

# FORUM

## *Genre-Based Instruction and Corpora*

**KEN HYLAND**

*University of East Anglia  
Norwich, UK; and  
Jilin University  
Jilin, China*

doi: 10.1002/tesq.3273

### INTRODUCTION TO GENRE

■ Genre, understood most simply, is a term for grouping texts together, representing broad rhetorical templates that writers draw on to respond to recurring situations: what users see as effective ways of getting things done using language. The importance of genre in second language writing instruction is that it allows teachers to understand and make explicit to students, the ways that texts can be written to achieve particular purposes. It is an approach to writing instruction that goes beyond helping learners to generate content, work through drafts or practice grammatical forms, and allows them to see texts as attempts to communicate with readers. It therefore counteracts any tendency to treat individual texts in isolation from others or to see grammar separately from use. In other words, genre helps teachers to theorize the common-sense labels we use to categorize texts and to unpack the ways they are structured and used; it provides a way of identifying the kinds of texts that students write in their target contexts and organizing courses to meet these needs.

Genre, then, is an approach to describing and understanding communication built on a social view of language in context. It is based on the idea that members of a community usually have little difficulty in recognizing similarities in the texts they use frequently and are able

to draw on these repeated experiences to read, understand, and perhaps write them relatively easily. This is because writing (and speaking) is based on mutual expectations with writers and readers acting on assumptions about what the other will recognize and expect. In other words, we assemble sense from a text by making connections to prior texts to anticipate the other's actions. Coherence is achieved by activating a relevant *schema* of prior knowledge which we share with others. We know immediately, for example, whether a text is an essay, a joke, or a recipe, but we can also recognize innovation, irony, and creativity. Genres can thus be seen as a kind of tacit contract between writers and readers, influencing the behavior of text producers and the expectations of receivers. It is the unfamiliarity of these genres which is one reason why writing is often difficult for L2 students.

We all have a repertoire of these schemata and develop new ones as we need them, but this does not mean that we simply follow well-worn tracks. We are all different and interpret genres in ways that best suit our goals and personalities so established patterns often form the basis of variations (Tardy, 2009).

While genre approaches differ considerably in the emphasis they give to text or context, the research methods they employ, and the types of pedagogies they encourage, text-analytic varieties have had the most impact in classroom applications. These linguistic approaches are influenced by Halliday's (2004) Systemic Functional (SFL) view of language as a system of choices that link texts to particular contexts through patterns of lexico-grammatical and rhetorical features (Rose & Martin, 2012). Pedagogically, the main inspiration for genre in classrooms was Bernstein's (1990) ideas advocating "a visible pedagogy" to provide a more equal distribution of knowledge in schools irrespective of students' socioeconomic and cultural background. The failure of 'progressive' process writing curricula to help disadvantaged children, therefore, inspired research to assist learners in controlling valued genres. Concepts such as the teaching and learning cycle (Rothery, 1994), the zone of proximal development (ZPD), and scaffolding (Vygotsky, 1978) were key to this work, as was the development of a metalanguage to help teachers better understand language and pedagogy.

A second, closely related, perspective on genre is that associated with English for Specific Purposes (ESP) which sees genre as a tool for understanding and teaching the kinds of writing required of nonnative English speakers in academic and professional contexts. The ability to function competently in a range of written genres is often a central concern for ESL learners as it can determine their access to career opportunities, positive identities, and life choices. As a result, ESP investigates the demands placed by academic or workplace contexts on

communicative behaviors and addresses a cross-cultural and L2 dimension of writing instruction.

Influenced by the seminal work of Swales (1990), ESP operates with a more restricted view of genre than SFL. Instead of seeing genres as the linguistic strategies for achieving broad rhetorical goals in a culture such as *narrative*, *argument*, *recount*, etc, ESP explores the specific practices of target groups. Recurrent classes of communicative events such as *case analyses*, *lab reports*, and *abstracts* are the genres that orchestrate everyday life in communities. Genre analysts therefore set out to offer descriptions of “typified acts of communication” based on the form and purposes of texts. The emphasis genre advocates place on sharedness, patterns, regularities, typification, and recurrence, however, leads critics to position genre instruction as discriminatory and its classrooms as sites of forced linguistic homogenization. But rather than being a simple tool of conformity, genres are an essential prerequisite for an informed understanding of the contingency of prestige discourses. Genres provide insights into the workings of the academy for learners and assist individuals to enter a community with understanding as well as the ability to change it.

In reality, genre is often taught recognizing that these are plans not molds and with the sensitivity to acknowledge that form and genre vary across contexts (e.g., Hyland, 2018; Tardy, 2009). Choice is made possible by constraint, by awareness of what options are conventionally available and what different choices mean to readers and hearers. When writers manipulate established forms this is usually a subtle redrawing of a genre, confined within the boundaries of what is recognized as conventional practices. It also involves a good grasp of the resources for creating meanings and a certain confidence to depart from convention. As Bakhtin (1986, p 80) has suggested, writers must be able to control the genres they use before they can creatively exploit them. So genres embody diversity and facilitate innovation, but innovation is only possible in the context of convention. Moreover, every time a genre is used it is performed in a unique way. They are not static forms but living, changing resources for users.

## GENRE AND CORPUS-BASED LANGUAGE INSTRUCTION

Corpora and genre are essential companions. Because corpora provide language data which represent a speaker’s experience of language in a particular domain, they offer evidence of typical genre patterns. They bring an empirical dimension to the study of writing and speech, so providing support for our intuitions, strength for our interpretations and generally allowing us to talk about genres with greater

confidence. This approach contrasts markedly with impressionistic methods of text analysis which tend to produce partial and prescriptive findings. Corpora have therefore informed genre instruction by providing students with naturally occurring, authentic data instead of invented examples, by revealing the phraseological basis of language use, contributing to curricula developments and teaching materials, and encouraging a more autonomous, inductive approach to language learning. Corpora are particularly valuable in classrooms where students are learning English for specific academic or occupational purposes as teachers are attempting to teach the genres their students most urgently need.

It is common to distinguish between *indirect applications* of corpora, where research findings feed into syllabuses, references works and teaching materials, and *direct applications*, which refer to hands-on involvement by teachers and learners in classrooms.

In terms of *indirect applications*, corpora have led to new descriptions of language which have ensured that learner dictionaries and reference grammars are now based on frequency of occurrence and typical collocational patterns. Corpora have also informed specialized word lists for such fields as environmental science (Liu & Han, 2015) and finance (Ha & Hyland, 2017). Notable textbooks have also emerged informed by corpus data. These Corpora here not only provide greater access to context and more examples. The 4-level *Touchstone* series, which draws on the Cambridge English Corpus, also helps learners to understand general principles for participating in conversational genres. Similarly, Swales and Feak's (2012) *Academic Writing for Graduate Students* employs data from the MICASE corpus and follows a task-based approach to develop a sensitivity to the language conventions of different advanced academic genres such as the conference poster and sections of the dissertation.

*Direct applications* of corpora in genre classrooms encourage students to engage with corpora in a more hands-on way. So-called data-driven learning (DDL) approaches, owe much to the work of Tim Johns (1991) who sought to replace instruction with discovery. DDL is essentially an inductive, consciousness-raising methodology that requires learners to discover facts about language use by analyzing authentic corpus examples. DDL activities encourage students to notice features of syntagmatic patterning which illustrates the typical genre constraints on individual items through the recurrent phrases they find in the data. Teachers can ask students to engage with raw corpus data or mediate the experience by prior selection of concordance lines to highlight target features, and this seems to be the preferred approach as students, on their own, are not always sure what they should look for, nor are they always sufficiently interested in

language to conduct research on it. So while students are asked to analyze the features of text extracts and develop an exploratory attitude toward rhetorical practices, this is often achieved by combining inductive activities (asking learners to infer a rule) with deductive activities (asking learners to apply a rule).

One example of corpus-informed instruction is Wong's (2019) DDL project for over 330 postgraduate students each year in a thesis writing course at the University of Hong Kong. The project involves a purpose-built, in-house tool designed to be easy to use, provide a large corpus, employ a straightforward query interface, and track learner usage. The corpus comprises 431 postgraduate theses of nearly 11 million words from 52 departments and rated "excellent" (the top 10% of all theses in the university). Users can search the whole database by sections, faculties, and disciplines. They can also narrow their searches by checking boxes to filter for discipline-specific lexis and phrases as well as by sections of the thesis (e.g., introduction, methods, etc.) to learn more about key features of theses in disciplinary writing.

The accompanying tasks are integrated into the existing materials of the course and based on the issues that students encounter in thesis writing (e.g., use of self-mention; identifying a research gap; using tenses in literature reviews, etc.). The DDL materials come after relevant language features are introduced and critical discussion of texts, and aim to raise students' awareness of key features of successful disciplinary writing. They make use of exemplars as well as student-driven discovery of the knowledge and skills required for good writing. The output provides concordance lines expandable to entire texts and has a range of data visualization options such as frequency hits, collocations, and item distributions allowing students to draw conclusions. A learning analytics function provides teachers with feedback on student use, giving information such as duration of use, query syntax, and types of searches. The results of 11,000 corpus queries show significant interdisciplinary trends and variation in the use of corpus functions, with students frequently going beyond the DDL materials to generate their own queries (Crosthwaite, Wong, & Cheung, 2019).

One problem with using corpora of this kind is that some learners may find the language of the corpus too difficult or the output too large for them to manage. One solution to this is to have each student build their own small do-it-yourself (DIY) corpus of a target genre in their own discipline. Charles (2017) has championed this approach in academic writing courses for multi-disciplinary classes in order to provide students with corpora specific to their own needs and this has been taken up in other contexts. The idea is that if students create their own corpora, they are familiar with the content and can more easily understand the language of the concordance lines while not

being overwhelmed with examples. These advantages were generally confirmed by both students and teachers in her postcourse questionnaire.

## **SUGGESTIONS FOR FUTURE RESEARCH**

Despite the documented strengths of using specialized corpora in EAP genre instruction (e.g., Viana & O'Boyle, 2022), they remain underused. DDL is still treated as an out-of-class activity and so most DDL studies are relatively small. In addition, for many teachers and learners problems remain concerning how to gain access to corpora, the lack of specialized corpora, and the complexity of corpus query tools. There is, then, a need for both larger-scale research into the effectiveness of corpus-informed instruction of genre and of the best ways to use it in classrooms. Some suggestions are given below.

### **Impact and Reception of Corpus Genre Instruction**

Not all learners are comfortable interacting with a target corpus and teachers may be unsure how to find corpus data, how to use the tools, what questions to ask, and how to interpret the output. There is therefore a need, using questionnaire and interview data, to better understand student and teacher perceptions of DDL materials, corpus platforms, and the effectiveness of DDL for disciplinary writing. What are teachers' and students' attitudes to corpus-based instruction, and what do they see as the greatest challenges and affordances of the approach? Moreover, do these attitudes vary with proficiency level, discipline, or teacher experience and what kind of training do they believe might be useful to better prepare them for using corpora?

### **Corpus Tool Improvements for DDL**

To assist improvements in the creation and application of DDL-specific corpus tools, developers and teachers need to consider the needs of learners and the ways in which software can most effectively contribute to their learning. By gathering feedback on existing tools and practices, learners can play an important role in improving the search and output tools they use and offering suggestions for how they might be improved. The fact that many students have grown up in a digital environment that exposes them to dynamic, visually rich, virtual worlds might help inform corpus use and lead to simpler, more motivating,

interfaces. Potential questions to follow here might be: What corpus interrogation tools do particular student groups find most engaging or useful? What tools are best able to integrate visual and textual materials? Is student language learning best assisted by graphical, tabular, or textual corpus outputs? Can corpora be more effectively integrated at the point of need, that is, when students are writing or speaking?

## Effective Classroom Practices to Teach Genre Using Corpora

Further research would also be useful on the use and effectiveness of different corpora and the kinds of tasks that are best able to exploit them in different contexts. What kinds of tasks work best with a homogenous class of learners from a single, specialist discipline, for example, and do these differ for a heterogeneous class from a range of specialist disciplines? How can we encourage learners to work independently with their own individualized, specialist corpus to discover discipline-specific language features and collectively to discover features that they can then discuss with others? There is also a need for studies investigating the longitudinal effects of corpora and DDL on writing in the medium to long term. Do students extend their use of target items beyond the DDL materials and into the drafting and revision stages of writing? There is also a need to move beyond corpus activities that focus on grammar and lexis to facilitate analysis of discourse or genre-level features, so how might this best be achieved?

**THE AUTHOR.** Ken Hyland is an Honorary Professor at the University of East Anglia. He has published 294 articles and 29 books on writing and academic discourse with over 83,000 citations on Google Scholar. According to the Stanford/Elsevier analysis of the Scopus database, he is the world's most influential scholar in language and linguistics.

## CONFLICT OF INTEREST

No funding or conflict of interest to report.

## REFERENCES

- Bakhtin, M. (1986). *Speech genres and other late essays*. Austin, TX: University of Texas Press.
- Bernstein, B. (1990). *The structuring of pedagogic discourse*. London: Routledge.

- Charles, M. (2017). Do-it-yourself corpora in the EAP classroom: Views of students and teachers. In K. Hyland & L. Wong (Eds.), *Faces of English*. London: Routledge.
- Crosthwaite, P., Wong, L. L. C., & Cheung, J. (2019). Characterising graduate students' corpus query and usage patterns for data-driven learning. *ReCALL*, 31(3), 255–275.
- Ha, A. Y. H., & Hyland, K. (2017). What is technicality? A technicality analysis model for EAP vocabulary. *Journal of English for Academic Purposes*, 28, 35–49.
- Halliday, M. A. K. (2004). *Introduction to functional grammar* ((3rd ed) ed.. revised by C. Matthiessen). London: Edward Arnold.
- Hyland, K. (2018). Sympathy for the devil? A defence of EAP. *Language Teaching*, 51(3), 383–399.
- Johns, T. (1991). Should you be persuaded—two examples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom concordancing. English language research journal* (4) (pp. 1–16). Birmingham, UK: The University of Birmingham.
- Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, 39(1), 1–11.
- Rose, D., & Martin, J. R. (2012). *Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney school*. London: Equinox.
- Rothery, J. (1994). *Exploring literacy in school English*. Sydney: Metropolitan East Dis-advantaged Schools Program.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Swales, J., & Feak, C. B. (2012). *Academic writing for graduate students* (3rd ed.). Michigan University Press: Michigan.
- Tardy, C. (2009). *Building genre knowledge*. West Lafayette, IN: Parlor Press LLC.
- Viana, V., & O'Boyle, A. (2022). *Corpus linguistics for English for academic purposes*. Oxford: Routledge.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, Trans.). Cambridge, MA: Harvard University Press.
- Wong, L. L. C. (2019). Implementing disciplinary data-driven learning for post-graduate thesis writing. In K. Hyland & L. L. C. Wong (Eds.), *Innovation and change in language education*. London: Routledge.