



A review of Register Variation Online (Biber & Egbert, 2018) — Cambridge University Press

Resenha de Register Variation Online (Biber & Egbert, 2018) — Cambridge University Press

Larissa Goulart¹

orcid.org/0000-0002-4190-4589
lg845@nau.edu

Received on: Aug. 12th, 2019.

Approved on: May 03rd, 2020

Published on: Jan. 11th, 2021.

1. Summary

Register variation online is aimed at language researchers interested in registers available online. The study describes register variation in the searchable web, in other words, texts that are publicly available when doing a Google search. Biber and Egbert (2018) argue that previous studies in web registers have focused on registers that emerged online and that only exist in the web, for example, tweets or Facebook posts. Nevertheless, when conducting a Google search on a random word -horse- they found that these registers are not the ones most commonly found by internet users. They also argue that previous research has focused on analyzing only one registers at a time. Therefore, they set out to investigate the different registers that emerge from a Google search, in order to identify their linguistic patterns. To do so, they conducted a multi-dimensional (MD) analysis in a corpus of online registers. Multi-dimensional analysis (Biber 1988) is an approach in which constellations of linguistic features are identified based on statistical co-occurrence patterns in texts. These patterns are interpreted as 'dimensions' of variation that are associated with the shared communicative functions of the co-occurring features. MD analysis demonstrates that linguistic features do not co-occur by chance, but rather because of shared underlying communicative functions that in turn correspond to shared situational features of the texts (Goulart et al. 2020). The authors also use keyword and key feature analysis to explore the register differences encountered online.

In chapter 1, Biber and Egbert (2018) motivate their study by first discussing how the internet has spread in the past years and how, consequently, different websites have emerged as more people have access to the internet. They mention that, as a consequence of the proliferation of new websites, there is a range of new registers that have not been studied and that might not be similar to printed registers. Thus, there is a gap in this field of research. The authors also argue that there are more web pages accessible online than there are items in the British Library and the Library of Congress; hence, it is important to study the linguistic characteristics of these texts that are readily available for users to access.



Artigo está licenciado sob forma de uma licença
[Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).

¹ Northern Arizona University (NAU), Flagstaff, AZ, US

In chapter 2, the authors describe the corpus used in the study, CORE - Corpus of Online Registers of English. CORE is a sample from the Global Web-based English (GlobWbE). The latter contains 1.9 billion words from 1.8 million Web documents. GlobWbE contains texts of different lengths, but for CORE, the authors excluded texts with fewer than 75 words. The average text length was 1.000 words. The texts that composed the corpus came from a Google search of 3-grams (combination of three words) that were among the most frequent in COCA.

In chapter 3, the texts classification into registers is described. The researchers used Mechanical Turks in order to classify the texts collected online into registers. This classification was first piloted several times in order to create a rubric that could be used by the software to perform the intended activity.

Chapter 4 discusses the results of the MD for all registers identified online. The MD included 57 linguistic features. As previously mentioned, they also conducted a keyword and key feature analysis. Nevertheless, the keyword analysis is presented only for registers that showed no variation in the MD analysis. Based on the MD, the authors have identified 10 dimensions of variation, with one of them being not readily interpretable. The dimensions identified were oral-literate, oral elaboration, oral clausal narrative vs nominal information, irrealis vs literate nominal information, reported communication, irrealis vs informational narration, procedural/explanatory discourse, nominal/literate stance, description of humans and non-technical explanation of description. Interestingly, the first three dimensions show the same loading patterns for almost all registers.

Chapter 5 discusses in detail the results of the MD for narrative registers. This is the most representative register in the corpus, and it is composed of many sub-registers - news report, personal blog, sports reports, historical articles, travel blogs, short stories, and other narratives. A surprising finding in this register is that news report and short stories shared a significant amount of linguistic features. By indicating that even though these registers have different communicative purposes, they share features of the same style of writing.

Chapter 6 analyzes opinion, advice and persuasion registers. The sub-registers in these registers are opinion blog, review, description with intent to sell, advice, religious blog/sermon, other opinion/persuasion, an interesting finding here is that opinion blogs scores near 0 in almost all dimensions. Since the MD did not show significant difference between sub-registers, the authors have also analyzed the keywords in this register. In this case, opinion blogs showed a difference between the keywords used when compared to other sub-registers. Opinion blogs rely on words relating to the status of knowledge, communicative acts, and general evaluation. Not surprisingly reviews and description-with-intent-to-sell also share many common features.

Chapter 7 looks at the informational descriptions, explanations and procedures, which contains varied sub-registers from how-to instructions to informational blogs. An interesting result from this, and most of the previous registers, is that the MD results show great variation among the sub-registers, even though their major communicative purpose is the same. These registers also have sub-registers that could have appeared in printed media too as recipes and research articles.

Chapter 8 deals with the oral registers, the sub-registers in this group represent informal language, even though they do not share the same communicative purpose. Almost all the sub-registers loaded similarly in nearly all dimensions, with the exceptions of lyrical, discussion, and interviews, which use linguistic features differently than the other sub-registers in the whole corpus.

Chapter 9 makes a case for readers to see web registers as a continuum, rather than discrete categories. Based on the classification of the texts, some hybrid registers were identified. In this chapter the authors also approach the texts using a cluster analysis to show how they fall in this continuum.

2. Evaluation

Biber and Egbert (2018) did an excellent work by describing the registers available online and discussing how the different communicative purposes motivate language variation in the use

of grammatical and lexical features. When it comes to the criteria used for the corpus compilation, however, it is unclear whether their goal was to explore registers that users encounter/read/produce the most, or registers that were available to them at the time of the research. If the goal was to analyze registers users read or produce the most, the authors do not provide evidence that their corpus met this criterium. On the other hand, if the goal was to analyze registers that were available at the time of the search, the reader might be interested in knowing how prevalent these registers are for internet users. From a user standpoint, it seems that – from the registers studied in this book – people access (to read or write) more often news reports and reviews, but most of the other texts read and written on the internet occur in websites that are not searchable, as Facebook, Twitter, Instagram, etc, and that were not included in this study. Furthermore, the criteria used for corpus compilation was a web search of 3-grams, but the authors do not specify if these 3-grams were all the most frequent 3-grams in COCA or if they were a sample of the most frequent 3-grams. While I do not believe this would have a great impact in the final corpus, it would be interesting to know for those of us who might want to conduct a similar research in other languages. Another issue to consider about corpus compilation is the use of Google as the search engine. Even though Google is the most popular search engine available right now, as a reader I would like further explanations for the reasons why not to use Bing or Yahoo Search, for example, especially because the authors state that the search algorithm used by Google plays a role on the results.

The main corpus, GlobWeB, from which CORE was extracted contained texts from twenty different English speaking countries, but CORE was sampled from only five countries, the US, UK, Canada, Australia and New Zealand, which constrains the varieties of English present in this study. This exclusion of other varieties of English might also impact the results, and might restrain the results extrapolation to teachers, for instance, who want to talk about registers online from an

English as a Lingua Franca perspective.

As expected, the definition of registers in the survey of online texts is somewhat confusing. While narratives, informational descriptions/explanations, opinions, etc, are analyzed as purposed, it could also be argued that these represent its communicative purpose, and that the sub-categories in each of them are registers of their own right. This also becomes clearer with the MD analysis of each sub-register that shows the variation between them. Another point to be considered is the “blog” category. It is worth considering what a blog is because – to a certain extent – everything on the internet could be a blog, but it could also be an online magazine, or an online newspaper – that has no printed version. One example we have of these hybrid registers is, for instance, the blog/newspaper Sul21 (<https://www.sul21.com.br/>), which started as a blog and has developed to be considered a newspaper. Although this example is very particular to the south of Brazil, there are other instances of websites that represent hybrid registers.

Even though the corpus compilation and classification seems to have some issues, the authors do a spectacular job of describing the language of the internet using a broad set of linguistic features and taking into account many registers. Although this book is not intended for teachers of English, they could benefit greatly from the descriptions of language provided in this book. For language researchers this is a groundbreaking study as the authors describe in detail the linguistic features that are prevalent in different registers, using not only an MD approach, but also key feature analysis to compare different registers.

References

- Biber, D. 1988. *Variation across speech and writing*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511621024>.
- Biber, D., & Egbert, J. 2018. *Register variation online*. New York: Cambridge University Press. <https://doi.org/10.1017/9781316388228>.
- Goulart, L., Gray, B., Staples, S., Black, A., Shelton, A., Biber, D. ... Wizner, S. in press. Linguistic perspectives on register. *Annual Review of Linguistics*, v. 6. <https://doi.org/10.1146/annurev-linguistics-011718-012644>.

Mailing address:

Larissa Goulart

Northern Arizona University

1899 San Francisco St, 86011,

Flagstaff, AZ, United States of America