



## Resenhas

**BIBER, Douglas; EGBERT, Jesse. 2018. *Register Variation Online*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316388228>.**

Por/By: Andressa Costa<sup>1</sup>

**Palavras-chave:** *variação de registro; registros online; análise multidimensional; inglês.*

**Keywords:** *register variation; online registers; multidimensional analysis; English.*

Douglas Biber é a grande referência nos estudos de variação de registro. Com a sua obra de 1988 *Variation across speech and writing*<sup>2</sup>, o autor traz uma nova perspectiva para o estudo da variação linguística por meio de corpora. Nesta perspectiva, técnicas estatísticas identificam grupos de características linguísticas que coocorrem sistematicamente nos textos. Considera-se que esses grupos ou fatores compartilham função comunicativa que são identificadas e interpretadas como dimensões de variação. Estudos multidimensionais de variação de registros vêm sendo aplicados a várias línguas como inglês (Biber, 1988; Xiao, 2009), espanhol (Biber et al., 2006), português (Berber Sardinha et al., 2014), coreano (Kim e Biber 1994), Somali (Biber e Hared,

1. Departamento de Linguística Aplicada e Estudos da Linguagem. Pontifícia Universidade Católica de São Paulo. São Paulo, SP – Brasil. <http://orcid.org/0000-0002-1742-7962>. E-mail: [acosta.andressa@gmail.com](mailto:acosta.andressa@gmail.com).

2. Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.



This content is licensed under a Creative Commons Attribution License, which permits unrestricted use and distribution, provided the original author and source are credited.

1994), Nukulaelae Tuvuluan (Besnier, 1988) e Gaelic (Lamb, 2008) tanto para análise da língua geral como para a descrição de registros (gêneros) específicos, como por exemplo, literatura de ficção (Egbert, 2012), cinema e televisão (Berber Sardinha e Veirano Pinto, 2017; Berber Sardinha e Veirano Pinto, 2019; Veirano Pinto, 2014), canções (Bértoli-Dutra, 2014), linguagem acadêmica (Biber, 2006; Gray, 2015) e Internet (Berber Sardinha, 2014; Biber e Egbert, 2016).

Um aspecto central nos estudos multidimensionais de variação de registro é definir quais registros devem ser incluídos na análise. Quando se fala em registros da Internet, geralmente se considera um grupo específico de registros: *blogs*, redes sociais, *websites*, fóruns, e-mails, *chat rooms*. Biber e Egbert mostram no livro *Register Variation Online* (2018), no entanto, que a Internet apresenta uma gama bem mais variada de tipos textuais do que os comumente referidos. Os autores apontam dois problemas relacionados às pesquisas acadêmicas sobre e com base em registros online:

1. Várias publicações acadêmicas que se ocupam com os tipos de textos próprios da web se concentram em textos específicos, mas a web não se restringe a apenas esses tipos de documentos.
2. Sabe-se muito pouco sobre as características linguísticas dos registros online encontrados a partir de sites de busca como o *Google*.

Para os autores, estudar o discurso dos registros da Internet encontrados a partir de buscadores se justifica do ponto de vista teórico porque isso incita repensar a noção tradicional de registro. Enquanto textos escritos publicados em mídia impressa podem ser facilmente descritos e categorizados em um determinado registro por meio de critérios externos, os textos online encontrados por meio de sites de busca na Internet apresentam pouca ou nenhuma característica que os defina claramente como pertencentes a um dado registro. Como consequência, grande parte dos corpora de Internet existentes não representa adequadamente os tipos textuais tipicamente encontrados nesses domínios porque esses corpora compõem-se apenas de registros que são determinados a priori, deixando de fora os documentos que não se encaixam em nenhuma categoria conhecida (cf. Biber e Egbert, 2018:12). Como tese central do livro, Biber e Egbert postulam que a

maioria dos documentos encontrados na Internet não são instâncias puras de um registro específico e que os registros podem às vezes ser interpretados como híbridos e expressam múltiplas intenções comunicativas. O principal objetivo do livro *Register Variation Online* é, portanto, suprir essa lacuna e fornecer uma descrição abrangente da variação linguística nos registros online.

Na introdução, os autores traçam uma breve descrição do desenvolvimento histórico do computador como meio de comunicação, mostrando os períodos e contextos (militar, acadêmico, esfera pública) nos quais os diferentes canais da comunicação mediada por computador (*chat rooms*, *Instant messages*, *websites*, mídias sociais) e os respectivos tipos de texto (e-mails, postagens de *website/blogs*/mídias sociais) surgiram, bem como o processo de expansão da web em termos de usuários e *websites* ao longo dos anos. Além disso, Biber e Egbert fazem uma apreciação dos diferentes repositórios de informações possíveis de se encontrar na internet, porém, acrescentam que grande parte dos textos *online* não são acessíveis ao público geral.

Uma consequência dessa nova forma de comunicação digital é o crescente interesse de linguistas em estudar a linguagem da internet. Biber e Egbert argumentam, no entanto, que grande parte desses estudos enfoca registros especiais que são comumente associados à Internet e à comunicação mediada por computador, tais como: e-mail, *IM messages*, *chat rooms*, interações em ambientes virtuais *online* (MUDs, or *Multi-User Dungeons*), *blogs*, fóruns de discussão, postagem de mídias sociais, entre outros. Eles argumentam ainda que uma busca rápida na Internet mostra que a variedade de registros encontrados nela é mais ampla do que o grupo de registros comumente analisados em estudos linguísticos, mas que ainda não há uma forma simples de determinar o seu conteúdo. Esse problema é ilustrado no livro com alguns exemplos de busca no *Google*.

Finalmente, na última parte da introdução, os autores definem mais precisamente os registros estudados no livro: trata-se de registros encontrados em buscadores da Internet e acessíveis ao grande público. Eles não incluem nem materiais armazenados na *deep web* nem a parte de internet destinada à comunicação privada, tais como mensagens de e-mails, postagens de redes sociais etc. Os autores apresentam os seguintes argumentos para justificar o estudo apresentado na obra:

- Os registros da Internet encontrados através de buscadores (*searchable web*) ocupam um lugar central na sociedade moderna, o que pode ser observado no uso do verbo *google*, que designa a prática extremamente comum de se utilizar o *Google* como buscador para se obter informação na Internet.

- Sabe-se muito pouco sobre as características linguísticas dos registros da Internet encontrados por meio de buscadores. Uma das questões centrais desse estudo é considerar a existência de registros híbridos e realizar análises exploratórias para investigar suas diferentes formas de hibridismo relativas às características situacionais e linguísticas.

- A maioria dos corpora escritos até o momento enfoca exclusivamente registros já estabelecidos tais como: romances, textos acadêmicos, livros de não ficção, artigos de revista e jornal etc., excluindo registros não publicados, porém disponíveis em formato eletrônico na Internet. Já o estudo apresentado no livro enfoca registros encontrados gratuitamente na Internet a partir de motores de busca propondo-se a preencher uma lacuna existente nos estudos relativos a corpora escritos.

No capítulo 2, os autores descrevem o método de coleta do corpus, o método de classificação dos textos em registros e os procedimentos usados para identificar e contar a ampla gama de características linguísticas em cada texto do corpus. O *Corpus of Online Registers of English* (CORE) foi coletado com o auxílio da ferramenta *HTTrack*. Ele contém 48.571 documentos e aproximadamente 54 milhões de palavras. A classificação dos textos da Internet em registros se constituiu, segundo os autores, em uma tarefa difícil por causa da ausência de indicadores externos que possibilitem a categorização clara destes textos. Outros problemas apontados por Biber e Egbert são a falta de concordância sobre os tipos de registros encontrados na Internet e a prevalência de registros híbridos. Esses problemas resultam em tentativas pouco bem-sucedidas em categorizar documentos *online* em registros. Por essa razão, os autores propõem um caminho alternativo com resultados bastante satisfatórios: eles pediram a usuários finais para classificar os textos da Internet em registros com base nas suas funções comunicativas e características situacionais. Em seguida,

realizaram uma série de estudos piloto para testar o quão confiável foi a classificação. No livro, os procedimentos dessa classificação são descritos em detalhe.

As análises do CORE têm como base um grande número de características linguísticas que podem ser fortemente associadas à variação de registros no discurso *online*. O texto é unidade de análise. Dentre as vantagens de usar o texto como unidade de observação, segundo os autores, estão: (1) a possibilidade de se computar frequências normalizadas das ocorrências e realizar estatística descritiva (medidas de tendência central e de variância) para cada registro, o que por sua vez possibilita observar a variação entre os registros e nos registros; (2) a possibilidade de se realizar procedimentos estatísticos mais avançados como a análise multidimensional e as análises das características principais (*key feature analysis*) e de palavras-chave (*keyword analysis*).

No capítulo três, os autores fazem um levantamento dos tipos de registros encontrados por meio de buscadores considerando em que medida cada tipo de texto pode ser associado ao domínio discursivo da Internet. Além disso, eles mostram o resultado da classificação dos registros pelos usuários finais, refletindo, a partir dessa classificação, sobre a complexidade da tarefa e da língua (no caso, o inglês) usada na Internet. Cada documento foi classificado em uma categoria de registro geral e em um sub-registro mais específico. Da análise do material coletado, resultaram quatro grandes registros gerais sob os quais os sub-registros foram agrupados. Esses registros são o foco dos capítulos 4 a 8 do livro.

O capítulo 4 apresenta a análise multidimensional dos registros identificados. Tal análise baseia-se nas características léxico-gramaticais identificados nesse projeto pelo *Biber Tagger*. Essas características estão listadas no apêndice A da obra. Após a etiquetagem dos textos com o *Biber Tagger*, a frequência normalizada das características léxico-gramaticais é computada pelo programa *Biber TagCount*. Essas frequências são usadas em uma análise fatorial para a identificação de agrupamentos de padrões de coocorrência complementar sistemática das características léxico-gramaticais entre os registros. Uma vez identificados, esses agrupamentos são interpretados comunicativa e funcionalmente e nomeados de acordo com a função comunicativa que

os subjazem. Os agrupamentos interpretados e nomeados são denominados dimensões de variação. A partir dessas dimensões, os registros são caracterizados linguisticamente.

O estudo descrito no livro identificou 9 dimensões de variação dos registros da Internet: Oral envolvido *versus* Letrado (Dimensão 1), Elaboração oral (Dimensão 2), Narrativa oracional oral *versus* Informação nominal letrada (Dimensão 3), Comunicação reportada (Dimensão 4), Narração irreal *versus* Narração informacional (Dimensão 5), Discurso procedural/explicativo (Dimensão 6), Posicionamento nominal/letrado (Dimensão 7), Descrição de seres humanos (Dimensão 8), Explicação não técnica ou descrição (Dimensão 9). Por meio dos escores médios atribuídos a cada registro em cada uma das dimensões de variação é possível traçar um perfil multidimensional para cada um dos registros, como por exemplo, as Dimensões 1 a 3 que diferiam entre registros da Internet orais e letrados. Nessas dimensões, registros como documentos líricos, diálogos de televisão e discussões têm um escore médio positivo bastante alto, o que indica que eles têm uma forte caracterização de oralidade. Em contrapartida, registros escritos informativos apresentam escore negativo nas três dimensões indicando ausência de marcação de oralidade e maior tendência para um estilo letrado.

Os capítulos 5 a 8 enfocam apenas as quatro categorias gerais de registros nas quais os sub-registros foram agrupados e descrevem detalhadamente as características linguísticas e funcionais que são distintivas desses registros:

- Registros narrativos da web (capítulo 5).
- Registros de opinião, conselho e persuasão da web (capítulo 6).
- Descrições informativas, explicações e procedimentos (capítulo 7).
- Registros orais da web (capítulo 8).

Para cada registro, Biber e Egbert identificam as palavras-chave<sup>3</sup> (*Keywords*), considerando não só a sua frequência no corpus, mas também sua dispersão nos textos que compõem os registros (*text-dispersion*

---

3. Resultados da comparação entre *corpus* de estudo e *corpus* de referência. Essa comparação elimina palavras com frequência relativa similar nos dois *corpora* de modo que restem as palavras cuja frequência é estatisticamente significativa. As que restam no corpus de estudo são denominadas 'palavras-chave positivas', enquanto as que sobram no corpus de referência, 'palavras-chave negativas.' (Tagnin, 2010: 359).

*keyness*). Esse método de medir a dispersão de palavras-chave pelos textos do corpus foi desenvolvido pelos autores para esse projeto e visa a medir cada palavra em relação a quantidade de textos nos quais elas ocorrem no *corpus* de referência e no *corpus* de estudo. Além disso, os autores descrevem as características léxico-gramaticais chave (*Key Feature Analysis*) nos registros considerando tanto as mais frequentes quanto as menos frequentes no corpus de estudo para identificar diferenças significativas entre um ou mais registros individuais e o restante do CORE. O método de análise das características léxico-gramaticais chave usado no projeto é explicado no subcapítulo 2.4 do livro. Finalmente, o capítulo 9 apresenta uma síntese dos achados do estudo, uma descrição do estado atual e futuro da pesquisa na área além de uma discussão acerca das implicações teóricas dos resultados desta pesquisa para estudos da variação de registro em outros domínios discursivos.

O livro *Register Variation Online* é uma obra de referência para aqueles que pesquisam sobre o inglês usado na Internet, e também para quem quer conhecer uma nova metodologia para estudar a língua em uso na Internet a partir de *corpora*. O estudo traz grandes contribuições tanto metodológicas quanto teóricas para a linguística de *corpus* e estudos de variação textual. No aspecto metodológico, ele é inovador ao desenvolver duas técnicas de medir a chavicidade (*keyness*)<sup>4</sup> que vão além da simples extração de palavras-chave. Grandes contribuições teóricas são: a descrição detalhada de registros online que até então não eram considerados nas pesquisas sobre registros da Internet e o reconhecimento da existência de registros híbridos tanto em relação às suas características linguísticas quanto situacionais.

## Referências

BERBER SARDINHA, Tony. 2014. 25 years later: Comparing Internet and pre-Internet registers. In: BERBER SARDINHA, Tony; VEIRANO PINTO, Marcia. (Eds.). *Multi-Dimensional Analysis, 25 years on. A tribute to Douglas Biber*. Amsterdam/Philadelphia: John Benjamins, 82-105.

4. Medida estatística que indica o quão significante é a frequência de uma palavra em um corpus específico a ponto de ela ser uma palavra-chave daquele corpus (nota da resenhista).

- BERBER SARDINHA, Tony; KAUFFMANN, Carlos; ACUNZO, Cristina M. 2014. Dimensions of Register Variation in Brazilian Portuguese. In: BERBER SARDINHA, Tony; VEIRANO PINTO, Marcia. (Eds.). *Multi-Dimensional Analysis, 25 years on. A tribute to Douglas Biber*. Amsterdam/Philadelphia: John Benjamins, 35-79.
- BERBER SARDINHA, Tony; VEIRANO PINTO, Marcia. 2017. American television and off-screen registers: A corpus-based comparison. *Corpora*, 12/1, 85-114.
- BERBER SARDINHA, Tony; VEIRANO PINTO, Marcia. 2019. Dimensions of variation across American television registers. *International Journal of Corpus Linguistics*, 24/1, 3-32.
- BÉRTOLI-DUTRA, Patrícia. 2014. Multi-Dimensional analysis of pop songs. In: BERBER SARDINHA, Tony; VEIRANO PINTO, Marcia. (Eds.). *Multi-Dimensional Analysis, 25 years on. A tribute to Douglas Biber*. Amsterdam/Philadelphia: John Benjamins, 149-175.
- BESNIER, Niko. 1988. *The linguistic relationships of spoken and written Nukulaelae registers*. *Language* 64 /4: 707-36.
- BIBER, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- BIBER, Douglas; DAVIES, Mark; JONES, James K.; TRACY-VENTURA, Nicole. 2006. Spoken and written register variation in Spanish: A multi-dimensional analysis. In: *Corpora* 1/1: 1-37
- BIBER, Douglas. 2006. *University Language: A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins.
- BIBER, Douglas; EGBERT, Jesse. 2016. Register Variation on the Searchable Web: A Multi-Dimensional Analysis. In: *Journal of English Linguistics* 44/2, 95-137.
- BIBER, Douglas; HARED, Mohamed. 1994. Linguistic correlates of the transition to literacy in Somali: language adaptation in six press registers. In: BIBER, Douglas e FINEGAN, Edward (Eds.) *Sociolinguistic Perspectives on Register*. Oxford: Oxford University Press, 182-216.
- EGBERT, Jesse. 2012. Style in nineteenth century fiction: A Multi-Dimensional Analysis. In: *Scientific Study of Literature* 2/2. Amsterdam/Philadelphia: John Benjamins, 167-198.
- GRAY, Bethany. 2015. *Linguistic Variation in Research Articles: When Discipline Tells Only Part of the Story*. Amsterdam/Philadelphia: John Benjamins.
- KIM, Yong-Jim; BIBER, Douglas. 1994. A corpus-based analysis of register variation in Korean. In: BIBER, Douglas e FINEGAN,



- Edward (Eds.) *Sociolinguistic Perspectives on Register*. Oxford: Oxford University Press, 157–181.
- LAMB, William. 2008. *Scottish Gaelic Speech and Writing: Register Variation in an Endangered Language*. Belfast: Cló Ollscoil na Banríona.
- TAGNIN, Stella. 2010. *Glossário de linguística de corpus*. São Paulo: HUB Editorial. Disponível em: [http://www.hubeditorial.com.br/site/recursos/5\\_glossario/glossario\\_423.pdf](http://www.hubeditorial.com.br/site/recursos/5_glossario/glossario_423.pdf). Acesso em: 13 set. 2018.
- VEIRANO PINTO, Marcia. 2014. Dimensions of variation in North American movies. In: BERBER SARDINHA, Tony; VEIRANO PINTO, Marcia. (Eds.). *Multi-Dimensional Analysis, 25 years on. A tribute to Douglas Biber*. Amsterdam/Philadelphia: John Benjamins, 109-147.
- XIAO, Richard. 2009. Multidimensional analysis and the study of world Englishes. In: *World Englishes* 28/4, 421-450.

Recebido em: 30/07/2019

Aprovado em: 06/02/2020