvolvido mostra sua relevância para a Fraseologia, especificamente o estudo das colocações da comunidade homossexual, visto que não há, até então, uma obra que contemple tais fraseologismos. Além disso, o resultado prático, isto é, a proposta de um *Glossário Bilíngue de Colocações da Comunidade Homossexual*, trará grande contribuição aos tradutores profissionais, tradutores aprendizes, terminólogos, terminógrafo, comunidade homossexual e demais interessados nessa temática, haja vista que tais combinações não são, geralmente, encontradas em dicionários de língua geral.

Sendo assim, um trabalho como este apresenta um impacto social à comunidade homossexual muito grande, já que pela primeira vez terão acesso a uma obra que demonstra as peculiaridades para uma comunicação oral e/ou escrita fluente em inglês e português do Brasil.

Ressaltamos que nosso estudo abre perspectiva de trabalho futuro, principalmente no que diz respeito às suas aplicações para o ensino-aprendizagem de língua estrangeira e para

tradução, para a área da Fraseologia/Fraseografia, principalmente no desenvolvimento de uma pedagogia do léxico baseada em *corpus*. Desta maneira, como mencionado, será disponibilizado o glossário em uma plataforma *on-line*, como parte do projeto “A compilação de materiais didáticos e glossários baseados em *corpora* e sua contribuição para uma Pedagogia do Léxico e da Tradução”, sob responsabilidade da Profa. Dra. Adriane Orenha Ottaiano.

Outrossim, como encaminhamento futuro, pretendemos desenvolver, durante o curso de Doutorado, a elaboração de atividades colocacionais baseadas no *Glossário Bilíngue de Colocações da Comunidade Homossexual*, tendo como público-alvo tradutores profissionais e aprendizes, comunidade homossexual e pessoas interessadas nessa temática, propondo a aplicação e a discussão de tais atividades, por meio de um questionário aberto.

REFERÊNCIAS BIBLIOGRÁFICAS

AGBAGO, A.; BARRIÈRE, C. Corpus construction for terminology. In: CORPUS LINGUISTICS CONFERENCE, 2005. *Proceedings...* Birmingham, 2005. p. 14-17.

BAKER, P. *Polari – the lost language of gay men*. New York: Routledge, 2002.

BALLY, C. *Traité de stylistique française*. 3. ed. Genève: Librairie Georg & Cie, Librairie C. Klincksieck, [1909], 1951.

BARBOSA, M. A. Da microestrutura dos vocabulários técnico-científicos. In: ENCONTRO NACIONAL DA ANPOLL, 4., 1989. *Anais...* Recife: ANPOLL, 1989.

BARRETT, R. The Homo-genius speech community. In: LIVIA, A.; HALL, K. *Queerly phrased: language, gender, and sexuality*. New York: Oxford University Press, 1997. p. 202-222.

_____. Supermodels of the world unite! Political economy and the language of performance among african-american drag queens. In: LEAP, W. L. (Ed.). *Beyond the lavender lexicon: authenticity, imagination and appropriation in lesbian and gay languages*. New York: Gordon and Breach Press, 1995. p. 207-226.

BARTSCH, S. *Structural and functional properties of collocations in English*. A corpus study of lexical and pragmatic constraints on lexical co-occurrence. Tübingen: Narr, 2004.

BEAUGRANDE, R. *New foundations for a science of text and discourse: cognition, communication, and the freedom of access to knowledge and society*. New Jersey: Norwood, 1996.

BECKER, R. *Gay TV and straight america*. New Jersey: Rutgers University Press, 2006.

BÉJOINT, H. *Tradition and innovation in modern English dictionaries*. Oxford: Oxford University Press, 1994.

BENSON, M. et al. (Ed.): *The BBI dictionary of english word combinations*. Revised edition. Amsterdam: John Benjamins Publishing Company, 1997.

BENSON, M.; BENSON, E.; ILSON, R. *Lexicographic description of english*. Amsterdam: J. Benjamins, 1986.

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

_____. *Pesquisa em linguística de corpus com o wordsmith tools*. Campinas: Mercado das Letras, 2009.

BETSCH, M. L. University conference focuses on ‘gay language’. *CNS News*, Reston, 7 jul. 2008. Disponível em: < <https://www.cnsnews.com/news/article/university-conference-focuses-gay-language>>. Acesso em: 13 jun. 2017.

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus linguistics: investigating language, structure and use*. Cambridge University Press: New York, 1998.

BIDERMAN, M T. C. Fundamentos da lexicologia. In: _____. *Teoria linguística: teoria lexical e linguística computacional*. São Paulo: Martins Fontes, 2001. p. 99-155.

BRONTSEMA, R. A Queer revolution: reconceptualizing the debate over linguistic reclamation. *Colorado Research in Linguistics*, University of Colorado, Boulder, v. 17, n. 1, p. 60-77, jun. 2004.

BUCHOLTZ, M.; HALL, K.. Theorizing identity in language and sexuality research. *Language in Society*, Cambridge, v. 33, n. 4, p. 469-515, 2004.

BURGER, H. *Phraseologie: eine einfuhrung am beispiel des deutschen*. Berlin: Erich Schmidt Verlag, 2007.

_____. *Phraseologie: eine einfuhrung am beispiel des deutschen*. Berlin: Erich Schmidt Verlag, 1998. Resenha de SOUZA, E. A. *Pandaemonium Germânicum*, 2001.

BUTLER, J. *Bodies that matter: on the discursive limits of "sex"*. New York: Routledge, 1993.

CAMERON, D.; KULICK, D. *The language and sexuality reader*. London: Routledge, 2006.

_____. *Language and sexuality*. Cambridge: Cambridge University Press, 2003.

_____. *Performing gender identity: young men's talk and the construction of heterosexual masculinity*. In: COATES, J. *Language and gender: a reader*. Oxford: Blackwell, 1998. p. 270-284.

CARTER, R. *Vocabulary – Applied linguistic perspectives*. New York: Routledge, 1998.

CARVALHO, G. L. *Expressões idiomáticas em dicionários escolares de língua portuguesa*. 2016. 253 f. Tese (Doutorado em Linguística Aplicada)-Centro de Humanidades, Universidade Estadual do Ceará, Fortaleza, 2016.

CASARES, S. J. *Introducción a la lexicografía moderna*. Madrid: SCIS, 1992.

_____. *Introducción a la lexicografía moderna*. Madrid: Consejo Superior de Investigaciones Lingüísticas, 1950.

CHAMBERS, J. K. *Sociolinguistic: theory linguistic variation and its social significance*. Revised Edition. Oxford: Wiley-Blackwell, 2009.

COATES, J. *Language and gender: a reader*. Oxford: Blackwell, 1998a.

_____. *Gossip revisited: language in all-female groups*. In: _____. *Language and gender: a reader*. Oxford: Blackwell, 1998b, 226-253.

CORPAS PASTOR, G. *Manual de fraseología española*. Madrid: Editorial Gredos, 1996.

COSERIU, E. *Lecciones de lingüística general*. Madrid: Gredos, 1981.

_____. *Teoría del lenguaje y lingüística general*. Madrid: Gredos, 1967.

COWIE, A. P. Speech formulae in English: problems of analysis and dictionary treatment. In: VAN DER MEER, G.; TER MEULEN, A. (Es.). *Making sense: from lexeme to discourse*. Groningen: Center for Language and Cognition, 2001. p. 1-12.

_____. Phraseology. In: ASHER, R. E. (Ed.) *The encyclopedia of language and linguistics*. Oxford: Pergamon, 1994. v. 6. p. 3168-3171.

_____ et al. *Oxford dictionary of english idioms*. Oxford: OUP, 1983.

_____. The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, Oxford, v. 2, p. 223-235, 1981.

CRIST, S. Duration of onset consonants. Gay male stereotyped speech. *Working Papers in Linguistics*, University of Pennsylvania, Pennsylvania, v. 4. n. 3, p. 53-70, 1997.

DARSEY, J. "Gayspeak": a response. In: CAMERON, D.; KULICK, D. (Ed.). *The language and sexuality reader*. New York: Routledge, 2006. p. 78-85.

DUBOIS, J. et al. *Dicionário de linguística*. São Paulo: Cultrix, 2006.

ECKERT, P. Communities of practice. In: BROWN, K.; ANDERSON, A. H. (Ed.). *Encyclopedia of Language and Linguistics*. Oxford: Elsevier, 2006. v. 2. p. 683-685.

_____; MCCONNELL-GINET, S. Think practically and look locally: language and gender as community-based practice. *Annual Review of Anthropology*, Palo Alto, v. 21, p. 461-490, 1992.

FERNANDO, C.; FLAVELL, R. *On idiom: critical views and perspectives*. Exeter: University of Exeter, 1981. (Exeter Linguistics Studies, v. 5)

FILLMORE, Charles. Inocence: a second idealization for linguistics. In: BERKELEY LINGUISTICS SOCIETY, 50., 1979. *Proceedings...* Berkeley, 1979.

FIRTH, J. R. Modes of meaning. In: _____ (Ed.). *Papers in Linguistics – 1934-1951*. Oxford: Oxford University Press, 1957. p. 114-118.

GAL, S.; IRVINE, J. The boundaries of languages and disciplines: how ideologies construct differences. *Social Research*, New York, v. 62, n. 4, p. 967-1001, 1995.

GLEDHILL, C. The discourse function of collocation in research article introductions. *English for Specific Purposes*, New York, v. 19, p. 115-135, 2000.

GRANT, R. *QaF Interview*. Retrieved 2 may 2011. Disponível em: <<http://bjfic.livejournal.com/2528384.html>>. Acesso em: 15 de Junho de 2017.

HALLIDAY, M. *On grammar*. Edited by J. Webster. London: Continuum International Publishing Group, 2005.

_____. Lexicology. In: _____ et al. *Lexicology and corpus linguistics*. London: Continuum, 2004. p. 1-22.

_____. Corpus studies and probabilistic grammar. In: AIJMER, K.; ALTENBERG, B. (Ed.). *English Corpus Linguistics*. London: Longman, 1991.

_____. *Lexis as a linguistic level*. In: BAZELL, C. E. et al. London: Longmans, 1966. p. 148-162.

_____; HASAN, R. *Language, context, and text: aspects of language in a social-semiotic perspective*. Geelong: Deakin University Press, 1985.

_____; _____. *Cohesion in English*. London: Longman, 1976.

HAUSMANN, F. J. *Einführung in die Benutzung der neufranzösischen Wörterbücher* (= Romanistische Arbeitshefte, Band 19). Tübingen: Max Niemeyer Verlag, 1977.

_____. Un dictionnaire de collocations est-il possible? *Travaux de linguistique et de littérature*, Strasbourg, v. 17, n. 1, p. 187-195, 1979.

_____. Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. *Praxis des Neusprachlichen Unterrichts*, Berlin, v. 31, p. 395-406, 1984.

_____. Kollokationen in deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels'. In : BERGENHOLTZ, H. ; MUGDAN, J. (Org.). *Lexikographie und Grammatik*. Tübingen: Niemeyer, 1985. p. 22-26.

_____. Le dictionnaire de collocations. In: _____ et al. *Wörterbücher, Dictionaries, Dictionnaires, erster Teilband*. Berlin: Walter de Gruyter 1989. p. 1010-1019.

_____. Le dictionnaire de collocations. In: _____ et al. *An International encyclopedia of lexicography*. v. 1. Berlin: Walter de Gruyter, 1990. p. 1010-1019.

_____. O dicionário de colocacões. Criterios de organizacão. In: *Actas do I Coloquio Galego de Fraseoloxía*. Santiago de Compostela: Centro Ramón Piñeiro, 1997. p. 63-81.

_____. Le dictionnaire de collocations – Critères de son organisation. In: GREINER, N. et al. (Ed.). *Texte und Kontexte in Sprachen und Kulturen*. Festschrift für Jörn Albrecht. Trier: WVT Wissenschaftlicher Verlag, 1999. p. 121-139.

HAYES, J. J. Gayspeak. In: CAMERON, D.; KULICK, D. (Ed.). *The language and sexuality reader*. London: Routledge, 2006. p. 68-77.

HOUAISS, A. *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objetiva, 2001.

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

HUNSTON, S.; FRANCIS, G. *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins Publishing Company, 2000.

IRVINE, J. T. What talk isn't cheap: language and political economy. *American Ethnologist*, Washington, v. 16, n. 2, p. 248-267, 1989.

JACOBS, G. Lesbian and gay male language use: a critical review of the literature. *American Speech*, New York, v. 71, p. 49-71, 1996.

KAISER, C. The Queerest show on earth. *New York Entertainment*, New York, retrieved. 2 may 2011. Disponível em: <<http://nymag.com/nymetro/arts/features/3788/>>. Acesso em: 02 jun. 2017.

KENNEDY, G. Amplifier collocations in the British National Corpus: implications for english language teaching. *TESOL Quarterly*, Washington, v. 37, p. 467-487, 2003.

_____. *An introduction to corpus linguistics*. London: Logman, 1998.

KILGARRIFF, A. et al. The sketch engine. In: WILLIAMS, G.; VESSIER, S. (Ed.). *Proceedings of the 11th Euralex International Congress*. Lorient: Universite de Bretagne-Sud, 2004. p. 105-116.

_____. et al. The sketch engine: ten years on. *Lexicography ASIALEX*, Hong Kong, v. 1, n. 1, p. 7-36, 2014.

KLARE, Johanne. Lexicología e fraseología no português moderno. *Revista de Filologia Românica*, Universidad Complutense, Madrid, n. 4, p. 355-360, 1986.

KOIKE, K. *Colocaciones léxicas en el español actual: estudio formal y léxico semântico*. Alcalá: Universidad de Alcalá/ Takushoku University, 2001.

KULICK, D. Gay and lesbian language. *Annual Review of Anthropology*, Palo Alto, v. 29, p. 243-285, 2000.

LABOV, W. *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics, 1966.

LAKOFF, R. *Language and woman's place*. New York: Harper & Row, 1975.

LEAP, W. L. *Word's out: gay men's english*. Minneapolis: University of Minnesota Press, 1996.

_____. Gay men's english: cooperative discourse in a language of risk. *Folklore*, New York, v. 19, n. ½, p. 45-70, 1993.

LECKIE-TERRY, H. *Language & context: a functional theory of register*. Edited by D. Birch. London, New York: Pinter, 1995.

LEECH, G. Corpora and theories of linguistic performance. In: SVARTVIK, J. (Ed.). *Directions in corpus linguistics: Proceedings of Nobel Symposium*. Berlin: Mouton de Cruyter, 1992.

_____. The state of the art in corpus linguistics. In: AIJMER, K.; ALTENBERG, B. (Ed.). *English corpus linguistics*. London: Longman, 1991. p. 8-29.

LEŚNIEWSKA, J. Collocations and second language use. *Studia Linguistica, Universitatis Jagellonicae Cracoviensis*, Cracóvia, v. 123, p. 95-105, 2006.

LEWIS, M. Language in the lexical approach. In: _____. *Teaching collocations*. London: Language Teaching Publications, 2000. p. 126-185.

LIVIA, A.; HALL, K. *Queerly phrased: language, gender, and sexuality*. New York: Oxford University Press, 1997.

MAIA, B. Making corpora – a learning process. In: BERNARDINI, S.; ZANETTIN, F. *I Corpora nella didattica della traduzione*. Bologna: CLUEB, 2000. p. 74-61.

MAHLBERG, M. *English general nouns: a corpus theoretical approach*. Amsterdam: John Benjamins Publishing Company, 2005.

MAKKAI, A. *Idioms structure in English*. The Hague: Mouton, 1972.

MARTIN, J. R. *English text: system and structure*. Amsterdam: John Benjamins Publishing Company, 1992.

MATUDA, S. *A fraseologia do futebol: um estudo bilíngue português-ínglês direcionado pelo corpus*. 2011. 304 f. Dissertação (Mestrado em Estudos Linguísticos e Literários)-Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2011.

MCENERY, T.; HARDIE, A. *Corpus linguistics: methods, theory and practice*. New York: Cambridge University Press, 2012.

_____; WILSON, A. *Corpus linguistics: an introduction*. Edinburgh University Press: Edinburgh, 2001.

MCINTOSH, C.; FRANCIS, B.; POOLE, R. (Ed.). *Oxford collocations dictionary for students of English (OCD)*. 2. ed. Oxford: Oxford University Press, 2009.

MEL'CUK, I. A. Collocations and lexical functions. In: COWIE, A. P. (Ed.). *Phraseology: theory, analysis, and applications*. Oxford: Clarendon Press, 1998. p. 23-53

MEYER, C. *English corpus linguistics: an introduction*. New York: Cambridge University Press, 2004.

MOON, R. *Fixed expressions and idioms in english*. Oxford: Clarendon Press, 1998.

MOONWOMON, B. Lesbian discourse, lesbian knowledge. In: LEAP, W. (Ed.). *Beyond the lavender lexicon: authenticity, imagination, and appropriation in lesbian and gay languages*. Australia: Gordon and Breach, 1995. p. 46-64.

MORRISH, L.; SAUNTON, H. *New perspectives on language and sexual identity*. New York: Palgrave Macmillan, 2007.

NESSELHAUF, N. *Collocations in a learner corpus*. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2005.

NOGUEIRA, L. C. R. *A presença de expressões idiomáticas (EIs) na sala de aula de E/LE para brasileiros*. 2008. 249 f. Dissertação (Mestrado em Linguística Aplicada)-Universidade de Brasília, Brasília, DF, 2008.

ORENHA-OTTAIANO, A. *A compilação de um glossário bilingue de colocações, na área de jornalismo de negócios, baseado em corpus comparável*. 2004. 246 f. Dissertação (Mestrado em Estudos Linguísticos e Literários)-Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2004.

_____. Algumas contribuições advindas da compilação de corpora especializados via Web e WebBootCat para a tradução, terminologia e fraseologia. In: TAGNIN, S. E. O.; VALE, O. *Avanços da linguística de corpus no Brasil*. São Paulo: Humanitas, 2008.

_____. The compilation of an online corpus-based bilingual collocations dictionary. In: CORPAS PASTOR, G. (Org.). *Computerised and corpus - based approaches to phraseology: monolingual and multilingual perspectives*. Genebra: Editions Tradulex, 2016. v. 1, p. 486-493.

_____. *Unidades fraseológicas especializadas: colocações e colocações estendidas em contratos sociais e estatutos sociais traduzidos no modo juramentado e não-juramentado*. 2009. 282 f. Tese (Doutorado em Estudos Linguísticos)-Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2009.

ORTIZ ÁLVAREZ, M. L. As expressões idiomáticas dentro da obra lexicográfica. *Revista Brasileira de Linguística*, São Paulo, v. 9, n. 1, p. 181-212, 1997.

PALMER, H. E. *Second interim report on english collocations*. Tokyo: Kaitakusha, 1933.

QUEEN, R. "I don't speak spritch": locating lesbian language. In: LIVIA, A.; HALL, K. (Ed.). *Queerly phrased: language, gender and sexuality*. New York: Oxford University Press, 1997. p. 233-256.

QUIRK, R. et al. *A comprehensive grammar of the english language*. London: Longman, 1985.

REY-DEBOVE, J. *Étude linguistique et sémiotique des dictionnaires français contemporains*. Paris: Mouton, 1971.

ROCHA, J. M. P. *Fraseologia jurídico-comercial: colocações especializadas no corpus UNCITRAL e proposta de glossário trilingue na área do Direito Comercial Internacional*.

2017. 292 f. Dissertação (Mestrado em Estudos Linguísticos)–Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2017.

RODGERS, B. *The queen's vernacular: a gay lexicon*. San Francisco: Straight Arrow Books, 1972.

RUIZ GURILLO, L. Aspectos de fraseologia teórica española. In: _____. *Cuadernos de filología*. Valencia: Universidad de Valencia, 1997. p. 1-125.

SÁNCHEZ, A. et al. *Cumbre. Corpus lingüístico del español contemporáneo. Fundamentos, metodología y aplicaciones*. Madrid: SGEL, 1995.

SANTAMARÍA PÉREZ, M. I. *Tratamiento de las unidades fraseológicas en la lexicografía bilingüe Español-Catalã*. 2000. 388 f. Tese (Doutorado em Letras)-Universidade de Alicante, Alicante, 2000.

SAUSSURE, Ferdinand de. *Curso de linguística geral*. Trad. Antônio Chelini, José Paulo e Izidoro Beinkstein. São Paulo: Cultrix, 1975.

SILVA, M. B. Uma palavra só não basta: um estudo teórico sobre as unidades fraseológicas. *Revista de Letras*, São Paulo, v. ½, n. 28, p. 11-20, jan.-dez. 2006.

SCOTT, M. *WordSmith Tools: version 6.0*. Oxford: Oxford University Press, 2012.

SINCLAIR, J. From theory to practice. In: LEECH, G.; MYERS, G.; THOMAS, J. (Ed.). *Spoken english on computer*. Harlow: Longman, 1995. p. 99-112.

_____. Corpus and text – basic and principles. In: WYNNE, M. (Ed). *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow Books: 2005. p. 1-16.

_____. *Corpus, concordance, collocation*. Oxford: OUP, 1991.

_____. *Looking Up. An account of the COBUILD project in lexical computing*. London: Collins Cobuild, 1987.

_____. The search for units of meaning. In: CICLE DE CONFERÈNCIES 95-96. LÈXIX, CORPUS I DICCIONARIS, 1996. *Proceedings...* Barcelona: Universitat Pompeu Fabra, Institut de Linguística Aplicada, 1996. p. 97-107.

_____; JONES, S.; DALEY, R. (1970). English Lexical Studies: report to OSTI on project C/LP/08. New Edition: Krishnamurthy (Ed.) *English Collocation Studies: the OSTI report*. Birmingham: Birmingham University Press, 2003.

STANLEY, J. Homosexual slang. *American Speech*, New York, v. 45, n. 1/2, p. 45-59, 1970.

STUBBS, M. Collocations and semantic profiles: on the cause of the trouble with quantitative methods. *Function of Language*, Amsterdam, v. 2, n. 1, p.1-33,1995.

TAGNIN, S. E. O. Collecting data for a bilingual dictionary of verbal collocations: from scraps of paper to corpora research. In: PRACTICAL APPLICATIONS IN LANGUAGE CORPORA, 1999. *Proceedings...* Lodz: Lodz University Press, 1999.

_____. *O jeito que a gente diz: expressões convencionais e idiomáticas*. São Paulo: Disal, 2005.

_____. *O jeito que a gente diz*. São Paulo: Disal, 2013.

TANNEN, D. *You just don't understand: women and men in conversation*. New York: Morrow, 1990.

TOGNINI-BONELLI, E. *Corpus linguistic at work*. Amsterdam: John Benjamins, 2001.

_____. Towards translation equivalence from a corpus linguistics perspective. In: SINCLAIR, J. et al. (Ed.). *Grammar patterns*. London: Collins COBUILD, 1996. p. 197-217.

_____; MANCA, E. Welcoming children, pets and guests: towards functional equivalence in the languages of 'agriturismo' and 'farmhouse holidays'. *TRADTERM*, São Paulo, v. 10, p. 295-312, 2004.

TRISTÁ, A. M. *Fraseologia y contexto*. La Habana: Editorial de Ciencias Sociales, 1988.

TVSUBTITLES. *Queer as folk*. Retrieved. Disponível em: <<http://www.tvsubtitles.net/tvshow-229-1.html>>. Acesso em: 20 maio 2013.

WOLFE, S. J.; RORIPAUGH, L. A. The (In)visible lesbian: anxieties of representation in the L word. In: AKASS, K.; MCCABE, J. (Ed.). *Reading the L word: outing contemporary television*. New York: I. B. Tauris & Company, 2006. p. 43-54.

WOOLARD, G. Collocation – encouraging learner independence. In: LEWIS, M. *Teaching collocations*. London: Language Teaching Publications, 2000. p. 28-46.

WRAY, A. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press, 2002.

ZULUAGA, A. *Introducción al estudio de las expresiones fijas*. Frankfurt: Peter D. Lang, 1980.