EDITORIAL

What is a register?

Accounting for linguistic and situational variation within – and outside of – textual varieties

Douglas Biber & Jesse Egbert
Northern Arizona University

Empirical studies of register variation have established the existence of functional correspondence between situation/context and language use. However, previous conceptualizations of register cannot adequately account for empirical findings which have revealed (i) situational and linguistic variation among texts within registers and (ii) texts that do not belong to a register. We propose an alternative conceptualization in which registers are culturally-recognized categories, as opposed to scientifically-defined categories. This allows us to describe registers for their typical characteristics as well as the variation among texts within register categories. It also allows us to account for the functional correspondence of texts that exist outside of register categories.

1. Introduction

It seems obvious that a clear definition for the term/concept of 'register' is the required foundation for a journal dedicated to 'register studies'. And so it is not surprising that the opening editorial in the inaugural issue of *Register Studies* begins with a section on 'Defining register and register variation' (Gray & Egbert 2019: 1–3). That discussion successfully launched the journal and set the stage for the numerous studies of registers and register variation published over the last several years.

The Gray and Egbert discussion is one of the most recent in a very long history of published papers that introduce and define the construct of 'register', usually in situational terms; for example:

> The term *register*, then, refers to text varieties that are defined by the situational characteristics of a text […] (Gray & Egbert 2019: 1)

https://doi.org/10.1075/rs.00004.bib | Published online: 7 March 2023
Register Studies 5:1 (2023), pp. 1–22. ISSN 2542-9477 | E-ISSN 2542-9485

The textbook *Register, Genre, and Style* (Biber & Conrad 2019) defines 'register' in a very similar way:

> A *register* is a variety associated with a particular situation of use (including particular communicative purposes) (Biber & Conrad 2019: 6)

Interestingly, though, the Biber and Conrad (2019) framework largely sidesteps the question of 'what is a register?', instead focusing on register as an analytical perspective; for example:

> we regard genre, register, and style as different approaches or perspectives for analyzing text varieties, *not* as different kinds of texts or different varieties (Biber & Conrad 2019: 15)

Earlier research similarly expressed the general conceptualization of 'registers' as situationally-defined text varieties, dating back at least to Biber (1988).[1] For example:

> ['registers' or 'genres' are] text categorizations made on the basis of external criteria relating to author/speaker purpose (Biber 1988: 68)

and

> ['registers' or 'genres' are] "categorizations assigned on the basis of external criteria", which include communicative purpose, relations among participants, mode, etc. (Biber 1988: 70)

This general conceptualization can be traced back to Reid (1956: 32), who introduces the notion of 'register' to describe the ways in which a person speak[s] (or write[s]) differently according to what may roughly be described as different social situations. Ferguson (1981: 11) similarly describes 'register variation' as "variation in structure […] depending on use".

In more general terms, this conceptualization was repeated by Halliday (1968: 141), who defined a 'register' as a "variety according to use", as opposed to a dialect, which is a "variety according to users". Halliday (1968: 149) goes on to note that "'registers' account for what people do with their language", and "differences in the type of language selected as appropriate to different types of situation".

---

1. The 1988 study has the added complication that it used the cover term 'genre' to refer to the textual varieties of interest. Biber and Conrad (2019) develop a framework that distinguishes between the 'register' perspective and the 'genre' perspective. But for our purposes here, the important point to emphasize is that these text varieties were defined in situational / communicative terms.

Although other early publications by Halliday conform to this general conceptualization, more recent work in the Systemic-Functional tradition argues that 'register' should be defined in linguistic (as well as situational) terms. For example:

> the linguistic features which are typically associated with a configuration of situational features – with particular values of the field, mode and tenor – constitute a REGISTER (Halliday & Hasan 1976: 22)

> A register is a cluster of associated features having a greater-than-random (or rather, greater than predicted by their unconditioned probabilities) tendency to co-occur (Halliday 1988: 162)

> register is a pattern of linguistic choices (Martin 2002: 57)

And in other publications, SFL scholars like Halliday, Hasan, and Matthiessen argue that 'register' should be defined in semantic terms:

> A register is a semantic concept. It can be defined as a configuration of meanings that are typically associated with a particular situational configuration of field, mode, and tenor. (Halliday & Hasan 1985: 38–39)

> Registers are the semantic configurations that are typically associated with particular social contexts, defined as we have defined them in terms of field, tenor, and mode. (Halliday & Hasan 1985: 42)

> Register variation is thus semantic variation [...] (Matthiessen 2019: 15)

This selective survey of previous research shows that there is disagreement regarding the specific definition of 'register'. However, despite the differences, there is one overarching characteristic shared by all of these definitions: they all entail the implied claim that 'registers' can be defined in terms of contextual (situational and communicative) or linguistic characteristics. As a result, such definitions logically preclude the possibility of variation among texts within a register, at least with respect to the defining characteristics.

Against this background, our main goals in the present editorial are to: (1) survey empirical research that documents the existence of linguistic and situational variation among texts within registers; (2) propose an alternative conceptualization of 'register' that accounts for the existence of register-internal variation; (3) discuss an unanticipated consequence of this alternative conceptualization: the existence of texts that do not belong to any register category; and, finally, (4) advocate for the description and interpretation of register-internal variation as an important component of all future studies of register variation.

2. Background: The TxtLx (Text-Linguistic) approach to register variation

As background to our argument, we are assuming the theoretical foundation for register studies developed in Biber and Conrad (2019), referred to as the TxtLx (Text-Linguistic) approach in more recent works (see, e.g., Biber 2019; Biber, Egbert, Keller & Wizner 2021b). There are two distinctive characteristics of the TxtLx approach:

1. Register descriptions are based on quantitative linguistic analysis of the individual texts contained in a corpus
2. Patterns of register variation are interpreted in terms of *functional correspondence*

TxtLx studies employ corpus-based analyses, with the explicit recognition that a corpus is a sample of texts that has been deliberately designed and collected to represent the target register (see Egbert, Biber & Gray 2022). The underlying theoretical claim is that texts are naturally-occurring units of communication (see Egbert & Schnur 2018), in contrast to a corpus, which does not exist until a researcher compiles a collection of texts. As a result of this perspective, each text is treated as an observation in TxtLx research designs. This methodological decision has major implications for the quantitative linguistic analyses: Rates of occurrence are computed for each linguistic feature in each text. Subsequently, the overall mean rates of occurrence are computed for all texts from a register, coupled with a computation of dispersion (usually a standard deviation), reflecting the extent to which there is linguistic variation among the texts within a register. Thus, the TxtLx research design directly enables the theoretical perspective that we are advocating in the present editorial: a descriptive focus on the variation among texts within a register.

The theoretical foundation of the TxtLx framework is the relationship among the three components of situation, function, and linguistic forms, illustrated in Figure 1 (see Biber & Conrad 2019: 6–10; see also Egbert & Biber 2016; Biber 2019). We refer to this relationship as 'functional correspondence', which is built on the theoretical claim that linguistic features are frequent and pervasive in texts from a register because they are functional – serving communicative functions related to the situational context of the register.

In previous studies, a TxtLx register description begins with analysis of the situational characteristics of the register, including consideration of the participant characteristics, relations among participants, channel, production circumstances, setting, and communicative purposes (see Biber 1988: 37–47; Biber 1994; Biber & Conrad 2019: 39–48). In past studies, the situational analysis of a register

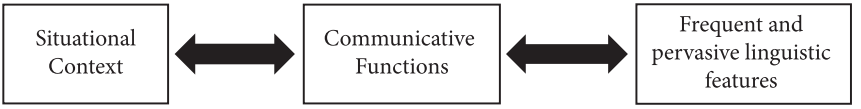

Figure 1. Visual representation of functional correspondence: The three-way relationship among situation, function, and linguistic characteristics

is based on four kinds of information: the analyst's own personal experiences and observations, interviews with experts who regularly produce texts from the register, previous descriptions of the register, and direct consideration of individual texts from the register (see Biber & Conrad 2019: 37–39). The second major step in a TxtLx register description is analysis of the lexico-grammatical features found in texts representing the register, including both a detailed linguistic analysis of tokens of the target feature, as well as a quantitative analysis of the distribution of linguistic features across texts. Finally, those linguistic/quantitative patterns of variation are interpreted functionally, based on the assumption that the linguistic characteristics of a register are derivative from the situational characteristics. The theoretical basis of this 'functional' interpretation in the TxtLx framework goes well beyond conventional or indexical associations between linguistic features and situational contexts. Rather, the underlying assumption is that linguistic features and variants are directly functional; that is, certain linguistic features are frequent and pervasive in texts from a register because they are required by the situational/communicative characteristics of the register, including the production circumstances (e.g., planned/edited versus real-time), interactivity, extent of shared space/time, and communicative purposes (see the discussion in Biber & Conrad 2019: 69–74).

The TxtLx methods for the quantitative analysis of linguistic features in texts and registers are well established (see, e.g., Biber & Conrad 2019: Chapter 3; and the survey of register studies in Barbieri & Wizner 2019). In particular, linguistic rates of occurrence are computed for each individual text in a corpus, enabling description of the dispersion of a linguistic feature across the texts of a register, in addition to the average rate of occurrence. That is, in a TxtLx design, registers can be compared for their average linguistic rates of occurrence (the extent to which they use a linguistic feature), and they can also be analyzed for their internal linguistic variation (the extent to which there are linguistic differences among the texts within a register).

In contrast to the methods for linguistic analysis, there has been less attention given to the methods for the situational analysis of registers in TxtLx studies. As a result, past situational analyses of a register have usually been limited in two key respects:

1. The situational analysis is usually based on discrete categories and characteristics
   a. Unlike quantitative linguistic analyses (based on the rate of occurrence of a feature in a text), situational analyses are usually based on the simple presence or absence of a characteristic (e.g., interactive, planned, informational purpose)
   b. Thus, in most situational analyses, there has been no consideration of the *extent* to which a text/register is interactive/planned/informational/etc.

2. The situational analysis is usually generalized to an entire register
   a. Unlike quantitative linguistic analyses, which are based on analysis of linguistic features in each text, situational analyses are usually carried out for an entire register
   b. Thus, linguistic analyses have been based on the expectation that texts within a register vary for the extent to which they use each linguistic feature. In contrast, situational analyses have been based on the assumption that all texts within a register share the same situational characteristics.

In summary, the situational analyses carried out in register studies have typically employed an inherently different and, we argue, weaker methodological approach than the linguistic analyses: the situational analyses are usually not empirical or replicable; they disregard the possibility that situational characteristics can be continuous, gradient attributes; they are usually based on general characteristics of an entire register rather than detailed situational analyses of each text; and they therefore disregard the possibility that the texts within a register might vary in their situational characteristics. As we show in the following section, recent empirical analyses of these considerations led to the need for an alternative conceptualization of 'register'.

3. Empirical research analyzing patterns of variation among texts within registers

Before we survey the empirical evidence relating to register-internal variation, it is important to emphasize that our goal in this editorial is not to challenge the utility or validity of previous TxtLx studies that have described and compared the typical characteristics of registers. Rather, we treat two findings from previous TxtLx research as the well-established foundation for our discussion here: (1) registers differ from one another in important and systematic ways with respect to their typical contextual and linguistic characteristics, and (2) these systematic

patterns exist because linguistic variation and situational variation exhibit functional correspondence: linguistic features are prevalent (or not) in a register as a direct reflection of the typical contextual characteristics of the register. The usefulness of the TxtLx framework for analyzing and interpreting contextual, linguistic, and functional differences between registers has been well documented (see, e.g., Biber & Conrad 2019, especially Appendix A; Goulart, Gray, Staples, Black, Shelton, Biber, Egbert & Wizner 2020). Biber (2012) showed that register is one of the most important predictors of linguistic variation, and Biber (2014) proposed a set of functional linguistic dimensions that are universal in registers across languages. In short, there is ample evidence to support the existence of registers and the functional correspondence between the contextual and linguistic characteristics of the texts that belong to them.

So, the main point of the present editorial is not that we should abandon TxtLx analyses of register variation. Rather, our point here is to focus on a complementary perspective: the existence of variation among texts within registers. This second perspective does not contradict the first perspective. Rather, registers can (and do) differ from one another in their typical characteristics, while at the same time permitting variation among texts within the register. The problem is that previous research has tended to give primary attention to the typical characteristics of registers, while mostly disregarding description of the variation among texts within registers.

This disregard of register-internal variation applies especially to the situational/communicative analyses of registers (see below), but to some extent, it also characterizes previous linguistic analyses of registers. Researchers working in the TxtLx framework might take exception to this characterization, because the quantitative-statistical analyses of linguistic features in a TxtLx study usually involve both consideration of mean scores and standard deviations. Mean scores are used to describe the typical quantitative-linguistic characteristics of a register, but differences in the mean scores across registers are evaluated relative to standard deviations, representing the extent of variation among texts within each register. Thus, in this sense, TxtLx studies have regularly paid attention to the variation among texts within registers. This is in fact one of the major strengths of TxtLx research designs, where each text represents one observation for the purposes of statistical analysis.

However, even though quantitative-linguistic variation among texts within a register has been a key component of the statistical analyses in TxtLx studies, researchers have rarely described and compared registers in those terms. That is, it has been rare to descriptively compare registers for the extent to which they exhibit internal linguistic variation, and rare to interpret those patterns of variability in functional terms.

One early exception to this generalization is Biber (1988), which devoted an entire chapter (Chapter 8) to the description of linguistic variation among texts within registers (referred to as 'genres' in the 1988 study). This analysis showed that there are important differences among registers in the extent to which they permit internal linguistic variation among texts. For example, academic prose is a register that exhibits a wide range of linguistic variation among texts, while personal letters is a register that exhibits relatively little internal linguistic variation among texts (see Biber 1988: 171–180). Figure 2 below illustrates these differences with respect to Dimension 3 ('Explicit versus Situation-dependent reference') from the 1988 multi-dimensional analysis.

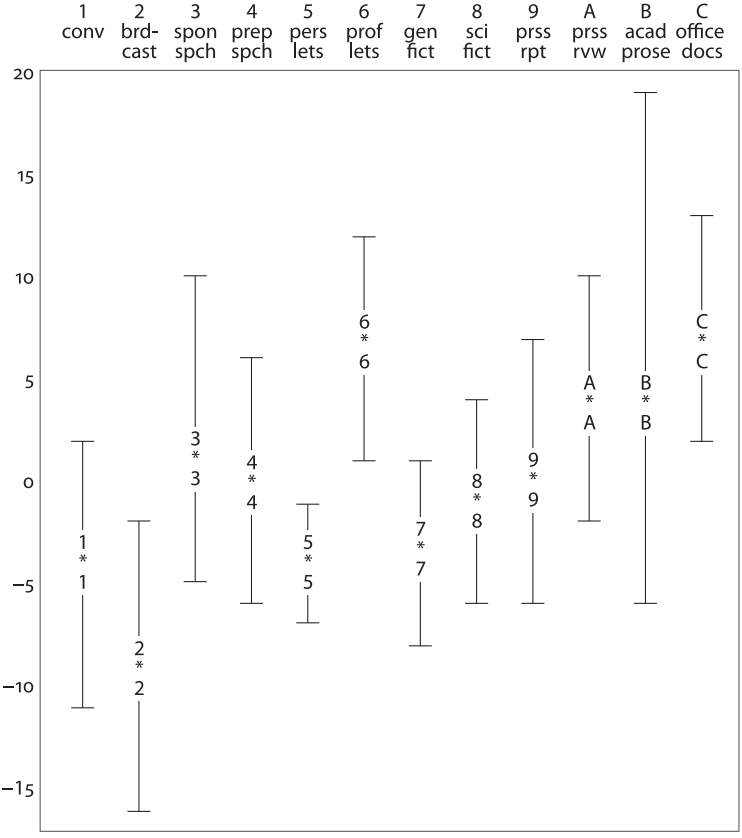

Figure 2. Spread of scores along Dimension 3 ('Explicit versus Situation-dependent Reference') for selected registers (* marks the mean score of each register)

Chapter 8 in Biber (1988) goes beyond the statistical analysis of register-internal variation, by discussing and interpreting the differences in variability across registers in functional terms (see p. 171ff). For example, the chapter discusses how some general registers – like academic prose – can exhibit extensive internal variation among texts because the register actually consists of several distinct sub-registers (e.g., science research articles versus humanities books). In contrast, the chapter discusses how "some other [registers] are simply not well-constrained or defined. For example, conversation shows large ranges on most dimensions, even though there are no clear-cut [sub-register] distinctions within conversation" (Biber 1988: 171).

The disregard of register-internal variation is even more evident when we consider previous analyses of the situational/communicative characteristics of registers. As we noted at the end of Section 2, this disregard is a natural consequence of the completely different methodologies used for the analysis of situational/communicative characteristics: relying on discrete categories (rather than continuous measures) and attempting to characterize entire registers (rather than analyzing the situational/communicative characteristics of each text). These methodological practices have essentially precluded the possibility of observing situational/communicative variation among texts within a register.

However, a few recent studies have applied a text-based approach to the analysis of situational/communicative characteristics, providing strong evidence for the existence of communicative variation among texts within registers. For example, Gray (2015) undertook a description of variation among the different types of academic research articles, distinguishing among six disciplines (e.g., history and biology) and three major research orientations (quantitative, qualitative, and theoretical). As part of that study, Gray also analyzed contextual characteristics of each of the 270 texts in her corpus, identifying register-internal variation for some characteristics (e.g., whether research articles in a discipline explicitly state the research questions).

Three other recent studies take this type of analysis a step further, coding each text for a range of communicative characteristics with the primary research goal of exploring the extent of register-internal variation: Goulart, Biber and Reppen (2022) focused on variation within university student registers; Biber, Egbert, Keller and Wizner (2021a) focused on variation within face-to-face conversation; and Biber, Egbert and Keller (2020) focused on variation within web registers.

The first of these studies, Goulart et al. (2022), coded 308 university student papers for their communicative purposes (e.g., to argue/express an opinion, to compare, to describe a tangible object, to give a procedural recount, to narrate a personal event, etc.). The coding further distinguished between the major purpose of a text versus the existence of other minor purposes. The goal of the analysis was

to compare the communicative characteristics of five major registers[2] of student writing, based on analysis of texts in the BAWE Corpus (essays, critiques, case studies, methodology recounts, and explanations; see Nesi & Gardner 2012).

Table 1. Major communicative purposes of texts within student writing registers (taken from Goulart et al. 2022, Table 6; numbers in parentheses give the actual frequencies of texts; percentages show the proportion of texts within each register)

	To explain	To argue	To propose	To give a procedural recount	To compare	To narrate
Essay (201)	48.5% (98)	47.3% (95)	.5% (1)	–	3.5% (7)	–
Critique (41)	36.5% (15)	46.3% (19)	7.3% (3)	7.3% (3)	–	2.4% (1)
Case Study (21)	47.6% (10)	19.0% (4)	28.5% (6)	4.7% (1)	–	–
Methodology recount (22)	9.0% (2)	9.0% (2)	4.5% (1)	77.2% (17)	–	–
Explanations (23)	86.9% (20)	4.3% (1)	–	4.3% (1)	4.3% (1)	–

The results showed that all five student registers exhibited considerable variation in purpose among texts. For example, Table 1 shows that 48.5% of the essays in this study had the major communicative purpose of 'to explain', while 47.3% of the essays had the major communicative purpose of 'to argue'. The student registers also differed from one another in the extent to which they exhibited such variation. For example, essays exhibited extensive variation among texts, while methodology recounts and explanations were more homogeneous.

The second of these recent studies, Biber et al. (2021a), had a different type of general research goal: to explore communicative variation within the single general register of face-to-face conversation. Egbert, Wizner, Keller, Biber, McEnery and Baker (2021) document the initial steps in that project, beginning with an operational definition of the textual units that comprise conversational interactions, referred to as Discourse Units (DUs):

2. These categories are referred to as 'genres' in Nesi and Gardner, but the term used does not change the fact that these are named categories recognized by students/teachers/researchers rather than bottom-up categories defined in situational or communicative terms.

1. Recognizably self-contained: A Discourse Unit has an identifiable beginning and end.
2. Coherent for its overarching communicative purpose, typically coupled with a single topic. Each DU has one communicative purpose (e.g., complaining about annoying co-workers).
3. Length requirement: A DU has a minimum of five utterances or 100 words.

Then, this operational definition was applied to segment 364 extended conversation transcripts (2,326,000 words) from the BNC-S 2014 corpus, resulting in a corpus of 17,828 coherent Discourse Units.

The next major step in this project was to develop a framework for analyzing the general communicative purposes of each DU (see Egbert et al. 2021). The framework was piloted and revised through multiple rounds of analysis, resulting finally in a coding rubric with nine major communicative purposes (e.g., Situation-dependent commentary, Joking around, Figuring things out, Sharing feelings and opinions, Describing or explaining the past). That framework was applied to code all 17,828 DUs in the conversation corpus, with two major methodological innovations:

a. Each DU was coded for the possibility of multiple communicative purposes
b. Each communicative purpose was coded on an ordinal scale, reflecting the extent to which the purpose was required for the communicative goals of the DU

Based on this coding, Biber et al. (2021a) describe the communicative characteristics of conversational DUs in a 9-dimensional space. The conversational DUs in the corpus occupied the full space of variation, with all 36 of the possible 2-way combinations of communicative parameters being attested. At the same time, certain combinations of communicative parameters were especially prevalent, and analysis of those characteristics was used to group DUs into 16 major conversational discourse types. As Figure 3 shows, some of these communicative combinations were much more commonly found than others. For example, over 25% of all DUs combine the major communicative purposes of time-neutral description and expressing personal feelings. In contrast, less than 3% of the DUs had a major purpose of giving advice. A much fuller description of the characteristics of these conversational discourse types is provided in Biber et al. (2021a). The main points for our purposes here, however, are the existence of extensive communicative variation within a general register like conversation, and an illustration of how such variation can be analyzed, resulting in systematic patterns of use that can be functionally interpreted.

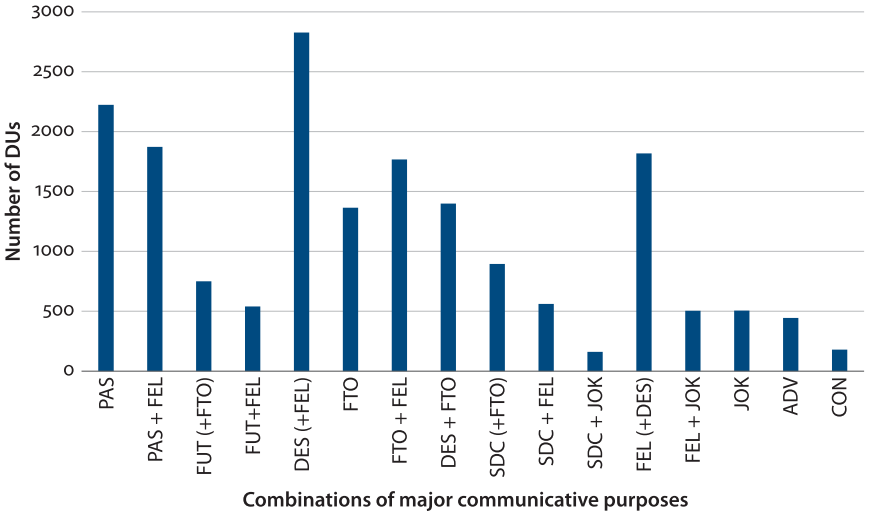

Figure 3. Combinations of major communicative purposes found in conversational discourse units

Note. PAS = Past time description; FUT = Future time description; DES = Time-neutral description; FEL = Feelings and opinions; FTO = Figuring things out; SDC = Situation dependent commentary; JOK = Joking around; ADV = Advice; CON = Conflict

Finally, the third recent study to focus on situational variation among texts within registers dealt with the discourse domain of the searchable web. This study grew out of a larger project to describe the patterns of register variation on the searchable web (see Biber, Egbert & Davies 2015; Egbert, Biber & Davies 2015; Biber & Egbert 2018). That study was based on a near-random sample of c. 50,000 documents from across the spectrum of the searchable web, referred to as CORE. Although the original goal of the study was to compare the linguistic characteristics of web registers, it quickly became apparent that many web documents have few or no external indicators of register category, making it difficult to classify the texts into register categories. These issues, and the resulting descriptions of web registers, are explored in detail in Biber and Egbert (2018). However, the important finding for our purposes here was the fact that coders were unable to fully agree on the register category of almost half of the documents in CORE (see Tables 3.3–3.7 in Biber & Egbert 2018).

This finding was the primary motivation for further exploration of web documents in a continuous-situational analytical framework, described in Biber et al. (2020). That follow-up study began with a set of 902 texts randomly selected from CORE. Those texts had been coded in the original study for their register categories (e.g., encyclopedia article, news article, opinion blog), but they were

recoded in the follow-up study for 23 communicative parameters (e.g., the text was pre-planned and edited; the author was an expert; the purpose was to narrate past events). This coding scheme was developed to represent the range of situational characteristics identified in previous frameworks, such as Biber and Conrad (2019, Chapter 2). However, that earlier framework was modified to achieve two important methodological innovations:

1. each situational parameter was coded as a quantitative (ordinal) variable on a 1–6 scale, to permit description of the extent to which a text realizes a given characteristic;
2. each value of some situational parameters was represented as a separate variable. This innovation is most important for the coding of communicative purpose. That is, rather than choosing a single communicative purpose for a text (e.g., narrative vs. explanatory vs. descriptive vs. persuasive), the coder was asked to evaluate the extent to which the text accomplishes each (and potentially all) of those different purposes.

These 23 situational/communicative variables were then analyzed using the methods of multi-dimensional analysis, resulting in two major situational dimensions: 'Personal opinionated discourse versus technical information supported with evidence' and 'Narrative, entertaining discourse versus other communicative purposes (explanatory, advice, or procedural discourse)'. Finally, following the methods used in other MD studies, situational dimension scores were computed for each text, allowing a comparison of registers.

For our purposes here, two major findings from this analysis are especially important:

1. all registers exhibited some degree of internal variation among texts; but
2. the registers differed from one another in the extent to which they exhibited such internal variation

These patterns are illustrated in Figures 4 and 5 below. For example, registers like encyclopedia articles and song lyrics had little internal variation among texts (see Figure 4), while registers like interactive discussions and news reports exhibited extensive variation in the communicative characteristics of texts (see Figure 5).

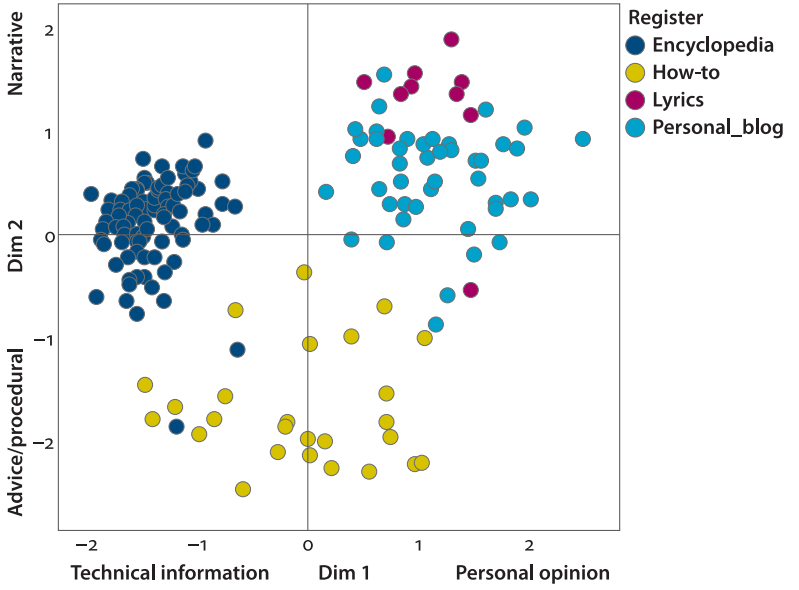

Figure 4. Registers that are relatively well-defined with respect to the situational dimensions (from Biber et al. 2020)

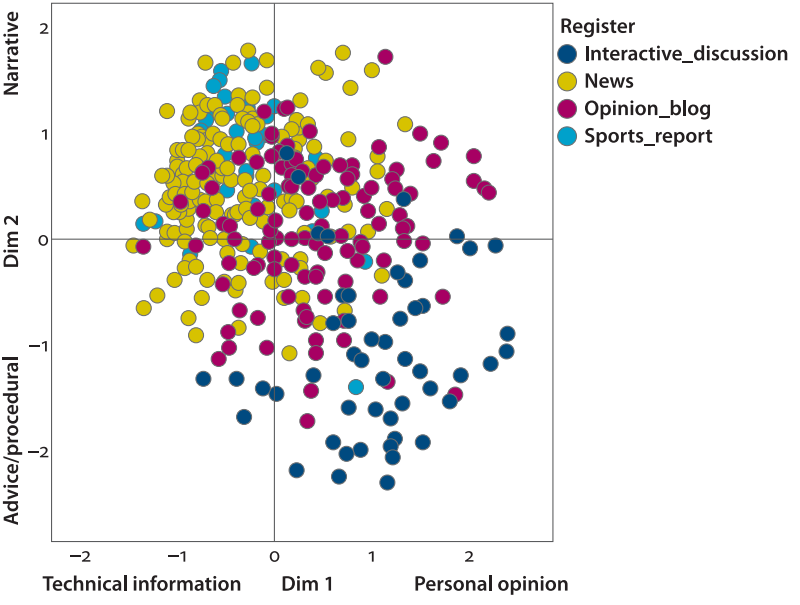

Figure 5. Registers that are not well-defined with respect to the situational dimensions (from Biber et al. 2020)

In summary, the present section has briefly surveyed the findings from previous empirical studies to document three major generalizations about language use in relation to registers:

1. Registers differ with respect to their typical characteristics, including both their linguistic characteristics and their situational characteristics.
2. At the same time, there are important differences among texts within registers in their typical linguistic characteristics and typical situational/communicative characteristics, and those differences can be interpreted in terms of functional correspondence.
3. It is possible to measure communicative characteristics (as well as linguistic characteristics) with continuous/quantitative variables, and to compare texts/registers with respect to those measures.

It is these three empirically-documented patterns of use that lead to the theoretical conceptualization of 'register' proposed in the following section.

4. So, what is a 'register'?

The problem that we hope to address in this editorial is the logical mismatch between theoretical conceptualizations of 'register' as text varieties that are defined by situational characteristics, versus the empirical research surveyed in the preceding section, showing the existence of extensive linguistic and situational/communicative variation among texts within registers. That is, logically, there should not be variation within a category with respect to the characteristics that define the category.

We believe that the root of this problem is the attempt to conceptualize 'register' as a scientific construct that can be defined in terms of necessary characteristics, and we argue instead that 'register' should be conceptualized as a cultural construct. Languages and cultures are organized in terms of cultural taxonomies. Such taxonomies consist of categories that are recognized by speakers of a language, even though it is not possible to provide a precise definition of the necessary characteristics for those categories. (In fact, individual speakers are likely to disagree on even the typical characteristics of specific categories.) For example, speakers of English recognize the categories of 'cups', 'mugs', and 'glasses' – but it is not possible to specify the necessary defining characteristics of those categories. It is certainly possible to describe the typical characteristics of 'cups', 'mugs', and 'glasses'. But the categories do not have a scientific basis. That is, these categories exist in the language/culture, even if we cannot specify their necessary defining characteristics.

This perspective is clearer if we contrast cultural taxonomies with scientific taxonomies. In a scientific taxonomy, the categories are derived inductively, based on empirical evidence, with the result that each category is defined in terms of its necessary characteristics. For example, a *fir tree* is a sub-category of *tree* with the following defining characteristics: an evergreen coniferous with flattened needle-like leaves that are attached singly to the branches, and cones that stand upright on the branches and disintegrate at maturity. There can be considerable variation among individual fir trees with respect to other characteristics (e.g., size, color of the bark), but not with respect to the characteristics that define the category.[3]

The important point for our purposes is that there is no variation among members within a scientific category with respect to the defining characteristics. Instead, if we encounter an item that has different characteristics, we would have discovered a new category. In contrast, cultural categories have no such requirements, because they are not based on a scientific specification of the necessary defining characteristics. And therefore, cultural categories can – and usually do – have considerable variation among individual members.

Our central argument here is that 'registers' are cultural categories, not scientific categories. These categories can be described for their typical situational and linguistic characteristics. But they are not defined in those terms. In fact, registers do not have definitions in terms of their necessary characteristics. Rather, cultures and languages evolve naturally in terms of such categorical organizations, without any scientific basis.

This distinction matters for two main reasons:

1. If registers are cultural categories, then we would predict the existence of variation among texts within the categories – with the implication that any comprehensive description of a register will include analysis and interpretation of the extent and nature of such variation.
2. If registers are cultural categories, then it is logically possible that some texts do not belong to any register.

In the following section, we briefly explore this second implication.

3. This portrayal of scientific taxonomies is an over-simplification, especially given recent research that documents variation within species. But the point here is the contrast with cultural taxonomies, which have no scientific basis at all.

5. Texts without a register

In a scientific taxonomy, every observation belongs to some category, because the categories are determined inductively. That is, if we encounter an observation that fails to conform to the characteristics of existing categories, then we would have discovered a new category, which would be defined by the characteristics of this new observation. As noted above, there are two consequences of this approach: (1) no variation within categories with respect to the defining characteristics, and (2) every observation belongs to some category.[4]

In contrast, a cultural taxonomy would not entail either of these consequences. Cultural categories exist simply because the culture recognizes the distinctions. But cultures do not specify necessary defining characteristics for those categories, and they do not specify any necessary requirement that all observations must belong to some category.

In Section 3, we summarized empirical evidence showing the existence of extensive variation within register categories. It turns out that empirical corpus-based studies have also uncovered the existence of texts that do not belong to any register. However, there are comparatively few studies of this type, because most corpora have been designed in a way that entirely ignores the existence of texts that do not belong to a register. That is, corpora are usually designed to represent one or more target register(s), and so the corpus includes only texts from those pre-selected registers. Thus, texts that fall outside of register boundaries are not documented.

An alternative approach would be to collect all texts from a discourse domain, and to then try to classify each text into a register – opening up the possibility that we might encounter texts that could not be classified: i.e., texts that do not belong to any register. The study of register variation on the searchable web introduced in Section 3 above (see Biber et al. 2015; Egbert et al. 2015) enabled exactly this possibility, because it was based on a near-random sample of documents from across the spectrum of the searchable web (see also Biber & Egbert 2018). Four raters coded each of the texts in the corpus to identify both the general register and more specific sub-register, based on the assumption that all texts would fall into a single register category. We quickly learned, however, that this assumption was problematic. Raters were able to agree on the general register of 70% of the texts in the corpus, and on the specific sub-register for c. 52% of the texts. But, at the same time, there were a large number of texts that did not clearly belong to any register

4. We noted in Note FN #3 above that this is a simplification of current scientific research. In particular, scientific taxonomies are currently considered to be much more complex than in earlier theory, accounting for mutations and other sources of variability within categories.

category (see Biber & Egbert 2018: 200–208). For example, 17.5% of the texts were coded as belonging to three different general register categories (i.e., with a 2–1–1 split among the raters), and 11.3% of the texts were coded as belonging to four different specific sub-registers (i.e., with a 1–1–1–1 split among the raters).

In short, many of the texts in the corpus were hybrid in nature, containing contextual characteristics of more than one category. As such, these could be regarded as texts that did not belong to any culturally-recognized register. Text Sample 1 below is an example. Raters were able to identify the communicative purposes of such texts, but unable to classify them into any culturally-recognized category.

Text Sample 1. An example web document with no rater agreement on register category (coded as Description-of-a-thing; How-to; Informational-persuasion; Description-with-intent-to-sell)

You do not need to have a Paypal account to use this service. Just click on the pay now button, look for where it says: "Don't have a PayPal account? Use your credit card or bank account (where available)" then click on the "continue" link next to that.

Eric Liddell Centre will not pass on your details to non associated companies. We comply with the Data Protection Act 1998. Information on the use of personal data by Eric Liddell Centre is available from: 15 Morningside Road, Edinburgh, EH10 4DP.

Rights and Responsibilities of Members

People admitted to membership will be expected to uphold the objects of the association, which are to further the welfare of the local community as an expression of Christian witness.

Paid up members have the right to:
- Attend and vote at the Annual General Meeting.
- Call an Extraordinary General Meeting (this requires a minimum of twenty members).

Every member agrees to pay a sum not exceeding œ1 in the event of the association being wound up.

A full description of the rights of members is contained in our constitution (memorandum and articles of association) which is available to download here.

As noted above, it is hard to find examples of texts-with-no-register in current corpora, because of the normal methods used to construct corpora. As a result, it is hard to assess the extent to which such texts exist in natural discourse. However, the Biber and Egbert (2018) study of the searchable web clearly documents the existence of such texts, and further suggests that they are more prevalent than we might believe. Future research is required to investigate these possibilities empirically.

6. Wrapping up

So, where does this leave us? We believe that the empirical studies surveyed here provide strong evidence supporting the validity of register studies. In terms of our theoretical framework, registers exist. They are not defined in terms of their linguistic characteristics; they are not defined in terms of their contextual characteristics; but they are culturally recognized, and as such, it is important to describe them. Two patterns are clearly established from previous research: (1) that registers have typical contextual and linguistic characteristics, making it possible and productive to analyze the patterns of variation across registers; and (2) that there is extensive contextual and linguistic variation among texts within registers, making it essential to include description of such variation in all register studies. In sum, we would argue that studies of registers have been and will continue to be meaningful research enterprises – with the caveat that all future studies of this type should include acknowledgement of the ways in which texts vary within the targeted registers.

Our goal in this editorial has not been to criticize or detract from the existing framework for exploring and describing variation between registers. Instead, we have proposed a complementary framework founded on the principle that registers are culturally-recognized categories that can be described, not scientific categories that can be defined. This opens up the possibility of situational/contextual and linguistic variation across texts within registers, as well as the possibility of texts without a register category. We believe that accounting for and analyzing these two phenomena will greatly enhance the quality and comprehensiveness of register studies.

We see future research progressing in three major directions:

1. cultural / sociological / historical / psycholinguistic research to further explore the nature of registers;
2. empirical studies on the nature of functional correspondence between situational/contextual and linguistic variation; and
3. empirical studies that account for other aspects of textual variation

The first of these might seem surprising, given our discussion in the preceding sections. But we freely acknowledge that our proposed framework fails to fully account for what a register is, how registers evolve historically, or why a culture recognizes one set of register distinctions as opposed to other possible distinctions. These are questions for psycholinguists, historical linguists, and socio/anthropological linguists. We believe that it is clear that registers exist, and we have argued that this is the only fact that we need in order to motivate studies of register variation. However, that fact should not be interpreted to mean that we

do not also need future research into the question of what registers actually are and how they function in societies and cultures.

Second, future research is required to much more fully investigate the nature of functional correspondence. For example, we know little about texts that do not belong to a culturally-recognized register. How prevalent are such texts? And how do we go about answering that question? It is relatively easy to collect texts from recognized registers, and we currently have solid methods for evaluating the representativeness of corpus samples for such categories (see Egbert et al. 2022). However, the methods for both the collection and evaluation of samples of texts-with-no-register is uncharted territory that needs to be developed in future research. And as a result of these methodological challenges, we know little about the nature of functional correspondence across different discourse domains. In particular, future research is required to investigate differences in the extent of functional correspondence observed within and across registers of different types, with the goal of explaining such differences with respect to different types of contextual variables and different types of linguistic variables.

Finally, future research is required to explore the ways in which 'register' variation and functional correspondence variation interact within the broader universe of textual variation. For example, it is clearly the case that not all linguistic variation can be accounted for by reference to either registers or functional correspondence. Some contextual variables are stronger predictors of linguistic variation than other contextual variables. And particular sets of contextual variables have functional relations with particular sets of linguistic variables. But beyond those specific relations, it might also be the case that some contextual variation does not predict linguistic variation, or that some linguistic variation cannot be accounted for in terms of contextual variation. For example, factors like individual style, dialect differences, or local linguistic context might all play a role in the goal of accounting for patterns of textual variation. Thus, the ultimate goal of this line of research is to consider the full range of possible factors, resulting in a comprehensive theory of textual variation.

Acknowledgements

We benefited from detailed comments on earlier drafts of this editorial from Susan Conrad, Bethany Gray, Larissa Goulart, Marianna Gracheva, Daniel Keller, Shelley Staples, Stacey Wizner, and Margaret Wood. In addition, we would like to thank Tove Larsson and Randi Reppen for their input during long discussions on these issues. Although any remaining unclarity is entirely our own fault, we believe that the criticisms and suggestions of these colleagues have greatly helped us to express our intended perspectives (and avoid confusing tangents).

References

Barbieri, F., & Wizner, S. (2019). Appendix A: Annotations of major register and genre studies. In D. Biber & S. Conrad (Eds.), *Register, genre, and style* [2nd edition] (pp. 318–349). Cambridge: Cambridge University Press.

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

Biber, D. (1994). An analytical framework for register studies. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 31–56). New York: Oxford University Press.

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, *8*(1), 9–37.

Biber, D. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, *14*(1), 7–34.

Biber, D. (2019). Text-linguistic approaches to register variation. *Register Studies*, *1*, 42–75.

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd edition). Cambridge: Cambridge University Press.

Biber, D., Egbert, J., & Davies, M. (2015). Exploring the Composition of the Searchable Web: A Corpus-based Taxonomy of Web Registers. *Corpora* 10(1), 11–45.

Biber, D., & Egbert, J. (2018). *Register variation online*. Cambridge: Cambridge University Press.

Biber, D., Egbert, J., & Keller, D. (2020). Reconceptualizing register in a continuous situational space. *Corpus Linguistics and Linguistic Theory*, *16*(3), 581–616.

Biber, D., Egbert, J., Keller, D., & Wizner, S. (2021a). Towards a taxonomy of conversational discourse types: An empirical corpus-based analysis. *Journal of Pragmatics*, *171*, 20–35.

Biber, D., J. Egbert, D. Keller, & S. Wizner. (2021b). Extending text-linguistic studies of register variation to a continuous situational space: Case studies from the web and natural conversation. In E. Seoane & D. Biber (Eds.), *Corpus-based approaches to register variation* (pp. 19–49). Amsterdam: John Benjamins.

Egbert, J., & Biber, D. (2016). Do all roads lead to Rome? Modeling register variation with factor analysis and discriminant analysis. *Corpus Linguistics and Linguistic Theory*, *14*(2), 233–273.

Egbert, J., Biber, D., & Davies, M. (2015). Developing a bottom-up, user-based method of web register classification. *Journal of the Association for Information Science and Technology*, *66*(9): 1817–1831.

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Cambridge: Cambridge University Press.

Egbert, J., Wizner, S., Keller, D., Biber, D., McEnery, T., & Baker, P. (2021). Identifying and describing functional discourse units in the BNC Spoken 2014. *Text & Talk*, *41*(5–6), 715–737.

Egbert, J., & Schnur, E. (2018). The role of the text in corpus and discourse analysis: missing the trees for the forest. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse* (pp. 159–173). New York: Routledge.

Ferguson, C. (1981). "Foreign talk" as the name of a simplified register. *International Journal of the Sociology of Language*, *28*, 9–18.

Gray, B., & Egbert, J. (2019). Register and register variation. *Register Studies*, *1*(1), 1–9.

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam: John Benjamins.

Goulart, L., Gray, B., Staples, S., Black, A., Shelton, A., Biber, D., Egbert, J., & Wizner, S. (2020). Linguistic perspectives on register. *Annual Review of Linguistics*, *6*, 435–455.

Goulart, L., Biber, D., & Reppen, R. (2022). In this essay, I will…: Examining variation of communicative purpose in university writing. *Journal of English for Academic Purposes*, *59*.

Halliday, M.A.K. (1968). The users and uses of language. In J.F. Fishman (Ed.), *Readings in the sociology of language* (pp. 139–169). The Hague: Mouton.

Halliday, M.A.K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of written English: Situational factors and linguistic features* (pp. 162–178). London: Frances Pinter.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M.A.K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Victoria: Deakin University.

Martin, J.R. (2002). Meaning beyond the clause: SFL perspectives. *Annual Review of Applied Linguistics*, *22*, 52–74.

Matthiessen, C.M. (2019). Register in systemic functional linguistics. *Register Studies*, *1*(1), 10–41.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.

Reid, T.B. (1956). Linguistics, structuralism and philology. *Archivum Linguisticum*, 8, 28–37.

Address for correspondence

Douglas Biber
Northern Arizona University
Flagstaff, AZ 86011-6032
USA

Douglas.Biber@nau.edu

Co-author information

Jesse Egbert
Northern Arizona University

Jesse.Egbert@nau.edu