Exploring Fictional Styles along Universal Dimensions of Register Variation

Douglas Biber and Jesse Egbert

Abstract Multi-dimensional (MD) analyses have been carried out to identify the linguistic parameters of register variation in many different discourse domains and many different languages (see, e.g., Biber 1988; 1995; 2014). Each MD study has identified linguistic dimensions that are peculiar to a particular language/discourse domain. However, the more theoretically interesting finding is that linguistically similar dimensions emerge in nearly all MD studies. Two of these dimensions are especially robust, making them strong candidates for universal dimensions of register variation: 1) a fundamental opposition between clausal/"oral" discourse versus phrasal/"literate" discourse, and 2) the opposition between "narrative" versus "non-narrative" discourse. It turns out that these same two functional parameters are fundamentally important in the discourse domain of fictional literature. The present paper overviews results of MD studies of English across discourse domains, and shows how these two universal dimensions are also fundamentally important in fictional literature.

Keywords Multi-dimensional analyses, register, literature

1. Introduction

One major focus of previous corpus-based research has been to describe the ways in which linguistic features vary across registers (e.g., conversation, classroom teaching, newspaper editorials). Such research can be carried out for many different research purposes, including a detailed description of a single register, comparing the patterns of register variation, or even describing patterns of variation across the styles of individual speakers or authors.

Biber, Douglas and Egbert, Jesse: Exploring Fictional Styles along Universal Dimensions of Register Variation, in: Hesselbach, Robert, et al. (Eds.): Digital Stylistics in Romance Studies and Beyond, Heidelberg: Heidelberg University Publishing, 2024, pp. 197–215. DOI: https://doi.org/10.17885/heiup.1157.c19372

Multi-Dimensional (MD) analysis is a methodological approach that has been applied to all of these general research goals. MD analyses have been conducted for many different discourse domains and many different languages. Using bottom-up statistical analyses, these studies have investigated specific patterns of register variation in several different discourse domains of English, as well as the more general patterns of register variation in numerous languages. Each study identifies linguistic dimensions that are peculiar to that particular language/domain. However, the more theoretically interesting finding is that linguistically similar dimensions emerge in nearly all of these studies. Two of these dimensions are especially robust, making them strong candidates for universal dimensions of register variation: 1) a fundamental opposition between clausal/"oral" discourse versus phrasal/"literate" discourse, and 2) the opposition between narrative versus non-narrative discourse. It turns out that these same two dimensions are important for distinguishing among the author styles employed in fictional novels. In the sections below, we first introduce the methodology of MD analysis, then briefly survey previous MD research studies with an emphasis on these potentially universal patterns of register variation, and finally present a new MD analysis of fictional novels, showing how these same two parameters of variation are important in that discourse domain.[1]

2. Overview of MD Analysis

The *Multi-Dimensional (MD)* analytical approach was originally developed to investigate the linguistic patterns of variation among spoken and written registers (see, e.g., Biber 1986; 1988; 1995). Studies in this research tradition have used large corpora of naturally-occurring texts to represent the range of spoken and written registers in a language. These registers are compared with respect to "dimensions" of variation (identified through a statistical factor analysis), comprising constellations of linguistic features that typically co-occur in texts. Each dimension is distinctive in three respects:

— It is defined by a distinct set of co-occurring linguistic features;
— It is associated with particular communicative functions;
— There are different patterns of register variation associated with each dimension.

The MD approach uses statistical factor analysis to reduce a large number of linguistic variables to a few basic parameters of linguistic variation: the "dimensions." In MD

1 Sections 2 and 3 are based on earlier summaries of the research goals, methods and findings of MD analysis, especially Biber (2014) and Biber (2019).

analyses, the distribution of individual linguistic features is analyzed in a corpus of texts. Factor analysis is then used to identify the systematic co-occurrence patterns among those linguistic features—the "dimensions"—and then texts and registers are compared along each dimension. Each dimension comprises a group of linguistic features that usually co-occur in texts (e.g., nouns, attributive adjectives, prepositional phrases). The dimensions are then interpreted to assess their underlying functional associations.

The first book-length MD analysis (Biber 1988) investigated the relations among general spoken and written registers in English, based on analysis of the LOB (Lancaster-Oslo-Bergen) Corpus (15 written registers) and the London–Lund Corpus (6 spoken registers). 67 different linguistic features were analyzed computationally in each text of the corpus. Then, the co-occurrence patterns among those linguistic features were analyzed using factor analysis, identifying the underlying parameters of variation: the factors or "dimensions."

After the statistical analysis is completed, dimensions are interpreted functionally, based on the assumption that linguistic co-occurrence reflects underlying communicative functions. That is, linguistic features occur together in texts because they serve related communicative functions. Table 1 summarizes the first two dimensions from the 1988 factor analysis, including a list of the most important linguistic features comprising each dimension as well as the interpretive functional labels.

Table 1 Summary of the major linguistic features co-occurring on Dimensions 1 and 2 from the 1988 MD analysis of register variation

	Dimension 1: Involved versus informational production	Dimension 2: Narrative versus non-narrative discourse
Positive features	mental (private) verbs, *that* complementizer deletion, contractions, present tense verbs, *WH*-questions, 1st and 2nd person pronouns, pronoun *it*, indefinite pronouns, *do* as pro-verb, demonstrative pronouns, emphatics, hedges, amplifiers, discourse particles, causative subordination, sentence relatives, *WH*-clauses	past tense verbs, 3rd person pronouns, perfect aspect verbs, communication verbs
Negative features	nouns, long words, prepositions, type/token ratio, attributive adjectives	present tense verbs, attributive adjectives

Each dimension can have positive and negative features. Rather than reflecting importance, positive and negative signs identify two groupings of features that occur in a complementary pattern as part of the same dimension. That is, when the positive features occur together frequently in a text, the negative features are markedly less frequent in that text, and vice versa.

For Dimension 1, the interpretation of the negative features is relatively straightforward. Nouns, word length, prepositional phrases, type/token ratio, and attributive adjectives all reflect an informational focus, a careful integration of information in a text, and precise lexical choice. The set of positive features for Dimension 1 is more complex, although all of these features have been associated with interpersonal interaction, a focus on personal stance, and real-time production circumstances. For example, first and second person pronouns, *WH*-questions, emphatics, amplifiers, and sentence relatives can all be interpreted as reflecting interpersonal interaction and the involved expression of personal stance (feelings and attitudes). Other positive features are associated with the constraints of real-time production, resulting in a reduced surface form, a generalized or uncertain presentation of information, and a generally "fragmented" production of text; these include *that*-deletions, contractions, pro-verb *do*, the pronominal forms, and final (stranded) prepositions.

Overall, Dimension 1 represents a parameter marking interactional, stance-focused, and generalized content (the positive features on Table 1) versus high informational density and precise word choice (the negative features). Two separate communicative considerations seem to be represented here: the primary purpose of the writer/speaker (involved versus informational), and the production circumstances (those restricted by real-time constraints versus those enabling careful editing possibilities). Reflecting both of these parameters, the interpretive label "Involved versus informational production" was proposed for the dimension underlying this factor.

A second major step in interpreting a dimension is to consider the similarities and differences among registers with respect to the set of co-occurring linguistic features. To achieve this, *dimension scores* are computed for each text, by summing the individual scores of the features that co-occur on a dimension (see Biber 1988: 93–97). Once a dimension score is computed for each text, the mean dimension score for each register can be compared across registers. For example, Figure 1 plots the mean dimension scores of registers along Dimension 1 from the 1988 MD analysis. The registers with large positive values (such as face-to-face conversations) have high frequencies of positive Dimension 1 features (e.g., present tense verbs, private verbs, etc.) combined with low frequencies of negative Dimension 1 features (e.g., nouns, prepositional phrases, etc.). Registers with large negative Dimension 1 values (e.g., academic prose, official documents) have the opposite linguistic characteristics.

The relations among registers shown in Figure 1 confirm the interpretation of Dimension 1 as distinguishing among texts along an oral/literate continuum. At the positive extreme, conversations are highly interactive and involved, with the language produced under real-time circumstances. And at the negative extreme, registers such as academic prose are non-interactive but highly informational in purpose, produced under circumstances that permit extensive revision and editing.

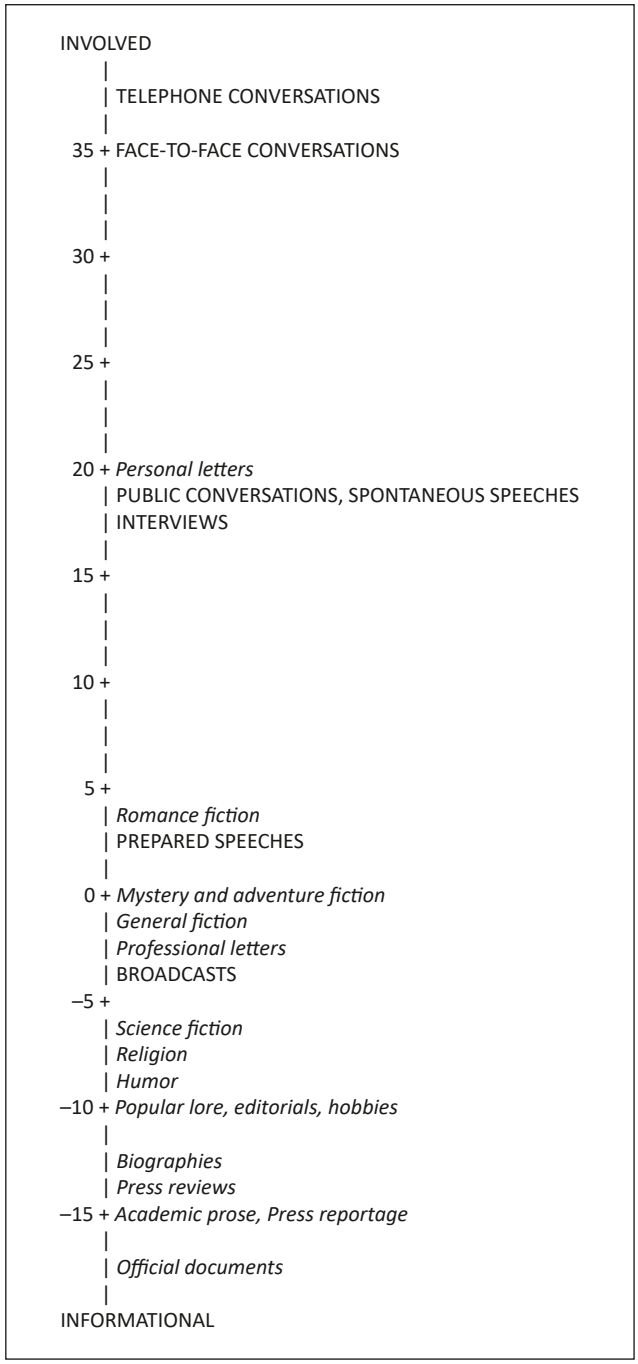

Fig. 1 Mean scores of registers along Dimension 1: Involved versus informational production. Written registers are in italics; spoken registers are in CAPS. ($F = 111.9$, $p < .0001$, $r^2 = 84.3$ percent), (adapted from Figure 7.1 in Biber 1988).

The overall comparison of spoken and written registers in the 1988 MD analysis requires consideration of all dimensions of variation, which each define a different set of relations among spoken and written registers. For example, Dimension 2 (see Table 1) is interpreted as "Narrative versus non-narrative concerns." The positive features—past tense verbs, third-person pronouns, perfect aspect verbs, communication verbs, and present participial clauses—are associated with past-time narration. In contrast, the negative features—present tense verbs and attributive adjectives—have non-narrative communicative functions. The distribution of registers along Dimension 2 supports the interpretation as narrative versus non-narrative concerns. All types of fiction have high positive scores on this dimension, reflecting their emphasis on narrating events. In contrast, registers which are typically more concerned with events currently in progress (e.g., broadcasts) or with building arguments rather than narrating (e.g., academic prose) have negative scores on this dimension.

The other dimensions in the analysis can be interpreted in a similar way. Overall, the 1988 MD analysis showed that English registers vary along several underlying dimensions associated with different functional considerations, including: interactiveness, involvement and personal stance, production circumstances, informational density, informational elaboration, narrative purposes, situated reference, persuasiveness or argumentation, and impersonal presentation of information.

Subsequent MD studies have shown that some of these dimensions turn out to be peculiar to English or to particular discourse domains (see Biber 1995; 2014). However, two linguistic parameters of variation have emerged consistently as dimensions across MD studies: a basic oral/literate parameter of variation, and a narrative/non-narrative dimension. The linguistic compositions and functional associations of these dimensions are remarkably stable across discourse domains (and languages, such as Spanish, Korean, Somali, and Czech; see Biber 2014 for an earlier survey), although the particular patterns of register variation differ according to the specific language/culture/discourse domain. In the following section, we briefly document the characteristics of these dimensions in previous MD studies of English, and then turn to an MD analysis of fictional styles in Section 4.

3. Universal Dimensions in Previous MD Studies of English

Numerous studies have undertaken MD analyses of particular discourse domains in English (as well as studies of different languages; see Biber 2014). Given that each of these studies is based on a different corpus of texts (representing a different discourse domain or different language), and based on increasingly comprehensive sets of linguistic features (as computational techniques for linguistic analysis have improved),

Table 2 The oral/literate dimension in selected MD studies of English

Discourse domain	Linguistic features defining Dimension 1	Register pattern along Dimension 1
University spoken and written registers; Biber (2006)	contractions, pronouns, present tense verbs, progressive aspect, time / place / stance adverbials, *that*-clauses, *WH*-clauses, adverbial clauses	service encounters, office hours, study groups, classroom teaching
	VERSUS	VERSUS
	nouns, nominalizations, attributive adjectives, prepositional phrases	textbooks, course packs, institutional writing
Conversational text types; Biber (2008a)	contractions, 1st and 2nd person pronouns, activity verbs	casual conversations
	VERSUS	VERSUS
	long words, abstract nouns, nominalizations, attributive adjectives, prepositional phrases	work-place conversations
Academic research articles across disciplines; Gray (2013)	pronouns, causative verbs, modals, stance and time adverbials, conditional adverbial clauses, *that* complement clauses, *wh*-clauses, *to*-clauses	theoretical philosophy
	VERSUS	VERSUS
	nouns, past tense verbs, prepositions, type-token ratio, word length; passives	quantitative biology, quantitative physics

it is reasonable to expect that they would each identify a unique set of dimensions. However, despite these differences in design and research focus, there are striking similarities in the dimensions that are uncovered across studies.

Most importantly, in nearly all previous MD studies, there is a dimension associated with an oral/literate opposition (cf. Biber 2014). Linguistically, this opposition is realized as two fundamentally different ways of constructing discourse: clausal versus phrasal. That is, across studies the "oral" pole of this dimension consists of verb classes (e.g., mental verbs, communication verbs), grammatical characteristics of verb phrases (e.g., present tense, progressive aspect), and modifiers of verbs and clauses (e.g., adverbs and stance adverbials). Interestingly, these "oral" features also include dependent clauses that function as clausal constituents, including adverbial clauses and finite complement clauses. In contrast, the "literate" pole usually consists of phrasal devices that function as elements of noun phrases, especially nouns, nominalizations, attributive adjectives, and prepositional phrases. Functionally, this dimension is interpreted as distinguishing between a personal/involved focus (personal stance, interactivity, and/or real-time production features) versus informational focus. And in nearly every case, this parameter is the first dimension identified by the statistical factor analysis (i.e., it is the most important factor, accounting for the greatest amount of shared variance).

Table 2 summarizes the composition of this oral/literate dimension in selected studies of English discourse domains. It is perhaps not surprising that Dimension 1

Table 3 Narrative dimensions in MD studies of particular discourse domains in English

Discourse domain	Linguistic features defining the narrative dimension	Register pattern along the dimension
University spoken and written registers; Biber (2006)	3rd person pronouns, human nouns, communication and mental verbs, past tense VERSUS concrete and quantity nouns	office hours, study groups VERSUS textbooks, course packs, institutional writing
Conversational text types; Biber (2008a)	past tense, 3rd person pronouns, communication verb + *that*-clause VERSUS present tense	narrative conversations VERSUS other conversations
Academic research articles across disciplines; Gray (2013)	past tense verbs, perfect aspect, communication verbs, 3rd person pronouns, time adjectives, etc. VERSUS technical nouns, passive voice verbs	history / political science / applied linguistics VERSUS theoretical / quantitative physics

in the original 1988 MD analysis was strongly associated with the oral/literate opposition, given that the corpus in that study ranged from spoken conversational texts to written expository texts. For the same reason, it is somewhat predictable that a similar dimension would have emerged from the study of spoken and written registers in world English varieties (Xiao 2009), and in the study of eighteenth-century general written and speech-based registers (Biber 2001).

However, it is more surprising that restricted comparisons of spoken and written registers would uncover a first dimension with a similar set of co-occurring linguistic features, associated with a similar opposition between oral and informational-literate registers, such as the studies of university spoken and written registers (Biber 2006), elementary school registers (Reppen 2001), and English as a second language (ESL) spoken and written exam responses (Biber, Gray, and Staples 2016).

The most surprising finding here is the existence of a similar first dimension in MD studies of registers from a single mode. Those include studies focused exclusively on spoken registers (e.g., call center interactions and conversations, Friginal 2009) as well as those focused exclusively on written registers (e.g., legal registers and research articles, Goźdź-Roszkowski 2011). In all of these cases, the linguistic composition of Dimension 1 is surprisingly similar, generally opposing verbs, dependent clauses, pronouns, and interpersonal features versus nouns and phrasal noun modifiers.

The second linguistic parameter that has emerged in all MD studies is a dimension associated with narration. Linguistically, this dimension is consistently defined by features like past tense verbs, 3rd person pronouns, human nouns, temporal adverbs, and communication verbs. In terms of register differences, this dimension distinguishes

narrative, time-organized descriptions of past-time events versus all other registers. Table 3 summarizes the narrative dimensions across a few MD studies of English.

4. Universal Dimensions in Fictional Novels

Two previous MD studies have focused specifically on the discourse domain of fictional novels: Biber's (2008b) study of novels in the nineteenth and twentieth centuries, and Egbert's (2012) study of nineteenth-century novels. The earlier study was based on 185 nineteenth- and twentieth-century novels collected from the Longman Corpus Network and from Project Gutenberg (comprising approximately 8.5 million words). Egbert's study then extended the earlier analysis by adding 100 fiction texts written by ten of the most famous nineteenth-century fiction authors (e.g., Louisa May Alcott, Charles Dickens, Henry James, Herman Melville, Mark Twain). Ten complete texts were collected for each author, making a total of approximately 10 million words.

Separate MD analyses were conducted in the two studies, with three factors extracted in the 2008 study and four factors extracted in the 2012 study. Table 4 shows that two of those dimensions are highly similar between the two analyses: Dimension 1, representing a basic oral/literate opposition, and Dimension 3, representing a 'narrative'/'non-narrative' opposition. Thus, the first dimension in both analyses shares many of the linguistic characteristics of the 'oral' versus 'literate' dimensions uncovered in other MD analyses, including verbs, adverbials, pronouns, and finite dependent clauses co-occurring as 'oral' linguistic features as opposed to nouns, attributive adjectives, and prepositional phrases co-occurring as 'literate' linguistic features. And the third dimension in both analyses shares many of the linguistic characteristics of the 'narrative' versus 'non-narrative' dimensions uncovered in other MD analyses, including past tense verbs and 3rd person pronouns as co-occurring 'narrative' features as opposed to present tense verbs.

At the same time, though, there are differences between the two analyses. For example, Dimension 1 in the Biber (2008b) study includes present tense, communication verbs, and 1st and 2nd person pronouns among the "oral" group of co-occurring features, while those features are grouped on to Dimension 3 in the Egbert (2012) study. These differences reflect the fact that 'oral' discourse in fictional novels (positive Dimension 1 characteristics) is also often dialogue and therefore 'non-narrative' discourse (negative Dimension 3). As a result, some interactive linguistic features tend to co-occur with present tense verbs marking non-narrative discourse. (In fact, 2nd person pronouns co-occur with present tense verbs on Dimension 3 in both MD analyses.)

Dimension 1 in the 2008 analysis is interpreted as 'Interactional (dialogue) versus informational (prose) focus.' This interpretation reflects the fact that the positive set

Table 4 The oral/literate dimensions and narrative-non-narrative dimensions in MD studies of fictional novels in English

Biber (2008b)	**Dimension 1: Interactional/involved versus informational focus**
	Features with positive loadings: – verbs: present tense – common verbs: mental, communication, pro-verb *do*, copula *be* – pronouns: 1st person, 2nd person – modals: possibility, necessity, prediction – adverbials: certainty – *that*-clauses: controlled by likelihood verbs, controlled by certainty verbs, controlled by communication – verbs, controlled by other mental/stance verbs – *that*-omission – *to*-clauses: controlled by desire verbs Features with negative loadings: – nouns: total nouns, place nouns, concrete nouns – prepositional phrases – adjectives: attributive – word length, type/token ratio – adverbials: place
	Dimension 3: Past versus present orientation (time and person)
	Features with positive loadings: – past tense verbs – perfect aspect verbs – 3rd person pronouns Features with negative loadings: – 2nd person pronouns – present tense verbs – contractions – nouns

of features are very similar to the set of interactive and involved co-occurring features found in face-to-face conversation. As a result, these features are common in novels that rely heavily on dialogue among characters. In contrast, the negative features are typical of informational written registers, and thus in novels, these features are characteristic of novels that rely heavily on descriptive or narrative prose. The functional interpretation of Dimension 1 in the 2012 study focuses on a slightly different opposition: thought presentation versus (informational) description.

Despite these differences, the similarities between the two analyses are strong in that both have strong dimensions associated with the 'oral'/'literate' opposition as well as the 'narrative'/'non-narrative' opposition. That is, Dimension 1 in both studies follows the pattern of the 'oral'/'literate' dimension in other studies, opposing a

Table 4 (*continued*)

Egbert (2012)	**Dimension 1: Thought presentation versus description**
	Features with positive loadings: – verbs: mental verbs, existence verbs, perfect aspect, possibility modals – pronouns: indefinite, *it* – adverbials: stance adverbials, general adverbs – dependent clauses: stance verb + *that*-clause; desire verb + *to*-clause; *WH*-clauses, *that*-clauses with complementizer deletion Features with negative loadings: – nouns – attributive adjectives – prepositional phrases
	Dimension 3: Narration versus dialogue [polarity reversed]
	Features with positive loadings: – past tense verbs – simple occurrence verbs – 3rd person pronouns Features with negative loadings: – present tense verbs – *have* as main verb – communication verbs – modal verbs – 1st and 2nd person pronouns – *WH*-questions

set of verbs, adverbs, and dependent clauses versus nouns, attributive adjectives, and prepositional phrases. And Dimension 3 in both studies follows the patterns of 'narrative'/'non-narrative' dimensions in previous studies, opposing past tense verbs and 3rd person pronouns versus present tense verbs.

It turns out that these dimensions are highly useful for distinguishing among the styles of fiction authors. Egbert (2012) focuses specifically on nineteenth-century authors. For example, with respect to Dimension 1, authors like James, Alcott, and Twain tend to rely on a clausal "oral" style, while authors like Melville and Kipling tend to rely on a more phrasal "literate" style. There are also important differences with respect to Dimension 3. For example, Hawthorne and Melville prefer a more narrative prose style, while Alcott and Twain rely much more heavily on present-time, interactive dialogue.

Biber (2008b) similarly identifies important differences among the prose styles of fiction authors in both the nineteenth and twentieth centuries. For example, with respect to Dimension 1, the children's novel *The House on Pooh Corner* by A. A Milne is extremely 'oral,' with frequent features reflecting personal involvement and interactivity. But these characteristics are not at all restricted to children's literature. For example, Forster's *A Room with a View* is nearly as marked for positive Dimension 1 features as *Pooh*.

1. Excerpt from *A Room with a View*
   E.M. Forster
   [present tense verbs, modals, and 1st and 2nd person pronouns are shown in italics]

   "*I want* so to see the Arno. The rooms the Signora promised *us* in her letter *would have looked* over the Arno. The Signora had no business to do it at all. Oh, it *is* a shame!"
   "Any nook *does* for *me*," Miss Bartlett continued; "but it *does seem* hard that *you shouldn't have* a view."
   Lucy felt that she had been selfish. "Charlotte, *you mustn't spoil me*: of course, *you must look* over the Arno, too. I meant that. The first vacant room in the front—"
   —"*You must have* it," said Miss Bartlett, part of whose travelling expenses were paid by Lucy's mother—a piece of generosity to which she made many a tactful allusion.
   "No, no. *You must have* it."
   "*I insist* on it. Your mother *would never forgive me*, Lucy."

At the other (informational) extreme of Dimension 1, we also find both children's literature (e.g., *The Tale of Peter Rabbit*, Beatrix Potter) and adult fiction (e.g., *Ulysses*, James Joyce). Surprisingly, the most "informational" novel in our corpus is a children's novel: Henry Williamson's *Tarka the Otter*, illustrated in Text Excerpt 2.

2. Excerpt from *Tarka the Otter*
   Henry Williamson
   [Nouns, attributive adjectives, and prepositional phrases are shown in italics]

   She ran *over the bullock's drinking-place* and passed *through willows to the meadow*, seeking *old dry grasses and mosses under the hawthorns growing by the mill-leat*, and gathering them *in her mouth with wool pulled from the over-arching blackberry brambles* whose *prickles* had caught *in the fleeces of sheep*. She re-

turned *to the river bank* and swam *with her webbed hind-feet to the oak tree*, climbed *to the barky lip of the holt*, and crawled within. *Two yards inside* she strewed her *burden on the wood-dust*, and departed *by water for the dry, sand-coloured reeds of the old summer's growth* which she bit off, frequently pausing to listen. *After several journeys* she sought *trout* by cruising *under water along the bank*, and *roach* which she found by stirring up *the sand and stones of the shallow* wherein they lurked.

Dimension 3 in the 2008 study also identifies important linguistic differences among novels. For example, children's novels like *Peter Rabbit* and *Tarka the Otter* (see excerpt 2) have a strong past narrative orientation (with large positive scores on Dimension 3). Adult novels like Virginia Woolf's *To the Lighthouse* also have a similar reliance on 3rd person past-time discourse, as illustrated in Text Sample 3:

3. Excerpt from *To the Lighthouse*
Virginia Woolf
[3rd person pronouns, past tense, and perfect aspect verbs shown in italics]

Nothing *happened*. Nothing! Nothing! as *she leant* her head against Mrs Ramsay's knee. And yet, *she knew* knowledge and wisdom *were stored* in Mrs Ramsay's heart. How then, *she had asked* herself, *did one* know one thing or another thing about people, sealed as *they were*? Only like a bee, drawn by some sweetness or sharpness in the air intangible to touch or taste, *one haunted* the dome-shaped hive, *ranged* the wastes of the air over the countries of the world alone, and then *haunted* the hives with their murmurs and their stirrings; the hives which *were* people. Mrs Ramsay *rose*. Lily *rose*. Mrs Ramsay *went*. For days there *hung* about *her*, as after a dream some subtle change is felt in the person one *has dreamt* of, more vividly than anything *she said*, the sound of murmuring and, as *she sat* in the wicker arm-chair in the drawing-room window *she wore*, to Lily's eyes, an august shape; the shape of a dome.

At the other extreme, many novels adopt a present-time focus, which tends to co-occur with 2nd person pronouns and contractions. It would be easy to suppose that this fictional style occurs in novels that include extensive dialogues among characters, given that actual face-to-face conversations also employ these same co-occurring linguistic features. However, it turns out that there are additional factors associated with this discourse style in novels.

In some novels, fictional dialogue does employ a present-time focus, similar to the norm in actual conversation. Thus, consider Text Sample 1 (above) and the following interaction from Joseph Heller's *Catch-22*:

4. Dialogue from *Catch-22*
Joseph Heller
[present tense verbs shown in italics]

"*Are* you crazy?" […] "I *suppose* you just *don't care* if you *kill* yourself, *do* you?"
"It*'s* my self"
"I *suppose* you just *don't care* if you *lose* your leg, *do* you?"
"It*'s* my leg"
"It certainly *is* not your leg!" […] "That leg *belongs* to the U.S. government."

Surprisingly, though, it is at least as common for fictional dialogue to occur with frequent past tense and perfect aspect verbs, being quite different from typical face-to-face conversation in this regard. Fiction authors rely on these past-time features in dialogue because they use dialogue to move the narrative story forward, and thus characters often report past events in their interactions (see the text excerpts in [5]).

5. Short dialogues from three novels, illustrating the dense use of past tense and perfect aspect in fictional interpersonal interactions.
[past tense and perfect aspect verbs shown in italics]

The Insidious Dr. Fu-Manchu
Sax Rohmer

"Ever *seen* one like it?" he *asked*.
"Not exactly," I *confessed*. "It appears to *have been* deeply *cauterized*."
"Right! Very deeply!" he *rapped*. "A barb steeped in the venom of a hamadryad *went* in there!"
[…]
"There's only one treatment," he *continued*, rolling his sleeve down again,
"and that's with a sharp knife, a match, and a broken cartridge.
I *lay* on my back, raving, for three days afterwards, in a forest that *stank* with malaria, but I should *have been* lying there now if I *had hesitated*.
Here's the point. It *was* not an accident!"

Masters of Space
E.E. Smith and E. Everett Evans

"Mr. Ashby, *did* you have your interspace rigs set?"
"No, sir. I *didn't think* of it, sir."

"Doctor Cummings, why *weren't* yours out?"
"I *didn't think* of such a thing, either—any more than you *did*," Sandra *said*.

The Highest Treason
Randall Garrett

"The Board of Strategy *asked* me to tell you," Tallis *continued*. "After all, my recommendation *was* partially responsible for the decision." […] "It *was* a hard decision, Sepastian—you must realize that. We *have been* at war with your race for ten years now. We *have taken* thousands of Earthmen as prisoners, and many of them *have agreed* to co-operate with us […]"

In contrast, it turns out that many novels—especially modern novels—employ a present-time style for narrative and descriptive prose, and as a result these novels have large negative scores for Dimension 2. Excerpt [6] illustrates this style:

6. *The Middleman*
Olen Steinhauer
[Present tense verbs in italics]

All day I *sit* by the lime green swimming pool, sun-screened so I won't turn black, going through my routine of isometrics while Ransome's indios *hack* away the virgin forests. Their hate is intoxicating. They *hate* gringos—from which my darkness *exempts* me—even more than Gutierrez. They *hate* in order to keep up their intensity.

I *hear* a litany of presidents' names, Hollywood names, Detroit names—Carter, *chop*, Reagan, *slash*, Buick, *thump—bounce* off the vines as machetes *clear* the jungle greenness.

We spoke a form of Spanish in my old Baghdad home. I always *understand* more than I *let on*.

For these same reasons, Dimensions 1 and 3 are not strongly related, having a Pearson correlation of only r = -0.19 for the 185 novels analyzed in the 2008 study. Figure 2 plots 15 of these novels in the two-dimensional space, illustrating this lack of a strong relationship. Although novels occupy much of the space, there is a noticeable absence of novels that have large scores for both dimensions. For example, *To the Lighthouse* has the largest positive score for Dimension 3 ("Past orientation") but a score near 0.0 for Dimension 1; *The Middleman* similarly has the largest negative score for Dimension 3 ("Present orientation") but a score near 0.0 for Dimension 1. And we find a very similar pattern with respect to Dimension 1: *Room with a View* and *The House on Pooh Corner*

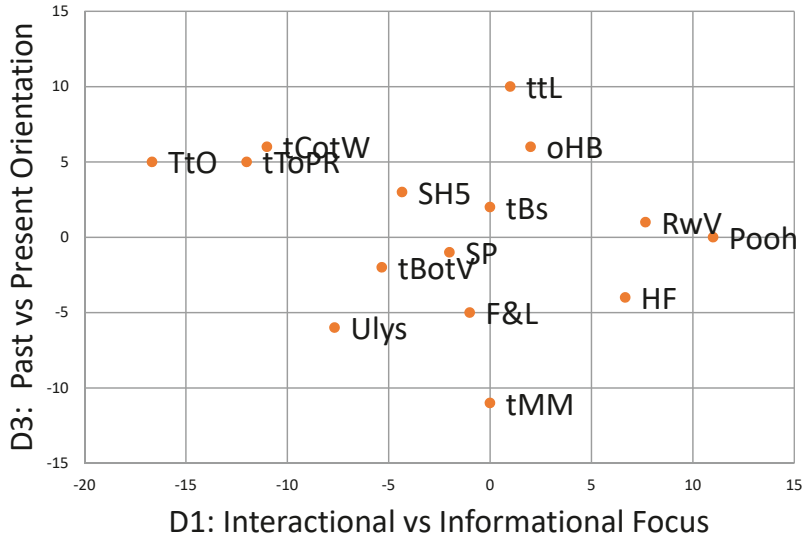

Fig. 2 Dimension 1 versus Dimension 3 scores for selected novels. Key: Pooh = *House on Pooh Corner*; RwV = *A Room with a View*; HF = *Huckleberry Finn*; oHB = *Of Human Bondage*; ttL = *To The Lighthouse*; tBs = *The Borrowers*; tMM = *The Middleman*; F&L = *Fear and Loathing In Las Vegas*; SP = *Mr Sammler's Planet*; SH5 = *Slaughterhouse-Five*; tBotV = *The Bonfire Of The Vanities*; Ulys = *Ulysses*; tCotW = *The Call of the Wild*; tToPR = *The Tale Of Peter Rabbit*; TtO = *Tarka the Otter*. (Biber/Egbert, CC BY)

both have large positive scores for Dimension 1 ("Interactional/involved focus") but scores near 0.0 for Dimension 3. The only exception to this generalization is *Tarka the Otter*, which has a very large negative score for Dimension 1 ("Informational focus") coupled with a moderately large positive score for Dimension 3 ("Past orientation"). In general, though, these two parameters are largely unrelated, indicating that authors can choose between a highly interactive/involved style versus a highly informational style independently of their choice between a highly narrative style versus highly present-time oriented style.

5. Conclusion

Both of the previous MD analyses of fictional novels identify an additional dimension that is peculiar to the discourse domain of fictional novels. In the 2008 study, this dimension was interpreted as "Concrete actions/events versus abstract description," opposing phrasal verbs, activity verbs, progressive aspect, concrete nouns, and place adverbials versus nominalizations, mental nouns, abstract nouns, long words, and attributive adjectives. And similarly, the 2012 study identified a third dimension with almost the same interpretation: "Abstract exposition versus concrete action." Authors like Sammler (in the 2008 study) and Kipling (in the 2012 study) relied heavily on the linguistic features associated with concrete action, while authors like Eliot relied heavily on the features of abstract description/exposition. Surprisingly, Twain's *Adventures of Huckleberry Finn* was also marked for the dense use of "abstract description" features.

This third dimension does not have a direct counterpart in MD studies of other discourse domains. Rather, it reflects a functional distinction that is important for distinguishing among fictional novels: those stories that are action-oriented versus novels with a much stronger focus on the description of people and places, which often also includes commentary on motives, actions, or society in general. In this regard, these MD studies of fictional novels are similar to all previous MD studies, in that they have all identified dimensions of variation that are specialized to a discourse domain or language. These specialized dimensions reflect the particular communicative priorities of each language/culture or domain of use.

From both theoretical and methodological perspectives, it is not surprising that each MD analysis would uncover specialized dimensions that are peculiar to a given language and/or discourse domain. After all, each of these studies differs with respect to the set of registers represented in the corpus for analysis, and the set of linguistic features included in the analysis. Given those differences, it is reasonable to expect that the parameters of variation that emerge from each analysis will be fundamentally different. And to some extent, this expectation is met, with specialized dimensions emerging in nearly all MD analyses.

However, given this background, the existence of universal dimensions of variation that emerge in nearly all MD studies is quite unexpected. Two of these dimensions are especially important, regardless of the discourse domain: a dimension associated with 'oral' versus 'literate' discourse, and a dimension associated with narrative discourse.

The robustness of narrative dimensions across languages and discourse domains indicates that this rhetorical mode is basic to human communication, whether in speech or in writing. Rhetoricians and discourse analysts have long argued for the central role of narration in communication. MD studies confirm that claim, showing the importance of this rhetorical mode in virtually all discourse domains (spoken and

written; interpersonal and informational; etc.). And, as we show in Section 4 above, this functional parameter is also important for distinguishing among fictional novels, both in the nineteenth century and in the twentieth century.

But the most surprising pattern discovered through MD analysis is the oral/literate opposition, which emerges as the very first dimension in nearly all MD studies. In studies based on general corpora of spoken and written registers, this dimension clearly distinguishes between speech and writing. However, other studies show that this is not a simple opposition between the spoken and written modes. In fact, this dimension emerges consistently in studies restricted to only spoken registers, as well as studies restricted to written registers.

In terms of communicative purpose, the 'oral' registers characterized by this dimension focus on personal concerns, interpersonal interactions, and the expression of stance. In contrast, 'literate' registers focus on the presentation of propositional information, with little overt acknowledgment of the audience or the personal feelings of the speaker/writer. Linguistically, this first dimension opposes two discourse styles: an 'oral' style that relies on pronouns, verbs, adverbs, versus a 'literate' style that relies on nouns and nominal modifiers. The oral style relies on clauses to construct discourse—including a dense use of dependent clauses. In contrast, the complexity of the literate style is phrasal.

It turns out that this same opposition is fundamentally important for distinguishing among the styles of fictional novels. Authors like Melville, and even some children's novels like *Tarka the Otter*, are notable for their dense reliance on an informational style employing phrasal grammatical features. At the other extreme, novels like *House on Pooh Corner* and *Room with a View* are marked by their highly 'oral' style relying on verbs, pronouns, adverbs and dependent clauses. As such, we have shown here how variation in the discourse domain of fictional novels is patterned in similar ways to all other discourse domains in that it reflects the two general functional parameters of 'oral'/'literate' discourse and 'narrative'/'non-narrative' discourse, while at the same time being organized with respect to additional functional-linguistic dimensions.

References

Biber, Douglas. 1986. "Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings." *Language* 62 (2): 384–414.

Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 1995. *Dimensions of Register Variation: A Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Biber, Douglas. 2001. "Dimensions of Variation among Eighteenth-Century Speech-Based and Written Registers." In *Multi-Dimensional Studies of Register Variation in English*, edited by Susan Conrad and Douglas Biber, 200–14. London: Longman.

Biber, Douglas. 2006. *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, Douglas. 2008a. "Corpus-Based Analyses of Discourse: Dimensions of Variation in Conversation." In *Advances in Discourse Studies*, edited by Vijay Bhatia, John Flowerdew, and Rodney H. Jones, 100–14. London: Routledge.

Biber, Douglas. 2008b. "Using Corpus-Based Analysis to Study Fictional Style: A Multi-Dimensional Analysis of Variation among and within Novels." Invited plenary lecture, International Society for the Empirical Study of Literature, University of Memphis.

Biber, Douglas. 2014. "Using Multi-Dimensional Analysis to Explore Cross-Linguistic Universals of Register Variation." *Languages in Contrast* 14 (1): 7–34.

Biber, Douglas. 2019. "Multi-Dimensional Analysis: A Historical Synopsis." In *Multi-Dimensional Analysis: Research Methods and Current Issues*, edited by Tony Berber-Sardinha and Marcia Veirano Pinto, 11–26. London: Bloomsbury.

Biber, Douglas, Bethany Gray, and Shelley Staples. 2016. "Predicting Patterns of Grammatical Complexity across Language Exam Task Types and Proficiency Levels." *Applied Linguistics* 37 (5): 639–68.

Egbert, Jesse. 2012. "Style in Nineteenth Century Fiction: A Multi-Dimensional Analysis." *Scientific Study of Literature* 2 (2): 167–98.

Friginal, Eric. 2009. *The Language of Outsourced Call Centers*. Amsterdam: John Benjamins.

Goźdź-Roszkowski, Stanislaw. 2011. *Patterns of Linguistic Variation in American Legal English: A Corpus-Based Study.* Frankfurt am Main: Peter Lang.

Gray, Bethany. 2013. "More than Discipline: Uncovering Multi-Dimensional Patterns of Variation in Academic Research Articles." *Corpora* 8 (2): 153–81.

Reppen. Randi. 2001. "Register Variation in Student and Adult Speech and Writing." In *Multi-Dimensional Studies of Register Variation in English*, edited by Susan Conrad and Douglas Biber, 187–99. London: Longman.

Xiao, Richard. 2009. "Multidimensional Analysis and the Study of World Englishes." *World English* 28 (4): 421–50.