

Register variation in understudied academic contexts

Larissa Goulart
Montclair State University / USA

Abstract – A major focus of register research has been language variation in academic discourse. These studies describe patterns of language use in spoken and written academic texts. Although there have been numerous studies of this type, most have focused on academic registers in English and on descriptions of published academic registers (e.g. textbooks, research articles, and abstracts). Much less work has been carried out on academic registers in other languages or unpublished academic registers. This special issue presents five studies describing the language patterns of understudied academic discourse in English (learners' writing and statutory law), as well as descriptions of published academic registers in languages other than English (Russian, Portuguese, and Arabic). We hope that the papers in this special issue will pave the way for future research in other understudied academic contexts.

Keywords – register studies; academic writing, understudied academic context

From a text linguistic perspective, registers are described as “text varieties that are defined by the situational characteristics of a text” (Gray and Egbert 2019: 1), with situational characteristics being used to refer to attributes, such as communicative purpose, mode of communication, addressor, etc. (see Biber and Conrad 2019: 40). Register studies seek to describe the linguistic profile of a group of texts and explain the use of these linguistic features in terms of their function in these texts. That is, register researchers believe that there is an intrinsic relationship between the situational characteristics and the linguistic profile of a text (or register). Therefore, a register investigation usually includes a situational analysis, a linguistic analysis, and a functional interpretation. This three-pronged analysis is usually referred to as the ‘register triangle’.

To date, the register approach has been applied to describe several discourse domains, from face-to-face conversations (e.g. Biber 2006; Quaglio and Biber 2006) to the language of popular science (e.g. Egbert 2016), to academic discourse (e.g. Gray 2015). In fact, Goulart and Wood (2021) show that academic registers are the most common types of discourse domain investigated using multidimensional analysis, a



method often used in register studies. Nevertheless, most of this research has focused on published academic registers (e.g. research articles, textbooks, and abstracts), especially on research articles. In Goulart and Wood's (2021) survey, the authors find 31 studies describing the language patterns of research articles alone. In addition, register studies have focused primarily on academic registers written in English. That is, few studies have examined the language patterns of academic discourse in languages other than English or in unpublished academic registers such as student writing and grant proposals, among others.

The articles in this special issue of *Research in Corpus Linguistics* seek to shed some light into the language patterns of understudied academic contexts from two different perspectives. First, we have a set of two papers that discuss language variation in understudied academic contexts in English student writing and in English legal language. Second, we have a set of three papers that describe the language of research articles in languages other than English (Portuguese, Russian, and Arabic).

In the first paper, **Magali Paquot, Damien De Meyere, Hilde Hasselgård, Tove Larsson, Signe Oksefjell Ebeling, Natalia Judith Laso, Hubert Naets, Larry Valentin, Isabel Verdaguer, and Sanne van Vuuren** describe the compilation process of the *Varieties of English for Specific Purposes dAtabase* corpus (VESPA) and discuss possible applications of this corpus to explore register variation in learner writing. VESPA is a collection of texts written by learners of English with different first languages across three disciplinary components: linguistics, business communication, and literature. Texts in the corpus are classified into the same register categories as texts in the *Michigan Corpus of Upper-level Student Papers* (MICUSP; Römer and O'Donnell 2011), allowing for comparisons between these two corpora. The authors conclude their paper with a case study illustrating the applications of VESPA to investigate register variation in learner writing. In the case study, the authors report on the results of a multidimensional analysis that compares the language profiles of argumentative texts and research papers in different corpora. Finally, they argue that VESPA can help us further understand the language patterns of a somewhat understudied academic register: that of university writing produced by English learners.

In the second paper, **Margaret Wood** discusses the language patterns of the understudied register of English statutory law. Her study examines the extent to which the language of statutory law differs from plain language. To this end, Wood conducts a

key-feature analysis comparing the use of several lexico-grammatical features between states statutes and a corpus of popular written language comprising news, sports reports, encyclopedia articles, historical and magazine articles. It is worth noting that the corpus of state statutory law is part of the *United States State Statutes* (CorUSSS) Corpus (Egbert and Wood, in preparation), which comprises the state codes for each of the 50 states in the United States. The results show that statutory language is denser in terms of clausal embedding, with more passive voice, prepositions, and *wh*- and *that* relative clauses than plain language. Such findings provide a broad overview of the language patterns encountered in state statutes.

In the third paper, **Marine Matte and Elisa Stumpf** examine the use of reporting verbs in Portuguese research articles. To date, few studies have investigated the language patterns of academic Portuguese (Hoffnagel 2010; Bessa 2011). The authors seek to bridge this gap by examining how authors in hard and soft sciences include sources in their research articles. With this goal in mind, Matte and Stumpf first retrieve occurrences of the word *autor* ‘author’ followed by verbs from the *Corpus of Portuguese for Academic Purposes* (CoPEP; Kuhn and Ferreira 2020). Secondly, they retrieve the most frequent verbs in both hard and soft sciences and analyze them in terms of their structural use (tense, mood, and aspect) and their semantic use (research, cognition, and discourse) in the corpus. The findings show that there is a considerable overlap between the verbs that introduce citations in hard and soft sciences and further suggest that sources are usually incorporated with verbs in the simple present or simple past tenses. Matte and Stumpf’s study is innovative in its approach to identify reporting verbs. In the conclusion, the authors highlight the need for further research dealing with the language patterns of academic Portuguese.

Examining the use of lexical bundles in research articles, **Maria Kostromitina** analyzes cross-linguistic transfer in writing between English as a second language (L2) by Russian native speakers and Russian as a first language (L1). To this end, the author compiles a corpus of research articles in educational psychology by L1 Russian writers, L2 English writers, and L1 English writers. Kostromitina retrieves lexical bundles from these three corpora and investigates the L2 English bundles that are mirrored in the L1 Russian and L1 English research articles. The results show that bundles produced by L2 English writers who are native speakers of Russian have a considerable overlap in form with the bundles extracted in the L1 Russian corpus. More importantly, the study shows

one possible approach to investigating language transfer when examining the language patterns of English learners.

In the last paper, **Mai Zaki** investigates the use of metadiscourse and rhetorical features in Arabic academic abstracts. The author compiles a corpus of 400 abstracts in Arabic from research articles and dissertations. Zaki analyzes the extent to which the use of metadiscourse features varies across 1) types of abstracts (dissertation or research articles) and 2) authors' gender (male, female, or mixed). The abstracts are annotated for their use of transition markers, frame markers, evidentials, endophorics, and code glosses, following Hyland's (2019) framework. The results show that engagement markers are quite frequent in Arabic abstracts. Interestingly, the study also finds that transition and frame markers are the most frequent types of metadiscourse within textual markers. This study provides insights into how Arabic academic writers use language features that can engage their readers with the text.

This collection of papers displays a range of different methods (key-features, multidimensional analysis, lexical bundles, etc.) used to describe the language patterns of understudied academic registers. We hope that these studies will motivate further research on other understudied academic registers that are central to academic life but rarely published, such as grant proposals, personal statements, or fellowship applications, among others.

REFERENCES

- Bessa, Jose Cezinaldo Rocha. 2011. (Re)pensando a citação em textos acadêmico-científicos. *Signum: Estudos da Linguagem* 14/2: 421–439.
- Biber, Douglas. 2006. *University Language*. Amsterdam: John Benjamins.
- Biber, Douglas and Susan Conrad. 2019. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Egbert, Jesse. 2016. Stylistic perception. In Paul Baker and Jesse Egbert eds. *Triangulating Methodological Approaches in Corpus-linguistic Research*, 167–182. London: Routledge.
- Egbert, Jesse and Margaret Wood. (In preparation). *Constructing and Designing a Specialized Corpus of Statutory Law (CorUSSS)*.
- Goulart, Larissa and Margaret Wood. 2021. Methodological synthesis of research using multi-dimensional analysis. *Journal of Research Design and Statistics in Linguistics and Communication Science* 6/2: 107–137.
- Gray, Bethany. 2015. *Linguistic Variation in Research Articles*. Amsterdam: John Benjamins.
- Gray, Bethany and Jesse Egbert. 2019. Register and register variation. *Register Studies* 1/1: 1–9.

- Hoffnagel, Judith C. 2010. A prática de citação em trabalhos acadêmicos. *Cadernos de Linguagem e Sociedade* 10/1: 71–88.
- Hyland, Ken. 2019. *Metadiscourse: Exploring Interaction in Writing*. London: Bloomsbury.
- Kuhn, Tanara Zingano and José Pedro Ferreira. 2020. O Corpus de Português Escrito em Periódicos-CoPEP. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 36/2: 1–42.
- Quaglio, Paulo and Douglas Biber. 2006. The grammar of conversation. In Bas Aarts and April McMahon eds. *The Handbook of English linguistics*, 692–723. Oxford: Blackwell.
- Römer, Ute and Matthew Brook O'Donnell. 2011. From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 6/2: 159–177.

Corresponding author

Larissa Goulart
Montclair State University
Department of Linguistics
1 Normal Ave.
NJ 07043 Montclair
USA
e-mail: goulartl@montclair.edu