



University of  
Texas Libraries



e-revist@s

Sumários.org



Centro Universitário Santo Agostinho

# revista fsa

[www4.fsnet.com.br/revista](http://www4.fsnet.com.br/revista)

Rev. FSA, Teresina, v. 16, n. 1, art. 10, p. 213-230, jan./fev. 2019

ISSN Impresso: 1806-6356 ISSN Eletrônico: 2317-2983

<http://dx.doi.org/10.12819/2019.16.1.10>

DOAJ DIRECTORY OF  
OPEN ACCESS  
JOURNALS

WZB  
Wissenschaftszentrum Berlin  
für Sozialforschung



## Gay Language: Colocações da Comunidade Gay

## Gay Language: Collocations of the Gay Community

### Guilherme Aparecido de Souza

Doutorado em Estudos Lingüísticos pela UNESP

Mestrado em Estudos Lingüísticos pela UNESP

E-mail: [conecctado@hotmail.com](mailto:conecctado@hotmail.com)

### Adriane Orenha-Ottaiano

Doutora em Estudos Lingüísticos pela UNESP

Docente na Universidade Estadual Paulista

Email: [adriane.ottaiano@unesp.br](mailto:adriane.ottaiano@unesp.br)

---

### Endereço: Guilherme Aparecido de Souza

Secretaria Municipal de Educação de São José do Rio Preto, E.M. Darcy Ribeiro. Rua Geraldo Barbosa de Oliveira, Jardim Santo Antônio, 15047166 - São José do Rio Preto, SP – Brasil.

### Endereço: Adriane Orenha-Ottaiano

Universidade Estadual Paulista Júlio de Mesquita Filho, Instituto de Biociências Letras e Ciências Exatas de São José do Rio Preto. Rua Cristóvão Colombo, 2265 Jardim Nazareth 15054000 - São José do Rio Preto, SP - Brasil.

Editor-Chefe: Dr. Tonny Kerley de Alencar Rodrigues

Artigo recebido em 20/09/2018. Última versão recebida em 10/10/2018. Aprovado em 11/10/2018.

Avaliado pelo sistema Triple Review: a) Desk Review pelo Editor-Chefe; e b) Double Blind Review (avaliação cega por dois avaliadores da área).

Revisão: Gramatical, Normativa e de Formatação



## RESUMO

Esta pesquisa tem como apoio o arcabouço teórico-metodológico da Linguística de *Corpus* e da Fraseologia, 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 um *corpus* comparável, composto de um *subcorpus* em inglês e um *subcorpus* português para atestar as palavras 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 foque 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, as well as in the literature on phraseological, 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 attest to the words most frequently 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 Community.

## 1 INTRODUÇÃO

O estudo das unidades fraseológicas (doravante UFs) comumente empregadas pela comunidade homossexual presentes no seriado *Queer as Folk* possuem grande importância para as áreas da Fraseologia, Fraseografia e para a Linguística de *Corpus*, uma vez que ainda há poucas pesquisas que envolvem o léxico, principalmente o léxico fraseológico da comunidade homossexual.

Barrett (1997, p. 202), em seu artigo *The Homo-genius Speech community*, 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), em *Queerly Phrased: language, gender, and sexuality*, mostram aos leitores o lugar da linguagem na vida dos seres humanos e sua integração dentro de um grupo social menos demográfico, tradicionalmente tratados por meio de estudos da variação.

Neste estudo, as autoras comprovam que a integração/aquisição, por todos seus membros, do léxico empregado do grupo Lésbicas, Gays, Bissexuais e Travestis (LGBT) analisado, faz com que o sentimento de “pertencer a” surja, uma vez que, por meio da linguagem, mais especificamente, por intermédio de colocações é possível identificar os pares, isto é, aqueles que apresentam características comuns.

Além disso, o resultado prático da investigação aqui proposta, ou seja, o glossário bilíngue de colocações da comunidade homossexual, poderá contribuir para tradutores, pesquisadores da área e, ainda, para a própria comunidade homossexual, dada a ausência de uma obra fraseográfica que trata de fraseologias desta comunidade específica.

Os interessados no estudo dos fraseologismos da comunidade homossexual não têm acesso a glossários de tais combinatórias de palavras, por exemplo, no Brasil, de que se tenha conhecimento. No exterior, tais estudos já vêm sendo desenvolvidos em outros países, como é o caso dos Estados Unidos.

A primeira coleção de artigos sobre questões da linguagem *gay* e *lésbica*, *Gayspeak: Gay Male and Lesbian Communication*, de Chesebro, foi publicado em 1981, seguido de *Queer words, Queer images: communication and construction of homosexuality* de William Leap em 1994. Livia e Hall (1997, p. 24-25), por exemplo, mostram que um dos primeiros glossários *gays*, *The language homosexuality: An American Glossary*, de Gershon Legman, lançado em 1941, é composto de 329 vocábulos.

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

Desse modo, acreditamos que este estudo acerca da extração e análise de colocações da comunidade homossexual, bem como a compilação de um glossário que contenha tais fraseologismos venham contribuir para melhorar a compreensão da linguagem desse grupo tão estigmatizado. Ademais, possibilitarão o acesso a uma obra fraseográfica rica em colocações nos dois idiomas enfocados, inglês e português.

Este estudo contribui ainda para o desenvolvimento de pesquisas na área de Fraseologia e da Linguística de *Corpus*. Além disso, auxilia na disseminação de tais fraseologismos para o aprendizado de léxico em língua inglesa, a partir de uma amostra de um glossário de vocábulos da comunidade homossexual contida no seriado *Queer as Folk*.

Para isso, abordaremos aspectos teórico-metodológicos necessários à extração de colocações mais frequentemente empregadas, por meio de um *corpus* paralelo bilíngue formado pelas transcrições das cinco temporadas da série *Queer as Folk* e, de um *corpus* comparável, composto de um *subcorpus* em inglês e um *subcorpus* português para atestar as palavras mais frequentemente empregadas pela comunidade homossexual, compilado pela ferramenta *Sketch Engine* (KILGARRIFF *et al*, 2004).

O referido seriado narra a história de cinco homens homossexuais sendo, portanto, um marco na luta pelos direitos LGBT, adquiridos pelo investimento em uma trama sem cunho pornográfico ou apelativo, mostrando homossexuais como pessoas comuns, vivendo seu cotidiano.

Acreditamos que, ao desenvolver um estudo focado nas colocações e baseado em *corpus*, passamos a observar a língua sob um novo aspecto, pois chamamos a atenção para a companhia mantida pelas palavras, ou seja, pelas combinações lexicais, pelas frequências e recorrências destas na língua em uso.

Desse modo, ao enfatizar e adotar a perspectiva que privilegie o falar natural no estudo das combinações lexicais (em especial as colocações) é necessário discorrer sobre os arcabouços teóricos citados abaixo, sendo eles: *Corpus*: definição, tipologia e contribuição; O conceito de colocação para os teóricos da Linguística de *Corpus*; Compilação do *corpus* e Levantamento das colocações da comunidade homossexual que levaram ao desenvolvimento deste artigo, bem como das colocações contidas na série *Queer as Folk*.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 *Corpus*: definição, tipologia e contribuição

A Linguística de *Corpus* é uma área interdisciplinar que está em grande desenvolvimento nas últimas décadas; apresenta estudos valiosos com a utilização de *corpora*. Para a elaboração de uma proposta fraseográfica, como este, que visa a elaboração de uma amostra de um Glossário de Colocações da Comunidade Homossexual.

Para tanto, far-se-á necessário, no presente artigo demonstrar qual acepção de *corpus* estamos utilizando:

*A corpus* is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as possible, a language or language variety as a source of data for linguistic research. (SINCLER, 2005, p. 2).

Como podemos observar na definição de Sincler, nosso *corpus* de estudo foi compilado com base em critérios de representatividade, sendo ele em formato eletrônico, a fim de que possa ser manuseado pelo *software*.

Entendemos que um *corpus* é uma compilação eletrônica e criteriosa de textos que ocorrem naturalmente com o objetivo de representar uma dada parcela de língua ou algum de seus aspectos mais pontuais, de forma a possibilitar uma análise linguística previamente delimitada.

A descrição da linguagem partindo da análise de um *corpus* de estudo para a elaboração de um glossário é uma atividade bastante comum nas pesquisas em Linguística de *Corpus*, entretanto, a descrição da linguagem com a finalidade de elaborar um Glossário de Colocações da Comunidade Homossexual não é algo que se encontra com facilidade, por tal motivo, faz-se necessária.

O *corpus*, desenvolvido é, portanto, essencial, para as diversas etapas de elaboração de um glossário; destacamos, também, a existência de vários tipos de *corpus*, dentre eles, vale ressaltar o *corpus* paralelo, objeto de nosso estudo, que consoante Frankenberg-Garcia (2008, p. 118):

É basicamente uma coleção extensa de textos naturais, selecionados de acordo com critérios específicos e armazenados em formato digital. Um *corpus* paralelo é, por sua vez, uma combinação de pelo menos dois sub-*corpora* alinhados entre si. Na sua acepção mais simples, podemos ter, de um lado, um sub-*corpus* composto de textos originais numa determinada língua (L1) e, do outro, um sub-*corpus* com os mesmos textos traduzidos para uma outra língua (L2).

Em nosso caso, observamos que o *corpus* compilado para esta pesquisa apresenta um conjunto de dados linguísticos pertencentes ao uso oral, por tratar-se das transcrições das cinco temporadas do seriado *Queer as Folk* e suas respectivas traduções, formando assim, nosso *corpus* paralelo de estudo.

Relacionado à noção de representatividade do *corpus*, temos a questão do tamanho de um *corpus* de estudo, para a qual não há uma definição clara o suficiente (TOGNINI-BONELLI, 2001, p. 57; BOWKER; PEARSON, 2002, p. 45). Nessa premissa, é apropriado, pautar o entendimento do tamanho dos *corpora* a partir da proposta de Bowker e Pearson (2002, p. 45-46):

Unfortunaly, there are no hard and fast rules that can be fallowed to determine the ideal size of a corpus. Instead, you will have to make this decision based on factors such as the needs of the Project, the availability of data and the amount of time that you have. It is very important, however, not to assume that bigger is always better. You may find that you can get more useful information from a corpus that is small but well-designed than from one that is larger but is not customized to meet your needs.

Iniciando assim, que a adequação do conteúdo de um corpus deve prevalecer sobre as questões acerca do tamanho. Os dados devem ser representativos do uso da linguagem que se deseja investigar, ponto fundamental para realizar uma pesquisa baseada em *corpus*; desde que o recorte ou a língua na qual temos o objeto de pesquisa seja representado pelo *corpus* de estudo.

A próxima seção tem por objetivo discutir algumas das questões fundamentais referente ao “Conceito de Colocação” adotado para o presente estudo, à luz da perspectiva da Fraseologia.

## 2.2 O Conceito de Colocação Geral

Segundo Woolard (2000, p. 29), as colocações são “palavras que são estatisticamente muito mais propensos a aparecer juntos do que acaso sugere”, aprender mais vocabulário não é apenas aprender novas palavras, mas também novas combinações, para tanto, faz-se necessário o estudo aprofundado das colocações em língua inglesa, aqui, especialmente das colocações comumente empregadas pela comunidade homossexual do seriado *Queer as Folk*, a fim de elaborar uma amostra de Glossário de Colocações da Comunidade Homossexual.

Há muitos anos, Firth (1957) define as colocações como “as palavras que mantem a companhia”, em suas relações com outras palavras. Outra definição poderia ser “a forma

como as palavras se combinam de forma previsível”. Partindo desse pressuposto, estudaremos as colocações encontradas em nosso *corpus* de estudo.

Lewis (2000, p.132) define as colocações como “a maneira em que as palavras coocorrem no texto natural em formas estatisticamente significativas”. Parece-nos uma definição inocente, mas um ponto muito importante precisa ser destacado: a combinação previsível é sobre a forma como as palavras coocorrem naturalmente que, a nosso ver, é a linguagem 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 pode usá-los facilmente e inconscientemente.

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

[...] 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.

A ideia principal na definição de Wray (2002) é que a linguagem estereotipada consiste em unidades de linguagem 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 coocorrendo 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”.

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.

A seguir, apresentaremos a base da tipologia apresentada por Orenha-Ottaiano (2004, p.33-34 e 2009, p.41-42), à luz da proposta de Hausmann (1985).

**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 básica:

→Adjetivo (colocado) + Substantivo (base)

**Adverbiais** – com três formas básicas:

→Advérbio (colocado) + Adjetivo (base)

→Verbo (base) + Advérbio (colocado)

→Advérbio (colocado) + Verbo (base)

Após a revisão teórica que embasou o presente artigo, que visa à elaboração de uma amostra de Glossário de Colocações da Comunidade Homossexual, seguiremos discorrendo sobre a metodologia empregada durante o desenvolvimento desse estudo.

### 3 METODOLOGIA

Para o levantamento e a análise das colocações, compilamos um *corpus* paralelo constituído 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.

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

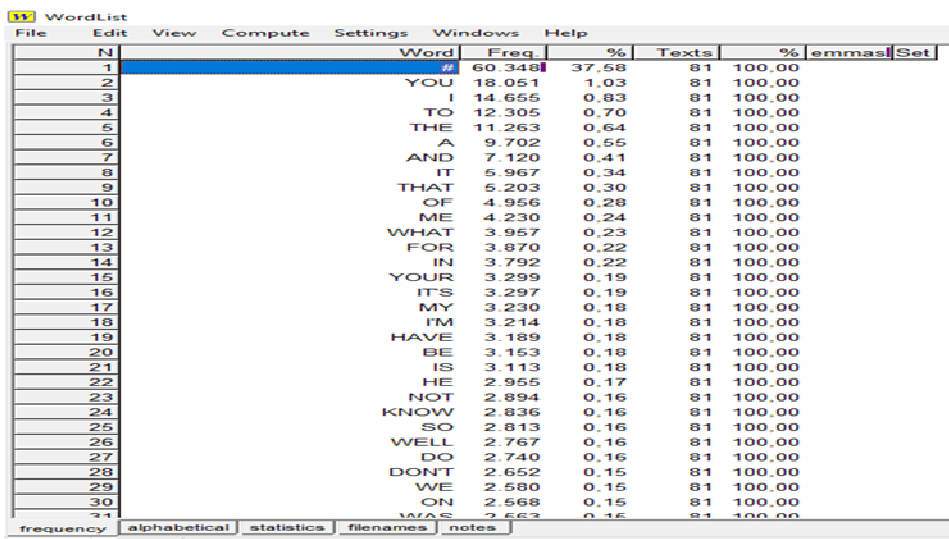
Conforme mencionamos, para Berber Sardinha (2004), o *corpus* é um objeto criado com fins específicos de pesquisa e de acordo com a sua extensão e o propósito da pesquisa. O *corpus* paralelo desta pesquisa é considerado médio, já que foi suficiente para o desenvolvimento desta pesquisa, pois é composto por 81 episódios, totalizando 5.177.124 palavras para o *subcorpus* em inglês, e 4.556.915 palavras para o *subcorpus* paralelo em português. 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*, o que permitiu o manuseio do *corpus* pelo referido *software*, favorecendo, assim a realização das análises baseadas na frequência e na coocorrência de palavras em *corpora*. Assim realizamos o levantamento das colocações.



A ferramenta apresenta três recursos: *WordList*, *Concord* e *KeyWords*. A *WordList* produz listas de palavras contendo todas as palavras do arquivo ou arquivos selecionados, elencadas em conjunto com suas frequências absolutas e percentuais.

**Figura 1 – Lista de palavras elaborado pelo recurso da *WordList*.**



N	Word	Freq	%	Texts	%	lemmas	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	I'M	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		

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

O programa é pré-definido para produzir, a cada vez, duas listas de palavras, uma ordenada alfabeticamente e outra classificada por ordem de frequência das palavras (com a palavra mais frequente encabeçando a lista).

A ferramenta *Concord* realiza concordâncias, ou listagens de uma palavra específica (o nódulo), juntamente com parte do texto onde ocorreu.

Figura 2 – Aba de concordâncias da ferramenta *Concord*.

N	Concordance	Set	Tag	Word #	t	#	os	#	os	t	#	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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.en.txt	7%
11	Remember, you were once a gay young thing yourself. 1191			12,354	1711	9%	0	8%	0	8%	12	hr	hdty.en.txt	77%
12	Remember, you were once a gay young thing yourself. 1190			12,341	1701	9%	0	8%	0	8%	12	hr	hdty.en.txt	77%
13	--> 00:33:56,620 And I have fought for gay rights more than you have or ever			11,493	1131	8%	0	2%	0	2%	12	hr	hdty.en.txt	72%
14	--> 00:33:52,658 And I have fought for gay rights more than you have or ever			11,475	1121	8%	0	2%	0	2%	12	hr	hdty.en.txt	72%
15	.where they accept and respect gay people. 1079 00:32:43,505 -->			11,146	0791	9%	0	0%	0	0%	12	hr	hdty.en.txt	70%
16	.where they accept and respect gay people. 1078 00:32:40,919 -->			11,135	0791	9%	0	0%	0	0%	12	hr	hdty.en.txt	70%
17	--> 00:30:10,519 It's his first time in a gay club. 1015 00:30:10,519 -->			10,464	0291	7%	0	6%	0	6%	12	hr	hdty.en.txt	66%
18	--> 00:30:08,851 It's his first time in a gay club. 1014 00:30:08,851 -->			10,452	0291	4%	0	6%	0	6%	12	hr	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.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	hdty.en.txt	25%
29	about losing your fans now that you're gay? 1153 00:38:01,506 -->			12,703	3721	0%	0	2%	0	2%	10	hr	hdty.en.txt	92%
30	about losing your fans now that you're gay? 1152 00:37:59,046 -->			12,688	3711	0%	0	2%	0	2%	10	hr	hdty.en.txt	91%
31	00:37:17,917 --> 00:37:20,688 Are you gay? What's that not to do with			12,257	3331	0%	0	0%	0	0%	10	hr	hdty.en.txt	88%

Fonte: PrintScreen da tela do *WordSmith Tools*.

Oferece também listas de colocados, isto é, palavras que ocorreram perto do nóduo. O sucesso da busca no *Concord* depende da especificação correta do termo de busca.

A *KeyWords* extrai palavras de uma lista, cujas frequências são estatisticamente diferentes (maiores ou menores) do que as frequências das mesmas palavras num outro *corpus* (de referência). Calcula também palavras-chave, que são chave em vários textos. Palavras-chave não são o mesmo que palavras “importantes”. O *corpus* de referência deve ser de língua geral e, preferencialmente, cinco vezes maior que o *corpus* de estudo para garantir que as palavras-chave sejam, realmente, especificidades do *corpus* de estudo. As palavras-chave são aquelas que ocorrem com mais frequência, estatisticamente, no *corpus* de estudo do que no *corpus* de referência.

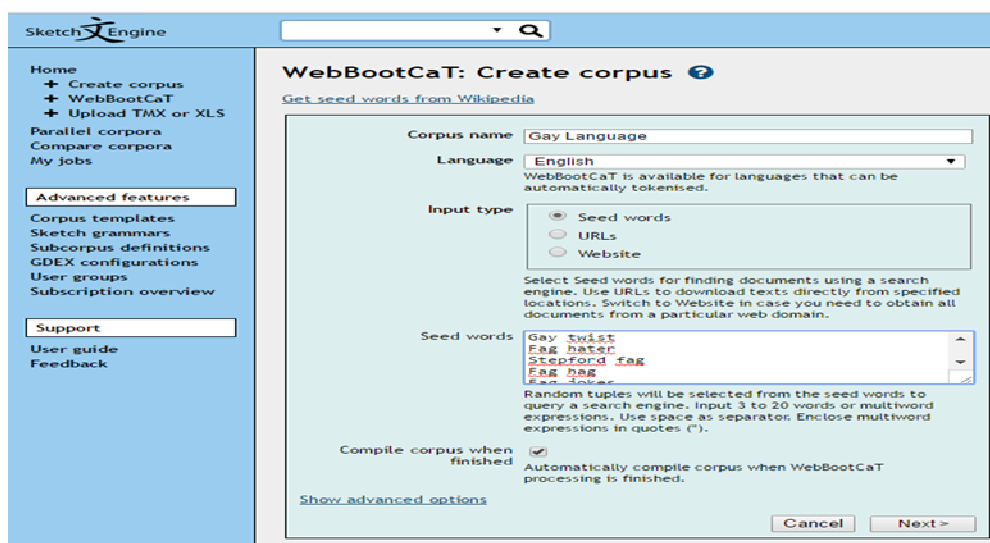
Após a seleção das palavras-chave, o passo seguinte foi o levantamento das colocações com o auxílio da ferramenta *Concord*, por meio das abas *concordance* e *collocates*, exemplificando, com a base *ass* (bunda).

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

Por meio desta ferramenta tivemos a oportunidade de compilar *corpora* comparáveis em língua inglesa e língua portuguesa para, assim, podermos aferir com precisão a utilização das colocações destacadas em nosso estudo.

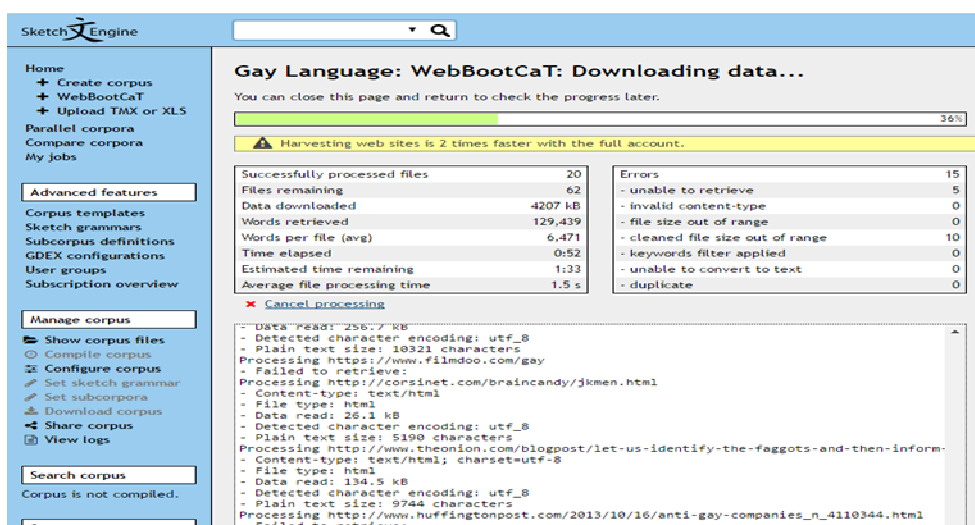
No próprio *Sketch Engine*, podemos clicar no ícone do *BootCat*, é uma interface que serve como guia na criação de *corpora* simples a partir da *web*, como podemos ver abaixo:

Figura 3 – Aba de compilação do *corpus*.



Fonte: PrintScreen da tela do *Sketch Engine*.

Figura 4 – A ferramenta fazendo *download* dos dados para a compilação do *corpus*.



Fonte: PrintScreen da tela do *Sketch Engine*.

Por meio desta importante ferramenta pudemos compilar os dois *corpora* comparáveis necessários para comprovar a veracidade de nossos dados, uma vez que a utilização de um *corpus* referência não demonstrou utilidade no nosso estudo, por tratar-se de colocações restritas que em sua grande maioria, são utilizadas, preferencialmente, pela comunidade *gay*. Na sequência, vamos discutir e analisar as frequências correspondentes a algumas colocações levantadas em nosso estudo.

## 4 RESULTADOS E DISCUSSÕES

### 4.1 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, um item lexical principal que remete ao léxico da comunidade homossexual, comumente empregado pelo referido grupo e que poderia formar colocações da língua geral, a saber: *ass*.

Conforme mencionado na metodologia, as linhas de concordância para a palavra de busca *ass*, geradas pela ferramenta *Concord*, mostram que se trata de uma palavra recorrente em nosso *corpus* paralelo, com a frequência de 244. Por meio da análise preliminar quanto à sua coocorrência, notamos que *ass* 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.

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 um item lexical selecionado por sua alta frequência no *corpus* paralelo, segundo mencionamos anteriormente: *ass*. As colocações geradas a partir deste item lexical irão compor a amostra de nosso Glossário de Colocações da Comunidade Homossexual.

O quadro 1, 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.

**Quadro 1 – Colocações adjetivas a partir da base *ass*.**

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

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 1 abaixo), no *subcorpus* paralelo do português, ou seja, de sua tradução – “traseiro apertado” (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).

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 1 – Estatísticas da colocação *tight ass*.**  
*Tight* (adjetivo) + Substantivo  
*Tight ass* – Traseiro apertado

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
08	08	05	0

A colocação adjetiva extraída de nosso *corpus* paralelo, como descrito na tabela 1 e exemplificada na figura 5 estrutura-se, sintagmaticamente em *adj.* + *noun*, em inglês. É formada pelo adjetivo em função atributiva (QUIRK *at 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” se trata da tradução da série.

A seguir, temos a figura 5 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 5 – Frequência de *tight ass* no *subcorpus* comparável em inglês**



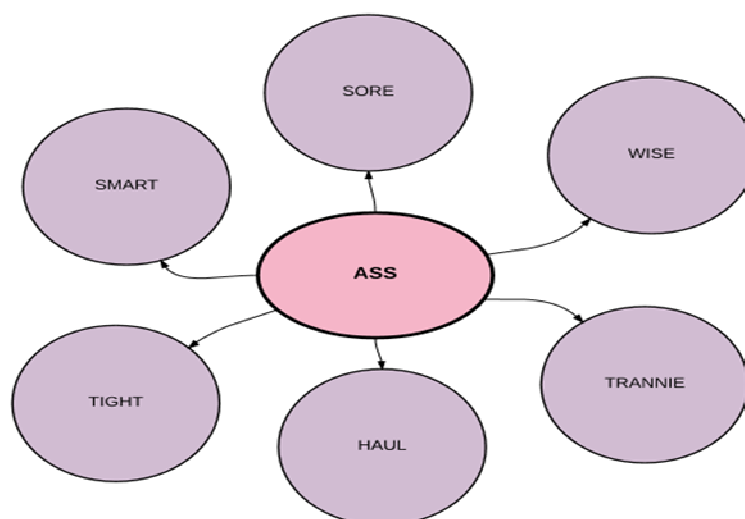
Fonte: Printscreen da tela do Sketch Engine.

Vemos nas linhas de concordância, presentes na figura 5, 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 1 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 acrescentá-los nos respectivos organogramas, porque esses itens aparecem em nossa amostra de Glossário. Vejamos a seguir o organograma 1:

### Organograma 1 – Colocações formadas pela base *ass*.



Fonte: Elaborado pelo autor.

Encontramos, no organograma 1, os itens lexicais de maior chavicidade em nossa pesquisa que formam colocações com *ass*.

## 5 CONCLUSÃO

O principal objetivo desse artigo foi tratar da extração e análise das colocações mais frequentes da comunidade *gay*, encontradas em nosso *corpus* de estudo, *Queer as folk*, a partir de uma palavra-chave mais frequente, *ass*, contribuindo, desse modo, para nossa pesquisa que investiga a linguagem *gay* (*gay language*).

Como resultado prático deste trabalho, objetivamos compilar um glossário bilíngue de colocações da referida comunidade, na direção inglês→português e português→inglês, com base no *corpus* paralelo bilíngue, constituído pelas transcrições das legendas em inglês e as traduções em português, dos episódios do seriado *Queer as Folk*.

Além do referido *corpus* paralelo, fez-se necessário a compilação de um *corpus* comparável, coletado da *web*, automaticamente, com o auxílio da ferramenta *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.

Tal estudo servirá como fonte de pesquisa para estudantes de língua inglesa, pesquisadores e interessados no estudo de Linguística de *Corpus* e da Fraseologia, especificamente das colocações da comunidade *gay*, além do público LGBT de modo geral, haja vista que tais combinações não são encontradas em dicionários de língua geral.

## REFERÊNCIAS

- BARRETT, R. The Homo-genius speech community. In LIVIA, A; HALL, K. **Queerly Phrased: Language, Gender, and Sexuality**. Ed. New York: Oxford University Press, p. 202-222, 1997.
- BERBER SARDINHA, T. **Linguística de Corpus**. São Paulo: Manole, 2004.
- BOWKER, L.; PEARSON, J. **Working with specialized language: a practical guide to using corpora**. London: Routledge, 2002.
- FIRTH, J. R. **Papers in Linguistics 1934-1951**. London: Oxford University Press, 1957.
- FRANKENBERG-GARCIA, A. Compilação e uso de corpora paralelos. In Stella E. O. Tagnin & Oto Araújo Vale (eds.), **Avanços da Linguística de Corpus no Brasil**. USP, São Paulo: Editora Humanitas, pp. 117-136, 2008,
- HAUSMANN, F. J. O dicionário de colocacions. Criterios de organización. In: **Actas do I Coloquio Galego de Fraseoloxía**. Santiago de Compostela: Centro Ramón Piñeiro, p. 63-81, 1997.
- HAUSMANN, F. J. **Kollokationen in deutschen** Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels'. In: BERGENHOLTZ, H.; MUGDAN, J. (Org.). **Lexikographie und Grammatik**. Tübingen: Niemeyer, p. 22-26, 1985.
- 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, p. 105-116, 2004.



LEWIS, M. **Teaching collocation**: Further developments in the lexical approach, p.126–154. Hove: Language Teaching Publications, 2000.

LIVIA, A; HALL, K. **Queerly Phrased**: Language, Gender, and Sexuality. eds. New York: Oxford University Press, 1997.

ORENHA-OTTAIANO, A. **A compilação de um glossário bilíngue de colocações, na área de jornalismo de Negócios, baseado em corpus comparável**. 246f. Dissertação (Mestrado em Estudos Linguísticos e Literários), FFLCH/USP, São Paulo, SP, 2004.

\_\_\_\_\_. **Unidades fraseológicas especializadas**: colocações e colocações estendidas em contratos sociais e estatutos sociais traduzidos no mundo juramentado e não juramentado. 2009. 282f. Tese (Doutorado em Estudos Linguísticos) – IBILCE, Universidade Estadual Paulista, São José do Rio Preto, 2009.

\_\_\_\_\_. 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., 2016.

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

SINCLAIR, J.M. **Corpus Concordance Collocation**. Oxford University Press, 1991.

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

TOGNINI-BONELLI, E. **Corpus linguistics at work**. Amsterdam/Philadelphia: John Benjamins, 2001.

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

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

**Como Referenciar este Artigo, conforme ABNT:**

SOUZA, G. A; ORENHA-OTTAIANO, A. Gay Language: Colocações da Comunidade Gay. **Rev. FSA**, Teresina, v. 16, n. 1, art. 10, p. 213-230, jan./fev. 2019.

<b>Contribuição dos Autores</b>	<b>G. A. Souza,</b>	<b>A. Orenha-Ottaiano</b>
1) concepção e planejamento.	X	X
2) análise e interpretação dos dados.	X	X
3) elaboração do rascunho ou na revisão crítica do conteúdo.	X	X
4) participação na aprovação da versão final do manuscrito.	X	X