

Os discursos feministas no Brasil e na Alemanha: um estudo contrastivo assistido por corpus sobre suas estruturas temáticas

Feminist discourses in Brazil and
Germany: a contrastive corpus-assisted
study of their thematic structures

Andressa Costa¹

1 Pesquisadora do pós-doutorado no departamento de Linguística aplicada e estudos da linguagem (LAEL) na PUC-SP. acosta.andressa@gmail.com.

Resumo: Esse estudo tem como finalidade analisar e contrastar os discursos de feministas na Alemanha e no Brasil. Os procedimentos metodológicos seguem a abordagem da linguística de corpus. A base de dados compõe-se de dois corpora compilados para esse projeto – um em alemão (feminaDE) e outro em português (feminaBR) - compostos por textos de diferentes registros/gêneros textuais escritos por feministas. A modelagem de tópicos (*topic modeling*) com o modelo LDA (*latent dirichlet allocation*) foi o método usado para identificar temas latentes nos dois corpora. LDA é um modelo probabilístico gerativo cuja ideia básica é que textos são uma mistura de tópicos latentes e cada tópico é caracterizado por uma distribuição de palavras (BLEI; NG; JORDAN, 2003, p. 996). Nesse modelo, as palavras são as variáveis observáveis e a estrutura de tópicos, *clusters* de palavras que coocorrem nos textos, são as variáveis escondidas. As variáveis observadas compõem-se de lemas de substantivos, verbos, adjetivos e advérbios. Os resultados mostram que há mais diferenças do que semelhanças em relação à composição dos tópicos. Mesmo havendo várias palavras-chave comuns nos dois corpora, elas coocorrem com diferentes palavras-chave criando diferentes estruturas temáticas. Por exemplo, apenas um tópico no corpus brasileiro apresenta estrutura temática similar a outros dois tópicos no corpus alemão.

Palavras-chave: Análise do discurso; Linguística de corpus; Estudos culturais; Linguística contrastiva; Feminismo; Língua Alemã; Língua Portuguesa

Abstract: This study aims to analyze and contrast the discourses of feminists in Germany and Brazil. The methodological procedures follow a corpus linguistics approach. The database consists of two corpora compiled for this project – one in German (feminaDE) and the other in Portuguese (feminaBR) – composed of texts from different registers/genres written by feminists. Topic modelling with the LDA (latent Dirichlet allocation) model was used to identify latent themes in the two corpora. LDA is a generative probabilistic model whose basic idea is that texts are a mixture of latent topics and each topic is characterized by a distribution of words (BLEI; NG; JORDAN, 2003, p. 996). In this model, the terms are the observable variables, and the topics, clusters of words that co-occur in texts, are the hidden variables. The observed variables are lemmas of nouns, verbs, adjectives and adverbs. The results show more differences than similarities regarding the composition of topics. Even though there are several common keywords in the two corpora, they co-occur with different keywords creating different thematic structures. For example, only one topic in the Brazilian corpus has a composition similar to other two topics in the German corpus.

Keywords: Discourse analysis; Corpus linguistics; Contrastive studies; Contrastive linguistics; Feminism; German, Portuguese.

1. Introdução

Os discursos feministas estão cada vez mais presentes nos espaços públicos de modo global. Na Alemanha, a presença feminista nos meios de comunicação parece ser mais frequente e ter maior eco do que no Brasil. Apesar disso, nota-se um crescimento de vozes feministas no Brasil nos últimos anos. Esses discursos não são homogêneos, embora haja temas que podem ser considerados típicos feministas como emancipação da mulher, aborto, violência contra a mulher, entre outros. E mesmo esses temas similares parecem ter diferentes pesos e ser abordados de diferentes maneiras nas culturas alemã e brasileira.

O presente estudo apoia-se em duas ideias centrais que relacionam os conceitos de cultura, língua e sociedade: a) a língua tem como funções primárias servir de base para o desempenho de atividades e identidades sociais e servir de suporte para afiliação cultural, institucional e de grupos sociais (GLEE 2005: 1); b) o uso padronizado da linguagem é resultado de atos de fala cooperativos e por isso determinados padrões de uso linguístico indicam determinados aspectos de atos de fala e podem, desse modo, ser interpretados como elementos centrais característicos de discursos (BUBENHOFER 2009: 43).

Partindo-se desses pressupostos, considera-se que os padrões linguísticos identificados neste estudo revelam aspectos que são culturalmente relevantes para feministas brasileiras e alemãs. Assim, o estudo orienta-se pelas seguintes perguntas de pesquisa: a) Quais são os principais temas nos discursos feministas nas culturas alemã e brasileira? b) Que diferenças e similaridades temáticas há nos discursos feministas nessas culturas? A abordagem metodológica combina linguística de corpus, análise do discurso e linguística computacional. Isso implica o uso de corpora como base de dados e técnicas estatísticas para identificar padrões nos textos. O corpus compõe-se de textos em alemão e português escritos por feministas. Esse estudo é o primeiro de uma série de estudos que tem como finalidade analisar e contrastar os discursos de feministas alemãs, brasileiras, britânicas e americanas sob diferentes aspectos a fim de identificar similaridades e diferenças discursivas.

2. Métodos

2.1. Os dados da análise

O presente estudo tem como base um corpus bilíngue, composto por textos escritos por feministas brasileiras e alemãs. Foi compilado em grande parte da internet com a ferramenta BootCat (BARONI/BERNARDINI 2004) e uma parte foi escaneada. A estrutura do corpus é apresentada na tabela 1:

Tabela 1: Estrutura do corpus

Nome do Corpus	FeminaBR	FeminaDE
Registros	Blog, Website, livro, revista, ensaio, entrevista/relatos, cartilha	Blog, Website, livro, revista, ensaios
Tamanho do corpus	286 textos	323 textos
	682.720 tokens	488.786 tokens
Período	de 2003 a 2020	de 2006 a 2020
Etiquetados	TreeTagger (1994)	TreeTagger (Schmid 1995)

A seleção do material foi feita a partir de um levantamento de autoras, organizações e grupos de mulheres que se autodenominam feministas. Definiu-se uma quantidade de no máximo 100 textos de cada registro, o que só foi possível para Websites, blogs e revistas. Por se tratar de corpus especializado, a quantidade de material disponível ainda é escassa para a construção de um corpus maior.

2.2. Modelagem de tópicos

Para a identificação de padrões temáticos usou-se a modelagem de tópicos com LDA (*Latent Dirichlet Allocation*), um modelo probabilístico desenvolvido para modelar dados textuais (BLEI/NG/JORDAN 2003). Segundo BLEI (2012), esse modelo parte do princípio de que cada documento é composto por múltiplos tópicos e cada documento apresenta esses tópicos em diferentes proporções. Nesse contexto, um tópico consiste em um *cluster* de palavras que coocorrem frequentemente nos textos. Os algoritmos, *topic models*, são usados para descobrir os temas principais presentes nos textos (BLEI 2012).

Neste estudo, a modelagem de tópicos foi realizada em R com o pacote *topicmodel* (GRÜN/HORNIK 2011). A unidade de observação foi o corpus todo de cada língua, já que o objetivo desse primeiro estudo é fazer uma comparação intercultural. As variáveis analisadas compõem-se dos lemas de substantivos, adjetivos, verbos e advérbios. A quantidade de tópicos a ser identificada pode ser determinada através de dois métodos:

1) coerência dos tópicos – examinando se as palavras nos tópicos fazem sentido. Assim, pode-se aumentar ou diminuir a quantidade de tópicos para torná-los coerentes.

2) medidas quantitativas – *Log-Likelihood* e *Perplexity*. O objetivo é maximizar *log-likelihood* negativo e minimizar *perplexity* (MCCRACKEN 2020).

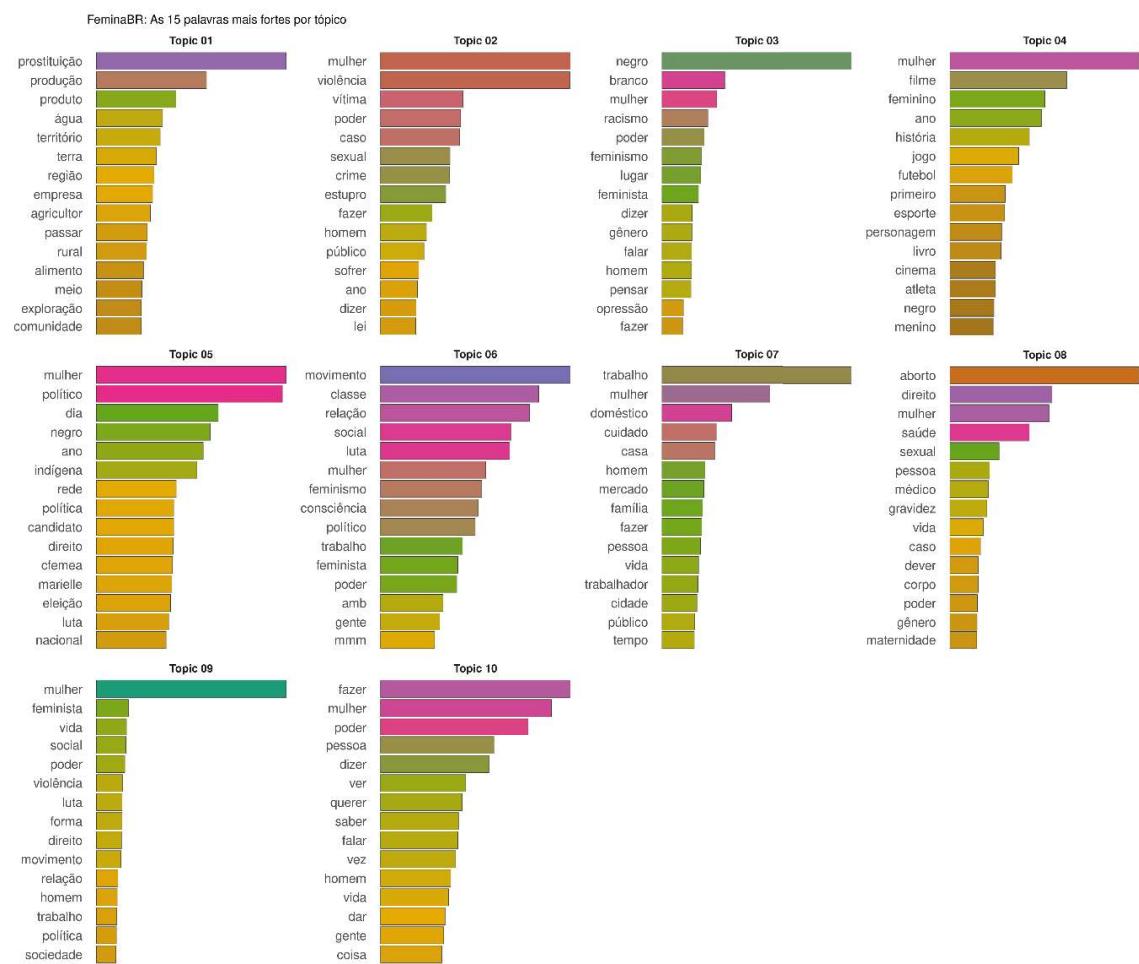
Os métodos quantitativos não foram muito úteis na determinação da quantidade de tópicos, pois os tópicos gerados, considerando-se os melhores índices de *log-likelihood* e perplexidade, não apresentaram uma composição interessante e possível de interpretar. O melhor índice do *log-likelihood* sugere 30 tópicos no corpus alemão (-1536747) e 29 tópicos no corpus brasileiro (-2189362.6), já o melhor índice da perplexidade sugere 27 tópicos no corpus alemão (2279.48) e 29 tópicos no corpus brasileiro (1339.93) como a quantidade mais adequada do ponto de vista estatístico. No entanto, além de ser um número muito elevado de tópicos, vários deles apresentam uma composição muito similar enquanto outros apresentam um conteúdo difícil de interpretar. Desse modo, decidiu-se analisar 10 tópicos, já que a sua composição é mais coerente e possibilita uma interpretação.

3. Resultados e discussão

A análise dos resultados baseou-se nas 15 palavras de maior peso de cada tópico. A identificação dos tópicos foi realizada por meio da modelagem de tópicos em R com o pacote *topicmodels* e teve os seguintes parâmetros:

```
feminaBR_mod = LDA(feminaBR_dtm, k=10, method="Gibbs",
control=list(alpha=0.001, seed=10005, burnin= 500,
delta= 0.1, iter=4000, thin= 100))
```

Nos gráficos 1 e 2 pode-se visualizar as 15 palavras mais marcantes de cada tópico. O gráfico 1 apresenta a estrutura temática do corpus do Português (feminaBR) e o gráfico 2 contém a estrutura temática do corpus alemão (feminaDE). Nos gráficos pode-se observar que há várias palavras-chave comuns aos dois corpora que ocorrem em diferentes constelações. Há também várias palavras específica de cada corpus.

Gráfico 1: as 15 palavras mais fortes por tópicos no corpus brasileiro

O tópico 1 do feminaBR compõe-se de palavras que fazem referência ao trabalho com a terra e à prostituição, trabalho sexual especialmente de mulheres. É o único tópico que não contém o termo ‘mulher’. Todos os outros tópicos fazem referência à mulher e a algum outro aspecto. A interpretação dos tópicos baseou-se na análise dos 15 textos mais representativos de cada tópico. Assim temos os seguintes focos temáticos nos discursos de feministas brasileiras:

Tópico 01: Modelo de desenvolvimento capitalista

Tópico 02: Violência contra a mulher

Tópico 03: Racismo, gênero e o ser feminista

Tópico 04: Cinema, literatura e futebol

Tópico 05: Engajamento político de mulheres

Tópico 06: Feminismo como movimento social

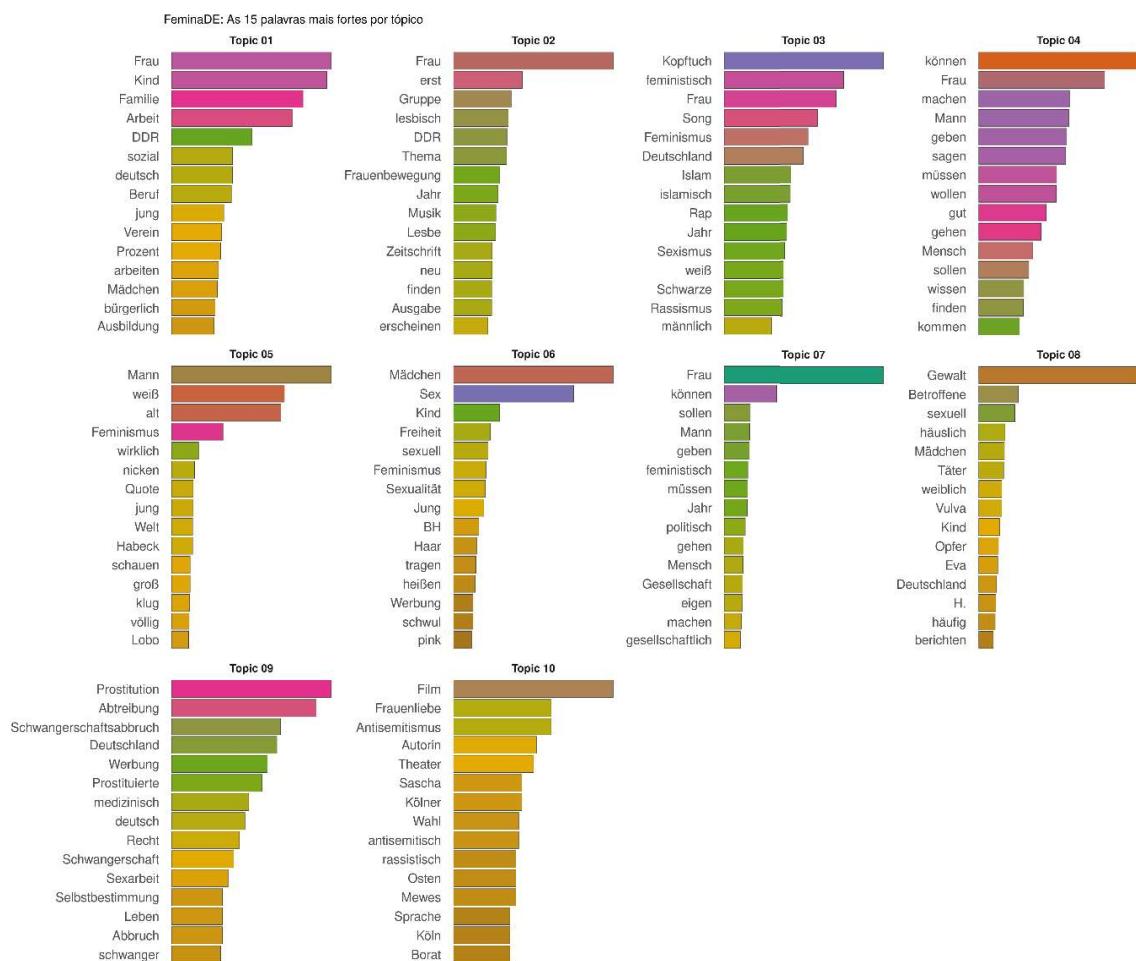
Tópico 07: Trabalho doméstico e mercado de trabalho

Tópico 08: Direitos reprodutivos e saúde da mulher

Tópico 09: Militância feminista

Tópico 10: Sobre pessoas e suas experiências

Gráfico 2: as 15 palavras mais fortes por tópicos no corpus alemão



No gráfico 2 pode-se observar que a palavra *Frau* que significa mulher aparece em 5 dos 10 tópicos entre as palavras de mais peso. Isso é uma das diferenças entre os dois corpora já que no corpus brasileiro a palavra mulher aparece em 9 dos 10 tópicos também entre as palavras de maior peso. Com base nos 15 textos mais representativos de cada tópico do corpus alemão, foram identificados os seguintes temas:

- Tópico 01: Família, educação e trabalho
- Tópico 02: Rede de mulheres, arte e música
- Tópico 03: Sexismo, racismo e feminismo pop
- Tópico 04: Sobre pessoas e suas experiências
- Tópico 05: Patriarcado, cota de mulheres e feminismo
- Tópico 06: Estereótipos sobre feminilidade e masculinidade
- Tópico 07: Engajamento e crítica social
- Tópico 08: Violência sexual e doméstica
- Tópico 09: Prostituição, aborto e autodeterminação
- Tópico 10: Atividades políticas e ações artísticas

Nos gráficos pode-se observar que há várias palavras-chave comuns aos dois corpora que ocorrem em diferentes constelações. Há também várias palavras específica de cada corpus. Ao comparar a estrutura temática dos dois corpora pode-se observar que há apenas 3 tópicos similares: o tópico 10 do feminaBR (Sobre pessoas e suas experiências) tem uma composição bastante similar à dos tópicos 4 (Sobre pessoas e suas experiências) e 7 (Engajamento e crítica social) do feminaDE. Esses tópicos são constituídos por substantivos como mulher/*Frau*, homem/*Mann*, pessoa/*Mensch*, alguns verbos e apresentam um caráter mais geral e neutro em comparação com os outros tópicos, com exceção do tópico 7 do corpus alemão que, além dos termos já mencionados, também contém termos como *feministisch*/feminista, *politisch*/político, *gesellschaftlich*/social. Os tópicos 10 do feminaBR e 4 do feminaDE contém 10 palavras comuns: fazer/*machen*, mulher/*Frau*, poder/*können*, pessoa/*Mensch*, dizer/*sagen*, querer/*wollen*, homem/*Mann*, dar/*geben*, gente/*Mensch*. Os 15 textos mais representativos dos dois tópicos abordam questões da vida em sociedade que mulheres e homens experienciam por conta do fato de serem mulheres e homens como por exemplo: ser mãe solteira ou pai solteiro, conciliar maternidade e profissão, a pobreza das mulheres que vivem sós. Por esse motivo, estes tópicos foram nomeados com o mesmo rótulo.

Outro aspecto interessante que se observa são as coocorrências do termo prostituição/*Prostitution* nos dois corpora. No feminaBR aparece como a palavra-chave mais forte no tópico 1 (Modelo de desenvolvimento capitalista) e coocorre com outros termos relacionados ao trabalho no campo. Já no

feminaDE, o termo *Prostitution* aparece como a palavra-chave mais forte no tópico 9 (Prostituição, aborto e auto-determinação) e coocorre com termos relacionados a gravidez e aborto, como se pode observar na tabela 2:

Tabela 2: Tópicos com coocorrências do termo prostituição/*Prostitution*

FeminaBR Tópico 1	FeminaDE Tópico 9
Prostituição	Prostitution/prostituição
Produção	Abtreibung/aborto
Produto	Schwangerschaftsabbruch/aborto
Água	Deutschland/Alemanha
Território	Werbung/Propaganda
Terra	Prostituerte/prostituta
Região	medizinisch/medicinal
Empresa	deutsch/alemão
Agricultor	Recht/direito
Passar	Schwangerschaft/gravidez
Rural	Sexarbeit/trabalho sexual
Alimento	Selbstbestimmung/Auto-determinação
Meio	Leben/vida
Exploração	Abbruch/interrupção
comunidade	Schwanger/grávida

Com base na análise dos textos mais representativos dos tópicos apresentados na tabela 2, observa-se que, os textos de feministas brasileiras, a prostituição está relacionada a trabalho produtivo e de exploração assim como a trabalho rural. Os textos mais marcados nesse tópico tratam do modelo capitalista e abordam diferentes aspectos que mostram como esse modelo afeta de modo negativo a vida das mulheres através do seu trabalho. Em contrapartida, os textos de feministas alemãs tratam do debate público sobre o aborto, especialmente relacionado a questões legais como os parágrafos 218 e 219 da constituição alemã, ou sobre exploração das mulheres através da prostituição na Alemanha. Isso mostra como o mesmo tema pode ser abordado de maneira diferente em cada cultura.

Outro exemplo é o racismo. Esse aparece no tópico 3 (feminaBR: Racismo, gênero e o ser feminista; feminaDE: Sexismo, racismo e feminismo pop) nos dois corpora, também com certa diferença. No feminaBR, está claramente associado às pessoas negras, mulheres e homens. Em contrapartida, o racismo engloba, no feminaDE, não só as pessoas negras, mas também pessoas muçulmanas, o que se pode observar no gráfico 2 no qual palavras relacionadas ao mundo islâmico têm maior peso nesse tópico.

Por fim, há no corpus alemão referências à DDR (antiga República Democrática da Alemanha) a à própria Alemanha. Contudo, não há referências ao Brasil no corpus brasileiro. Em contrapartida, pode-se considerar o tópico 5 do feminaBR tipicamente brasileiro, uma vez que ele contém as palavras indígena, Marielle, cfemea (organização feminista brasileira) e o tópico 4 também, pois futebol tem uma forte associação com cultura brasileira. O tópico 5 do feminaDE parece ser típico da cultura alemã porque fala do velho homem branco e das cotas (de mulheres em posições de liderança) e o tópico 10 também por conter o termo antisemitismo, algo fortemente associado à cultura alemã. Todos os outros tópicos contêm várias palavras-chave comuns aos dois corpora, mas em diferentes composições, o que dever ser influenciado por questões específicas de cada cultura. Estas questões culturais e históricas necessitam de um estudo mais aprofundado.

Os resultados sugerem que os discursos de feministas nas culturas brasileira e alemão abordam questões relacionadas à condição de mulher na sociedade que se constituem, por um lado, em denúncias contra a violência, preconceitos raciais, sexuais e de gênero, exploração sexual e, por outro lado em crítica social que pode ter como objetivo conscientizar a sociedade sobre ideias preconcebidas (a mulher tem que se comportar segundo regras impostas pela sociedade) e comportamentos preconceituosos (que julgam e condenam mulheres por não seguirem regras impostas). A maior parte dos textos trata da mulher heterossexual e alguns poucos consideram lésbicas e mulheres trans. Apenas no tópico 2 (Rede de mulheres, arte e música) do corpus alemão aparecem as palavras *lesbisch* (adjetivo lésbica) e *Lesbe* (substantivo lésbica) e no tópico 6 (Estereótipos sobre feminilidade e masculinidade) a palavra *schwul* (gay).

4. Considerações finais

O objetivo deste primeiro estudo foi obter uma visão geral sobre os temas abordados nos discursos de feministas alemãs e brasileiras. Os resultados mostraram que os discursos feministas não são homogêneos, apenas três tópicos apresentam similaridade temática: tópicos 4 e 7 do corpus alemão, e tópico 10 do corpus brasileiro.

Com base na análise dos dados, pode-se concluir que, nas duas culturas, o feminismo tem um caráter de engajamento social, político e intelectual. Além disso, nota-se que as duas culturas compartilham temas que, todavia, são abordados em diferentes contextos como vimos com o termo ‘prostituição’ e ‘aborto’. Os tópicos refletem questões centrais influenciadas pelo contexto social, histórico, político de cada cultura.

Referências bibliográficas

- BAKER, PAUL. *Using Corpora in Discourse Analysis*. London, New York: continuum, 2006.
- BARONI, MARCO; BERNARDINI, SILVIA. BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of LREC 2004*. Disponível em: https://home.sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_2004.pdf
- BLEI, DAVID. Probabilistic Topic Models. *Communications of the acm*, vol. 55 no. 4, 2012, pp. 77-84. doi:10.1145/2133806.2133826.
- BLEI, DAVID; NG, ANDREW; JORDAN, MICHAEL. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3: 993-1022, January 2003. Disponível em: <http://www.cs.columbia.edu/~blei/publications.html>.
- BUBENHOFER, NOAH. *Sprachgebrauchsunter. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin, New York: de Gruyter, 2009.
- GEE, JAME P. *An Introduction to Discourse Analysis: Theory and Method*, 2nd Ed. London, New York: Routledge, 2005.
- GRÜN, BETTINA; HORNIK, KURT. topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, Volume 40, Issue 13, 2011, pp. 11-30. Disponível em: <https://www.jstatsoft.org/article/download/v040i13/480>.
- McCRACKEN, CEL. *TED Talks: AI and Topic Modelling*. RPubs, 2020 Disponível em: <https://rpubs.com/CelMcC/645438>
- SCHMID, HELMUT. Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland, 1995.
- SCHMID, HELMUT. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, 1994.
- Scripts da modelagem de tópicos em R: <https://github.com/cristinnebr/feminims.git>