

The Authors Respond: Strengths and Goals of Multidimensional Analysis

DOUGLAS BIBER

*Northern Arizona University
Flagstaff, Arizona, United States*

SUSAN CONRAD

*Portland State University
Portland, Oregon, United States*

RANDI REPPEN

*Northern Arizona University
Flagstaff, Arizona, United States*

PAT BYRD

*Georgia State University
Atlanta, Georgia, United States*

MARIE HELT

*California State University
Sacramento, California, United States*

■ We welcome thoughtful discussion of the multidimensional (MD) research approach, including proposals to make the methodology even more effective than it is. However, Ghadessy's critique of MD analysis is flawed in four major ways. First, it does not acknowledge the strengths of MD analysis or address whether alternative approaches can retain those strengths; second, it does not reflect an understanding of the linguistic and statistical bases of MD analysis; third, it incorrectly assumes that MD analysis has claimed to be an all-inclusive approach to linguistic variation; and, finally, much of the critique is based on personal preference for linguistic analysis rather than external considerations such as the uses for the analysis.

The main point of Ghadessy's critique seems to be the desirability of including additional linguistic characteristics—especially discourse features—in future MD analyses. We fully agree with this general goal, and we have called for similar extensions in nearly every published MD study. For example, in the article under review, we note that “additional features—including rhetorical and lexical features—also deserve attention” (p. 43). However, we do not believe that the desirability of future extensions constitutes a “major deficiency” in MD methodology or diminishes the important contributions of past MD research.

Although Ghadessy raises several specific criticisms in his review, he does not address or acknowledge the strengths of MD analysis:

- It takes advantage of corpus-based research methodologies to investigate overall patterns of register variation and achieve more generalizable results than in other methodologies.
- It is based on the analysis of large corpora, in this case representing university language use, including long texts, multiple texts from each register, and a large range of spoken and written university registers.
- The use of computer programs and interactive software tools makes it possible to analyze a wide range of linguistic features in each text; in the present case, 67 different grammatical and syntactic features are covered.

In contrast, most register studies undertaken with other approaches have focused on a single register, a small number of linguistic features, and a smaller number of texts. Because of these differences, we claimed that our study “reports results of the most comprehensive linguistic analysis of academic language to date” (p. 11). We stand by this claim, and we would argue that the strength of alternative approaches should be evaluated by these criteria as well.

Ghadessy seems most bothered by our claim to be “comprehensive” (pp. 13, 43). He seems to assume an absolute meaning for the term, something like *all-inclusive*. As a result, he concludes that failure to include discourse features in MD analyses is a “major deficiency,” a “problem,” and a “flaw,” and that such analyses cannot claim to be comprehensive.

However, we use the term *comprehensive* as a gradable adjective, with the meaning “of large scope; covering much; inclusive” (*Random House Webster’s College Dictionary*, 2nd ed., s.v. “comprehensive,” first definition). We believe that the analysis is comprehensive by that definition: The study is based on analysis of a 2.7-million-word corpus, containing 423 texts from 10 different spoken and written registers. Further, we analyzed the distribution of 67 different linguistic features in each text, including grammatical classes (e.g., nouns, verbs, adjectives, pronouns), meaningful grammatical distinctions (e.g., past tense, perfect aspect, passive voice), and syntactic/clause-level features (e.g., relative clauses, adverbial clauses, complement clause types). MD studies are some of the most inclusive analyses of register variation yet attempted, and the omission of discourse features does not diminish that accomplishment.

Several other criticisms raised by Ghadessy seem simply to reflect his preference for a different analytical framework. Ghadessy points out that we do not adopt the framework of systemic functional linguistics (SFL) and do not analyze registers with respect to “field,” “tenor,” and “mode.” In addition, we analyze texts as “products” rather than focusing on meaning-making “processes.” We agree with these statements, but we do

not regard them as criticisms. They are simply reflections of the research constructs and goals of MD analysis and the ways they happen to differ from SFL. We see SFL as a complementary theoretical framework, with different analytical techniques and research goals from our own. We would welcome a synthesis of perspectives and methods, but we do not think that either approach should abandon its own goals and priorities in that effort.

Finally, Ghadessy makes several incorrect statements that reflect misunderstandings or misrepresentations. The first has to do with the definition and identification of registers. Ghadessy characterizes MD analysis as “Biber et al.’s MD theory for register identification.” However, in the MD approach, registers are named varieties in a culture, defined in situational terms, like conversation, letters, textbooks, and lectures. As we explain (p. 10, Footnote 1), registers in this view are not necessarily well defined linguistically; that is, there can be important linguistic differences among texts within a register. Similarly, registers are not necessarily distinguished from one another in their linguistic characteristics. For example, university textbooks and newspaper prose are completely different registers, but they are similar in many of their linguistic characteristics. MD analysis was developed to analyze those linguistic characteristics, addressing the extent to which any two registers are similar or different along multiple linguistic dimensions of variation. Ghadessy is not correct in stating that these dimensions are used to define or identify registers; rather, the dimensions are used to compare the linguistic characteristics of predefined registers. A complementary analytical approach—analyzing text types—identifies text categories that are well defined in linguistic terms (see the references cited on p. 10, Footnote 1, in our article).

Second, Ghadessy repeatedly states that MD analysis has an impoverished linguistic basis. He states that it is based on “the frequency of some grammatical categories, such as nouns, verbs, adjectives, pronouns, and articles”; that it is restricted almost exclusively to “single words or phrases”; and that it consists of “a list of words and their collocation probabilities.” On the flip side, Ghadessy suggests that MD analysis completely disregards clause-level features. Such descriptions are erroneous at best. In fact, MD analysis includes as wide a range of relevant grammatical classes and distinctions as possible within its corpus-based approach, including grammatical classes, analysis of grammatical/syntactic function, and analysis of many clause types and variants. These features include word classes like nouns, verbs, adjectives, and pronouns. Within most classes, MD analyses further include grammatical/syntactic distinctions. For example, verb phrases are analyzed for their grammatical function, distinguishing past and present tense; simple, perfect, and progressive aspect; and active and passive voice. Adjectives are subclassified

according to their syntactic function as attributive or predicative. More importantly, in contradiction to Ghadessy's claims, MD analysis includes many clause-level features, both main clause features (e.g., questions, clausal coordination) and dependent clause features (e.g., adverbial clauses, relative clauses, complement clauses). Further, dependent clauses are analyzed for specific syntactic type and variants (e.g., *that* relative clause with subject gap, *wh-* relative clause with object gap, passive participial postnominal [reduced relative] clause; *that* complement clause controlled by verb; *wh-* complement clause; causative vs. concessive vs. conditional adverbial clauses). These features are clearly listed in Table 1 (pp. 15–16). We find it difficult to imagine how Ghadessy could have read our article and come away with the impression that MD analysis consisted of “a list of words and their collocation probabilities.”

The MD analytical framework is not static, and we therefore welcome the addition of new features. For example, we have recently developed a new MD model of register variation among university spoken and written registers (see Biber, in press; Biber et al., in press). This analysis includes many additional linguistic features, like lexical bundle features (e.g., preposition initial lexical bundles, such as *in the form of*; *wh-*initial lexical bundles, such as *what you're saying is*; see Biber, Johansson, Leech, Conrad, & Finegan, 1999, chapter 13) and many lexicogrammatical features (e.g., mental verbs controlling *that* complement clauses and verbs of desire controlling *to* complement clauses).

Many discourse features—including characteristics of theme and theme progression, as discussed by Ghadessy—have thus far been prohibitively time-consuming to identify and count reliably in a large corpus. This difficulty helps to explain why Ghadessy's (1999) study of thematic features is based on only 1,286 clauses representing about 15,000 words.¹ However, although we currently lack the analytical tools

¹ Ghadessy cites his 1999 study as evidence that inclusion of thematic features will disconfirm the results of previous MD analyses. In that study, Ghadessy claims to have used a cluster analysis for this purpose. However, an examination of that study suggests that Ghadessy fails to understand the statistical basis of MD analysis.

In the MD framework, text types are identified through a multivariate statistical procedure called *cluster analysis*; this procedure groups observations that are maximally similar with respect to certain quantitative variables. In MD studies, the observations are texts, and the quantitative variables are the dimension scores (see Biber, 1995, chapter 9).

In contrast, Ghadessy (1999) claims to have performed a cluster analysis in a study in which the observations were individual clauses and the variables were nominal characteristics (e.g., whether the theme type of the clause was simple theme, multiple theme, or textual theme). A statistical cluster analysis is not possible in this case because there are no quantitative variables that could be used to group the observations. Instead, Ghadessy appears to have relied on a visual inspection of a cross-tabulation table, which showed the proportional use of each thematic clause type in each register. From a statistical point of view, the analysis carried out by Ghadessy is in no way comparable to a cluster analysis, and the results are in no way comparable to those achieved through MD analysis.

that permit generalizable, empirical corpus investigations of their distribution, we agree that thematic features are likely to be important register characteristics, and we welcome efforts to develop analytical techniques.

Until a utopian time when a single methodological approach can cover all aspects of language use, we are left with examining register variation through a variety of approaches. It is our position that all approaches conducted in principled ways should be welcome. We need to sit down to a banquet, if you will, rather than expecting any single dish to include all ingredients, and rather than criticizing a dish for not conforming to another chef's priorities.

REFERENCES

- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, D. (in press). Variation among university spoken and written registers: A new multi-dimensional analysis. In C. Meyer & P. Leistyna (Eds.), *Corpus analysis: Language structure and language use*. Amsterdam: Rodopi.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., et al. (in press). *Representing language use in the university: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus* (TOEFL Monograph). Princeton, NJ: Educational Testing Service.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. Harlow, England: Pearson Education.
- Ghadessy, M. (1999). Textual features and contextual factors for register identification. In M. Ghadessy (Ed.), *Text and context in functional linguistics* (pp. 125–139). Amsterdam: Benjamins.