Computer Assisted Language Instruction Consortium

calico journal (online) issn 2056–9017

Article

Generating Genre-Based Automatic Feedback on English for Research Publication Purposes

Stephanie Link[1], Robert Redmon[2], Yaser Shamsi[3], and Martin Hagan[4]

Abstract

Artificial intelligence (AI) for supporting second language (L2) writing processes and practices has garnered increasing interest in recent years, establishing AI-mediated L2 writing as a new norm for many multilingual classrooms. As such, the emergence of AI-mediated technologies has challenged L2 writing instructors and their philosophies regarding computer-assisted language learning (CALL) and teaching. Technologies that can combine principled pedagogical practices and the benefits of AI can help to change the landscape of L2 writing instruction while maintaining the integrity of knowledge production that is so important to CALL instructors. To align L2 instructional practices and CALL technologies, we discuss the development of an AI-mediated L2 writing technology that leverages genre-based instruction (GBI) and large language models to provide L2 writers and instructors with tools to enhance English for research publication purposes. Our work reports on the accuracy, precision, and recall of our network classification, which surpass previously reported research in the field of genre-based automated writing evaluation by offering a faster network training approach with higher accuracy of feedback provision and new beginnings for genre-based learning systems. Implications for tool development and GBI are discussed.

Affiliations

[1] Oklahoma State University, USA.
Corresponding author: steph.link@okstate.edu

[2] Oklahoma State University, USA.
email: robert.redmon@okstate.edu

[3] Oklahoma State University, USA.
email: yaser.shamsi@okstate.edu

[4] Oklahoma State University, USA. email: mhagan@ieee.org

vol 41.1 2024 1–28

Received: 23 August 2023
Accepted after revision: 21 January 2024

calico journal

https://doi.org/10.1558/cj.26273

Stephanie Link, Robert Redmon, Yaser Shamsi, and Martin Hagan

Keywords: genre-based feedback; automated writing evaluation; natural language processing; transformer network; writing for publication.

1. Introduction

Advancements in artificial intelligence (AI) offer new opportunities and challenges for supporting second language (L2) writers in generating and disseminating knowledge. One affordance is through the use of automated writing evaluation (AWE) for improving the quality of academic essay writing (e.g., Dizon & Gayed, 2021; McCarthy et al., 2022). Recent discussion has emphasized that AI can change the landscape of L2 English for research publication purposes (ERPP), requiring scholarly writers to find new avenues of leveraging the affordances of AI-mediated technologies in ethical ways that maintain the integrity and morality of science and scholarship (Habibie & Starfield, 2023). This advancement toward AI-mediated ERPP, however, requires technological developments that move beyond earlier AWE research on general essay writing and into socially situated writing practices (Burstein et al., 2016), thus extending from a focus on lower-order subconstructs of writing (e.g., grammar, usage, mechanics, style) to higher-order writing subconstructs (e.g., rhetorical effectiveness).

Currently, very few technologies provide AWE feedback for ERPP. Those that exist align well with genre-based instruction (GBI), forming a shift from traditional AWE to *genre-based AWE*. Genre-based instruction is a dominant pedagogical approach in graduate and L2 writing programs, and is most frequently based on Swales' (1981) Create-A-Research-Space (CARS) model, which denotes genres in terms of *moves* or "rhetorical units that perform a coherent communicative function" (p. 228) and functional *steps*, or sub-moves related to the communicative purpose of the move (Dudley-Evans & St John, 1998). The CARS model has motivated technology developers to construct feedback systems, also known as genre-based AWE, which classifies lexico-grammatical patterns into moves and steps, provides feedback on rhetorical organization, and thus demonstrates how meaning is communicated in rhetorically effective ways. One of the first genre-based tools to emerge was Anthony and Lashkia's (2003) tool called Mover, which analyzes research article abstracts and splits them into move-labeled sentences. More recently, Cotos (2014) and Knight et al. (2020) developed the Research Writing Tutor and AcaWriter, respectively. These genre-based AWE tools provide asynchronous feedback on move/step usage to reflect the rhetorical conventions of select genres. Expectedly, developments in genre-based AWE are limited because they require domain-specific content and applicable pedagogies which are both labor intensive and difficult to model computationally. Therefore, current developments in genre-based AWE tend to generalize feedback based on relatively small datasets of representative, domain-specific texts. With the rapid advancements of machine learning and natural language processing (NLP), emerging techniques can push the boundaries of genre-based AWE affordances and offer more rapid development and expanded GBI affordances.

In this article, we investigate how a combined approach of using NLP and large language models can be used to more rapidly and accurately train a genre-based feedback engine to classify sentences based on move/step rhetorical conventions. Large language models are the algorithmic bases for many disruptive technologies, such as ChatGPT by OpenAI. Despite their growing popularity, large language models remain under-explored in AWE research, and thus offer new possibilities for advancing genre-based AWE research. We describe a streamlined workflow of classifying sentences based on genre features derived from a large dataset of 40,000 research articles from the Elsevier Open Access CC-BY Corpus (Kershaw & Koeling, 2020). We fine-tuned our pre-trained network to classify sentences by using an iterative process of training the network, modifying network miscategorizations, and retraining the network to improve accuracy. Our approach was integrated into Dissemity, a GBI system with evaluative and suggestive AWE feedback on ERPP, which offers new beginnings for AI-mediated ERPP. Our work can be integrated into a genre-based pedagogical framework to support a vast range of scholarly writers across disciplines.

2. Review

2.1 Feedback and Genre-Based Instruction

Feedback is broadly defined as any corrective, suggestive, or evaluative information on performance or understanding that an agent (e.g., teacher, self, technology) provides to a writer (Hattie & Timperley, 2007). Although the benefits of corrective feedback are occasionally debated (Truscott, 1996, 1999), automated feedback has demonstrated impact on lowering (meta)cognitive barriers (Gayed et al., 2022; Ranalli et al, 2017), supporting revision and retention of grammatical concepts (Link et al., 2022), enhancing learner engagement during the writing process (Saeli et al., 2023; Zhang, 2020), among other benefits (see Zhai & Ma, 2022, for a meta-analysis). Genre-based AWE research has revealed additional affordances, including the attention to and enhancement of linguistic features to fulfill rhetorical purposes (Cotos & Huffman, 2013; Feng & Chukharev-Hudilainen, 2022). These genre-based AWE studies have derived from an English for specific purposes (ESP) approach to genre studies (Swales, 1990), enabling our research to extend previous work along similar lines of inquiry.

Although there are interrelated approaches to genre studies (see Hyon, 1996; Johns, 2002), the ESP approach is arguably the most influential in ESP and English for academic purposes research (Bhatia, 1993; Johns, 2002; Swales, 1990). It is also widely adopted in teaching discipline-specific writing (Flowerdew, 2015; Swales & Feak, 2012). Additional schools of thought in genre studies include New Rhetoric and systemic functional linguistics. The New Rhetoric, New Literacy, and Academic Literacies tradition focuses on the socio-rhetorical climates that influence writers' choices and rhetorical structures as they relate to audiences and purposes (Devitt, 2004; Freedman, 1993). Systemic functional linguistics (Halliday, 1994) is a theory of language that centers around the notion of language function in a social context (i.e., what language does and how it does it within the constraints and affordances of a situational and cultural context). The ESP approach often defines genre as any socially recognized discourse that adheres to language use and formal conventions delimited by goal-oriented, communicative purposes and the demands of social-rhetorical contexts (Bahktin, 1986, Johns, 2002; Swales, 1990, 2004).

Applying this definition to practice, GBI aims to develop learners' genre knowledge, which includes an understanding of how a genre functions in particular contexts and situations to meet the needs of the target discourse community (Tardy, 2009). Genre knowledge can contribute to rhetorical flexibility, so that learners can respond to various writing demands. To build genre knowledge, rhetorical consciousness-raising tasks (Swales, 1990) are a hallmark of ESP genre-based training, designed to elevate learners' awareness of genre features and their function (Hyland, 2007; Paltridge, 2019; Swales, 1981, 1990). This explicit instruction includes the analysis and reproduction of the rhetorical structure in model texts (Bhatia, 1993; Hyland, 2007).

The CARS model used within a complementary genre-based pedagogical framework (Swales & Feak, 2012) has far-reaching success in both first and second language learning environments (Cheng, 2008; Tardy, 2006), offering implications for training scientific writers from various fields and language backgrounds. While much of the move/step analysis process should be performed inductively by analyzing successful writing to support the transfer of knowledge to other genres, bottom-up analysis, such as lexical approaches and some corpus-based pedagogies, can be complementary (Flowerdew, 2015) by offering detailed descriptions of how lexis and phraseology contribute to move structure.

Although there has been some debate about the role of explicit genre-based teaching (Freedman, 1993; Williams & Colomb, 1993), empirical studies have argued that explicit teaching can enhance language awareness and improved knowledge of genre-specific language choices (Yasuda, 2011), increase attention to rhetorical parameters that shape a genre (Cheng, 2008, 2011), and utilize knowledge of move structure to organize writing (Huang, 2014). The positive outcomes of GBI have motivated a few technological developments to focus on move structure analysis for publication purposes, marking the start of genre-based AWE development.

2.2 Genre-Based Automated Writing Evaluation

Altogether, genre-based AWE has received limited attention in the field of genre studies. However, there are far-reaching implications, since genre-based feedback can be used at all stages of a rhetorical reading-to-write process (Cheng, 2008) to raise awareness of genre-based writing conventions and expand genre knowledge. Mover (Anthony & Lashkia, 2003) was the first tool to explore genre-based text classification. Other systems have emerged more recently with genre-based feedback capabilities—the Research Writing Tutor (Cotos, 2014) and AcaWriter (Knight et al., 2020). Table 1 provides a comparative overview of two tools that directly address ERPP.

While results from the support vector machine approach in Cotos and Pendar (2016) seemed to improve on the performance of Anthony and Lashkia's (2003) NaïveBayes classifier in most step-level classifications, the researchers acknowledged the limits of each approach and felt it prudent to investigate the influence of context on classifications. Recent advances in AI have made machine learning techniques, particularly the use of neural network models, increasingly accessible and progressively viable for genre-based AWE researchers to explore the influence of context. Neural networks are fundamentally mathematical models designed to learn from data provided to them in a process referred to as "training." Fiacco et al. (2019) have shown that neural network models can be trained to recognize specific genre-based features more accurately and reliably than traditional rule-based approaches, approaching even human levels of accuracy when classifying step categories (average = 77%). Although exact step-level accuracy is not reported, findings from Fiacco et al. (2019) suggest that context-aware approaches may improve domain modeling and should thus be explored further.

Table 1:
Comparison of Genre-based AWE Classifiers for English for Research Publication Purposes (ERPP)

Tool	Classifier	Genre	Move/step categories	Accuracy
Mover[a]	NaïveBayes classifier	Research article abstracts	Move 1: Establish a territory	
			– Claim centrality	28%
			– Generalize topics	82%
			– Review previous research	Unknown
			Move 2: Establish a niche	
			– Counterclaim	Unknown
			– Indicate a gap	17%
			– Raise questions	Unknown
			– Continue a tradition	Unknown
			Move 3: Occupy the niche	
			– Outline purpose	Unknown
			– Announce research	92%
			– Announce findings	66%
			– Evaluate research	57%
			– Indicate research article structure	Unknown
Research Writing Tutor[b]	Support vector machine classifiers	Research article introductions[c]	Move 1: Establishing a territory	
			– Claiming centrality	67.9%
			– Making topic generalizations	70.4%
			– Reviewing previous research	86.7%
			Move 2: Identifying a niche	
			– Indicating a gap	75.2%
			– Highlighting a problem	64.7%
			– Raising general questions	50.0%
			– Proposing general hypotheses	66.3%
			– Presenting a justification	68.9%
			Move 3: Addressing the niche	
			– Introducing present research descriptively	50.6%
			– Introducing present research purposefully	78.6%
			– Presenting research questions	84.6%
			– Presenting research hypotheses	74.2%
			– Clarifying definitions	100.0%
			– Summarizing methods	44.6%
			– Announcing principal outcomes	51.4%
			– Stating the value of the present research	39.8%
			– Outlining the structure of the paper	92.0%

[a] Based on Anthony and Lashkia (2003).

[b] Based on Cotos and Pendar (2016).

[c] Research Writing Tutor addresses all sections of a research article. Relevant to the current study are results for Introduction sections.

2.3 Advancements in AI for Genre-Based Automated Writing Evaluation

While machine learning is a rapidly advancing domain of practice, much remains to be learned regarding its applicability to genre-based AWE. The effectiveness of a neural network model can be seen as a product of both the data used and the kind of network trained. The model described by Fiacco et al. (2019) starts with a manually labeled, 900-article corpus used to train a neural network without predefined parameters for understanding general language use. The recent emergence of pre-trained language models—such as GPT, BERT, and ELECTRA—are significant departures from traditional networks. Pre-trained language models are networks pre-trained on large quantities of unlabeled general domain language data, which can be "fine tuned" for specific NLP tasks (Howard & Ruder, 2018). In a sense, the pre-trained language model provides the network with a basic understanding of what language looks like, and the data for fine tuning teach the network to perform a specific task.

Existing pre-trained language models generally include two categories: language models and masked language models. Language models, such as GPT, process text from left to right and only use the context to the immediate left of a token to predict a given output. Masked language models are advantageous because they process text bi-directionally (left to right, right to left), using the context on both sides of a token to predict output. The output prediction, however, is based on only a small subset of words, limiting the amount that can be learned from a sentence. To this end, the ELECTRA network offers novel strides in NLP tasks (Clark et al., 2020), outperforming other networks (e.g., GPT and RoBERTa) when given the same model size, data, and compute (e.g., processing power, memory, storage). The ELECTRA network utilizes "replaced token detection (RTD)" (p. 1) that involves jointly training two transformer models: the generator and the discriminator. The generator is a rather standard masked language model. The discriminator, most novel to the ELECTRA, learns from all all language tokens, rather than a small subset, as a step toward learning the language represented in the data. After pre-training, the generator is no longer used, and the discriminator can be fine-tuned with additional, more focused input (such as genre-based rhetorical categorizations) to perform on downstream tasks. Integrating pre-trained language models into genre-based systems can thus offer new opportunities for enhancing the quality of feedback for ERPP. Therefore, this study investigated the following research question: *With what level of accuracy, precision, and recall can a neural network be trained to classify sentences within a genre-based framework?*

3. Methodology

Figure 1 introduces the model architecture for genre-based sentence classification. Data were derived from an open collection of peer-reviewed journal articles (Kershaw & Koeling, 2020). Natural language processing tools, referred to hereafter as Wrangler NLP, were developed to streamline the fine-tuning process, which served as the input to the sentence classifier. Iterative training and testing of the neural network produced output in the form of genre-based sentence classifications that were analyzed for accuracy. Model details are elaborated in the following sections.

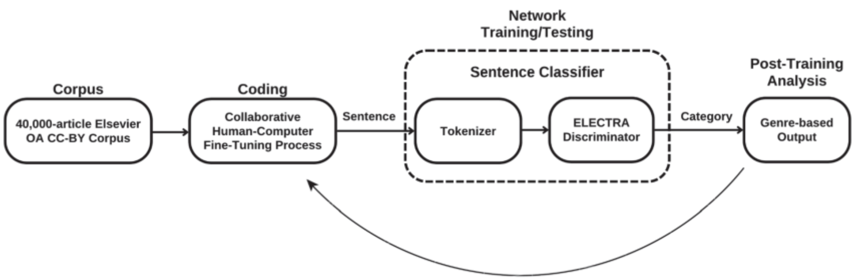

Figure 1: Model architecture for genre-based sentence classification of research articles.

3.1 Corpus Data Description

The Elsevier OA CC-BY Corpus is an open access corpus of 40,001 scientific research articles from 27 scientific subject classifications. See Kershaw and Koeling (2020) for a discussion about representativeness and the discipline naming scheme. The corpus contains articles published by Elsevier since 2014 and is covered by CC-BY 4.0 license. The data used for this project consist of sentences extracted from the Introduction sections of the corpus (Table 2).

Table 2:
Description of Introduction Corpus (*N* = 40,001 articles)[a]

Discipline	Number of articles	Number of sentences	Number of words
General	310	150	29,729
Agricultural and Biological Sciences	3,985	62,432	1,876,865
Arts and Humanities	1,014	17,653	531,481
Biochemistry, Genetics and Molecular Biology	8,417	109,082	3,007,886
Business, Management and Accounting	1,002	19,446	541,283
Chemical Engineering	2,196	18,862	490,266

(*Continued*)

Table 2 (*Continued*)

Discipline	Number of articles	Number of sentences	Number of words
Chemistry	2,749	30,100	808,625
Computer Science	3,004	48,912	1,254,795
Decision Sciences	530	3,278	91,114
Earth and Planetary Sciences	2,764	56,041	1,741,304
Economics, Econometrics and Finance	1,081	30,055	805,300
Energy	2,845	36,619	987,060
Engineering	5,962	69,299	1,795,747
Environmental Science	6,241	104,463	3,096,125
Immunology and Microbiology	3,258	33,571	963,713
Materials Science	4,008	48,899	1,283,808
Mathematics	1,561	17,634	438,745
Medicine	9,225	121,547	3,463,865
Neuroscience	3,277	69,403	2,127,540
Nursing	310	4,753	145,499
Pharmacology, Toxicology and Pharmaceutics	2,233	41,248	1,179,043
Physics and Astronomy	3,927	43,935	1,145,430
Psychology	1,796	42,803	1,310,115
Social Sciences	3,623	65,325	1,930,825
Veterinary	1,010	3,589	108,269
Dentistry	43	595	15,941
Health Professions	774	1,793	47,389

[a] Each article can belong to multiple disciplines based on ASJC (All Science Journal Classification) codes.

3.2 Coding Scheme for Human–Computer Fine-Tuning Process

Introduction sections were analyzed for steps associated with three overarching moves (see Appendix). This framework draws on the simplicity of the CARS model and the refinement of Cotos et al.'s (2016) cross-disciplinary Introduction move/step model, resulting in an eight-category step organization (Table 3). This simplified framework intends to lessen the cognitive load on novice researchers (Singh, 2014) when using our technology for reading articles and analyzing genre-based feedback.

Human–computer processing was performed in two stages: (1) development, testing, and validation of regular expressions (regex) for automatic pattern markup, and (2) manual annotation of sentences based on functional step categories (Figure 2).

Regex is a pattern-matching language that can be used to locate and manage strings of text. Regex development started with a bottom-up genre-based analysis by using the Introduction move/step framework to identify representative lexico-grammatical patterns from a random sample of research articles.

Table 3:
Move/Step Classification

Move	Step	Shorthand	Coding category
Establish a territory	Provide background	background	0
	Claim centrality	centrality	1
Identify a niche	Present justification	justification	2
	Problematize research	problem	3
Address the niche	State contribution	contribution	4
	Outline study	outline	5
	Announce purpose	purpose	6
	Highlight study specifics	specifics	7

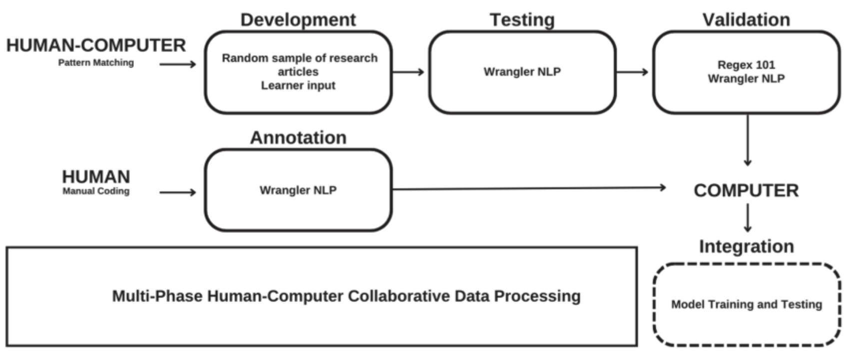

Figure 2: Collaborative Human–Computer Fine-tuning Process.

Patterns were input into the Wrangler NLP to evaluate the pattern markup against the whole corpus (see below), which led to the addition and revision of regex. For example, our initial observation was that the linguistic patterns similar to "important implications" functioned regularly to "Address the niche—State contribution," resulting in the following regex.

Original regex pattern:
/((ha(s|ve)) (\S)?implication(s)?/

However, we found that this expression functions differently across moves/steps, as shown in examples 1–2 (emphasis added to show lexico-grammatical patterns).

Example 1 (Identify a niche—Problematize research): This lack of clarity about what climate change skepticism actually is <u>has important implications</u>.

Example 2 (Address the niche—State contribution): Our findings can be used to build the evidence base for [...], which has important implications for the HIV/AIDS epidemic.

Example 1 shows how "has important implications" functions to problematize research due to a "lack of clarity." In Example 2, the inclusion of "Our findings …" connects the initial pattern to the study's contribution. Thus, the regex was revised to ensure precise identification of study contributions, as shown here.

Revised regex pattern:
/([Tt]h(ese|is)|[Oo]ur)(finding(s)?|research|result(s)|study|work)(\S*)?((ha(s|ve)|hold|possess|yield)(s)?) (\S)?implication(s)?/

Regular expressions were validated using RegEx 101 (https://regex101.com) and Wrangler NLP to ensure that sentences classified by each expression were indicative of the respective step. Figure 3 illustrates how Wrangler NLP was used in the regex testing/validation process. The left side shows the regex for each step category. By selecting an expression, the sentence matches from the Introduction corpus are highlighted on the right side. New or modified regex can be entered into the search bar at the top right, which then auto-populates new matches.

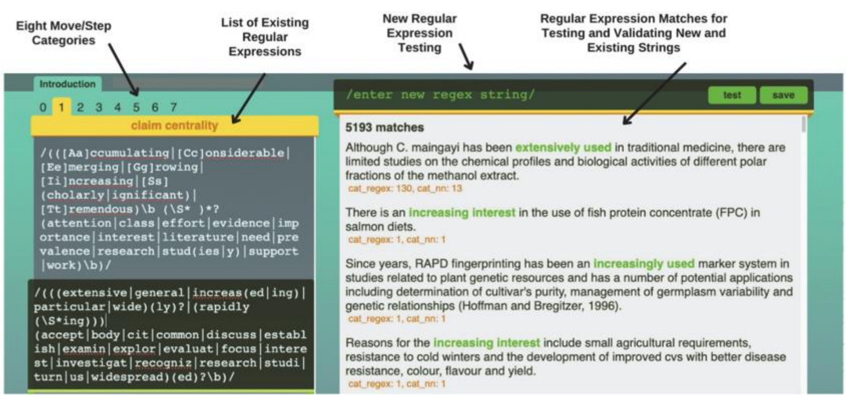

Figure 3: Wrangler NLP for testing and validating regular expressions for collaborative pre-processing.

Manual annotation was also performed in Wrangler NLP. Figure 4 shows how each sentence was coded in context with one or more steps to account for sentences with a combination of rhetorical strategies. Two coders were involved in the calibration process. Using the Introduction coding scheme, both coders

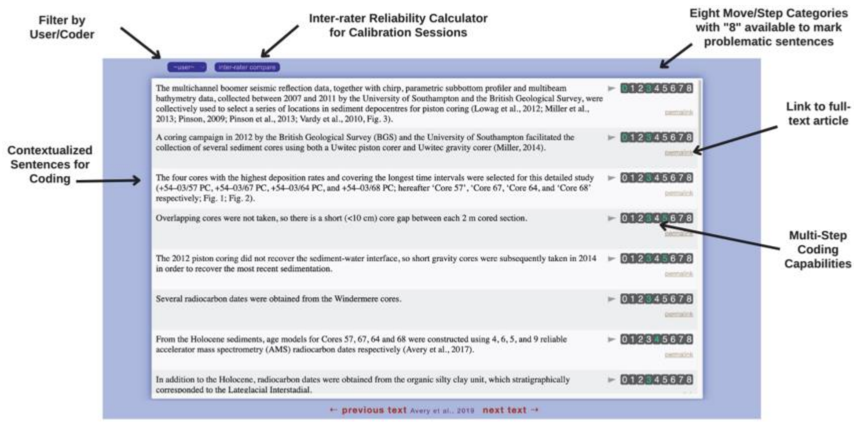

Figure 4: Wrangler NLP for manual annotation of steps in contextualized sentences.

analyzed five texts, and the Wrangler NLP automatically calculated the percentage agreement. The coders met to discuss discrepancies, and the calibration continued until inter-rater reliability was over 80% in perfect agreement; at that time, the second coder continued to code the remaining data.

3.3 Network Training and Testing

For model training and testing, the Introduction corpus was processed using the regex, the human labels, and a combination of both. After one round of training and post-training analysis (described below), the regex was modified and human labels were added to update the dataset and improve the accuracy of output. Table 4 shows the human and regex groups for the final dataset.

Table 4:
Distribution of Regex and Human Labels for Network Training

Dataset	Number of unique sentences
Regex	192,969
Human	24,832
Human, regex union	200,594
Human, regex intersection	17,207

For training, we divided the dataset into two parts: a training set and a test set. The test set was never used (either directly or indirectly) during training;

errors were computed on the test set after training was completed to predict network generalization performance for future data. Because a single test set might not be fully representative, we retrained the network multiple times, with a different test set randomly held out each time in a five-fold cross-validation process. We trained five different networks. For each network, a different 20% of the data were used for the test set. In this way, all the original sentences were used in one of the five test sets.

The final neural network used to classify the sentences is the ELECTRA small discriminator (Clark et al., 2020), which is a transformer network. We began with a pre-trained network downloaded from the Huggingface Transformers Library (https://huggingface.co)and fine-tuned it on our training sets. Before sentences were input to the network, they passed through the default tokenizer for the ELECTRA small discriminator, also downloaded from the Huggingface Transformers Library. We trained the network for the multi-label case, since each sentence could be assigned to more than one step.

3.4 Post-Training Analysis

Figure 5 illustrates four subgroups of test set data that were used to check the accuracy of sentence classification. The region marked HO represents sentences with "human only" labels, thus forming a baseline standard for evaluation of network performance. The RO region contains "regex only" labels. The HRA region contains sentences with both human and regex labels, and the labels agree. The HRD region contains sentences with both human and regex labels, and the labels disagree. We analyzed these four regions individually at completion of training to provide insights into the quality of human and regex labels, and to determine how to improve results.

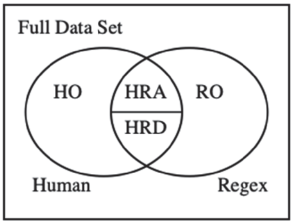

Figure 5: Four subgroups of test set data for analysis of sentence classification accuracy. HO = human only labeled sentences; RO = regex only labeled sentences; HRA = human and regex labeled sentences and labels agree; HRD = human and regex labeled sentences and labels disagree.

After training the networks, we computed for each step category both precision and recall, which are standard evaluation metrics used in machine learning to determine the performance of a classifier. Precision measures the accuracy of network predictions, while recall measures the completeness of positive predictions. Accuracy was calculated based on Exact match, Jaccard, and Hamming accuracies, which again are standard metrics for evaluating the performance of machine learning models with multi-label classifications (Park & Read, 2019). Table 5 defines the performance measures for sentence classification.

Table 5:
Performance Measures for Network Classification of Sentences

Symbol	Name/definition	Formula
P	Positives	
N	Negatives	
FP	False positives, type I errors	
FN	False negatives, type II errors	
TN	True negatives	
TP	True positives	
	Precision	TP(TP + FP)
	Recall	TP/(TP+FN)
	F1-score	Average precision + Recall

4. Results: Accuracy, Precision, and Recall

Table 6 shows the Exact match, Jaccard, and Hamming accuracies on the training and test sets. Accuracies are slightly lower on the testing set, but still very close to the training set. The network has an exact match (matching all labels) on 96.4% of the testing sentences, which can be compared to some extent to Fiacco's (2019) Cohen's kappa results, showing the accuracy of step-level prediction to be 75.1%.

Table 6:
Full Dataset Performance

Training set		**Testing set**	
Statistic	Value	Statistic	Value
Exact match accuracy	99.7%	Exact match accuracy	96.4%
Jaccard accuracy	99.2%	Jaccard accuracy	97.3%
Hamming accuracy	99.9%	Hamming accuracy	99.5%

Table 7 shows the test set precision, recall, and F1-score for each category. The precision for seven of the eight categories is over 96%, and the minimum precision is 88.8%, which is for category 2—Justification. The recall for six of the eight categories is over 98%, and the minimum recall is 93.6%, which is also for category 2—Justification.

Table 7:
Full Test Set Error Statistics Across Five-Fold Cross-Validation Process

Step category	Precision	Recall	F1-score
0 Background	98.7%	98.8%	98.8%
1 Centrality	98.1%	99.2%	98.7%
2 Justification	88.8%	93.6%	91.2%
3 Problem	97.0%	98.9%	97.9%
4 Contribution	94.4%	97.1%	95.8%
5 Outline	97.6%	98.6%	98.1%
6 Purpose	96.6%	98.0%	97.3%
7 Specifics	98.2%	98.6%	98.4%
Average	97.2%	97.9%	97.0%

The results for the HO group on the test set are the key to judging the network performance. These are sentences that were not labeled by regex, and the network did not have access to these sentences during training. We would expect the network performance on these data to be similar to its performance on new sentences. Table 8 shows the precision, recall, and F1-score for each category.

The average precision across all categories is 77.4%. In categories 5—Outline and 6—Purpose, the precision is also close to 100%. The average recall across all categories is 79.6%. In categories 5–Outline and 6—Purpose, the recall is also close to 100%. Categories 5 and 6 are the categories that have the most example sentences based on regex and human labeling, suggesting that additional human–computer processing will improve the network.

Overall, the results map close to Fiacco et al. (2019), who found an average of 77% precision and 77% recall across all step categories compared to our 77.4% and 79.6%, respectively. A lack of reported data in Fiacco et al. (2019) does not make it possible to compare step-level accuracies directly to the results found here. However, it is more important to align the current results with Cotos

Table 8:
Test Set Error Statistics for Human only (HO) Labeled Sentences Across Five-Fold Cross-
Validation Process

Step category	Precision	Recall	F1-score
0 Background	84.0%	77.2%	80.4%
1 Centrality	49.3%	66.5%	56.6%
2 Justification	74.9%	78.0%	76.4%
3 Problem	57.0%	66.3%	61.3%
4 Contribution	66.7%	64.2%	65.4%
5 Outline	99.2%	99.5%	99.3%
6 Purpose	99.0%	98.8%	98.9%
7 Specifics	89.5%	86.3%	87.9%
Average	77.4%	79.6%	78.3%

and Pendar (2016), which represents the current model used in the existing Research Writing Tutor (see Table 9).

Table 9:
Comparison Between Research Writing Tutor and Dissemity Step-Level Accuracy

Research Writing Tutor—step accuracy[a] (%)		Dissemity—step accuracy (%)	
Move 1: Establishing a territory		Move 1: Establishing a territory	
Claiming centrality	67.9	Claim centrality	49.3
Making topic generalizations + Reviewing previous research	78.6	Provide background	84
Move 2: Identifying a niche		Move 2: Identifying a niche	
Indicating a gap + Highlighting a problem + Raising general questions + Proposing general hypotheses	64.1	Problematize research	57
Presenting a justification	68.9	Present a justification	99
Move 3: Addressing the niche		Move 3: Addressing the niche	
Introducing present research descriptively + Introducing present research purposefully	64.6	Announce purpose	89.5

(*Continued*)

Table 9 (*Continued*)

Research Writing Tutor—step accuracy[a] (%)		Dissemity—step accuracy (%)	
Presenting research questions + Presenting research hypotheses + Clarifying definitions + Summarizing methods + Announcing principal outcomes	71.0	Highlight study specifics	74.9
Stating the value of the present research	39.8	State contribution	66.7
Outlining the structure of the paper	92.0	Outline study	99.2

[a] Step-level accuracies from Table 1 were averaged to show direct comparison with step- level accuracies in the present study.

This comparison shows that Dissemity outperforms Research Writing Tutor's step-level classifications in most categories, with the exception of "Claim centrality" and "Problematize research." Although these two steps will require additional attention as development continues, the results provide evidence that the ELECTRA classifier can provide reliable classification relative to comparative technologies, and is thus a viable option for integrating into a new genre-based AWE tool.

4.1 Dissemity Integration for Genre-Based Feedback Generation

The results offer evidence that sentence classification can provide positive implications for feedback generation. We thus integrated the neural network into Dissemity—for disseminating research with clarity (https://dissemity.com)—a genre-based learning system for supporting novice and emerging researchers with ERPP. The system contains a series of interactive, interconnected modules grounded in GBI that guide users through the reading-to-write process on their path to publication. The first is *Discover*, which introduces an inductive reasoning process to orient learners to socio-rhetorical patterns in published texts. The second is *Learn*, containing instructional videos, quizzes, and note-taking options. The third is *Analyze*, which enables users to inductively and deductively evaluate the argument structure in published research articles. The fourth is *Explore*, which is used to identify and archive high-frequency lexico-grammatical patterns that enable writers to communicate meaning and purpose based on the conventions of a discipline. The final module is *Write*, used to stimulate written production with the aid of resources from other modules.

The Analyze and Write modules contain automated feedback derived from the ELECTRA network, which provides step-level classifications of each sentence the user analyzes or writes. The regex helps to underline indicative patterns that can be archived or explored further in the Explore module. In cases of multiple network predictions, the step category with the highest confidence interval is assigned to a sentence. If the system is confident in a secondary prediction, that category is also assigned to represent the multifunctionality of discourse units. Dissemity feedback in the Analyze module might be: "You correctly marked this sentence as *identify the niche*. Good job!" In the Write module, feedback might be: "Dissemity is confident that this sentence is being used to *problematize the research*. Consider using more specific language to add clarity."

In the Analyze module (see Figure 6), users find a model article from the corpus or individually upload articles. After importing the article into the module, users can select each sentence and assign it a primary move and step that are associated with contrasting colors to help visualize the argument structure of the text. During this reading process, Dissemity underlines functional units to help users notice move/step patterns and provides evaluative feedback in the form of red or green checks. By selecting each check mark, users receive a pop-up window containing the feedback and access to other modules for increasing understanding of the feedback. When the system is not confident of the accuracy in the user's annotation, no check mark is provided to minimize distractors.

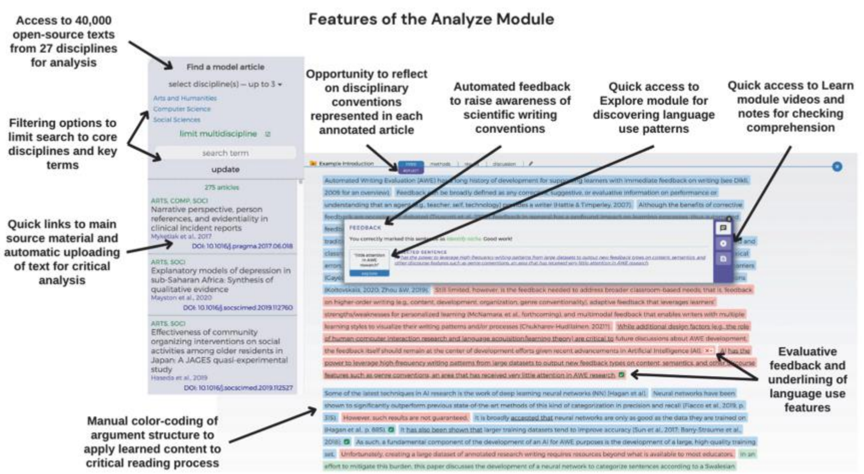

Figure 6: Dissemity Analyze module with automatic feedback on learner annotation of disciplinary texts.

In Write, Dissemity provides three forms of feedback: (1) automated color coding to represent moves/steps and the argument structure of a manuscript; (2) section-level feedback on the comparison between the user's manuscript and conventions in target disciplines; and (3) sentence-level feedback on the user's communicative intentions at the micro-level (Figure 7). In the feedback window, users see the percentage of move/step representation, along with the number of sentences and words in their manuscript compared to the corpus. Dissemity uses this information to provide suggestive feedback. For example: "The most notable difference between your manuscript and the norms for Computer Science is that you have more blue." To address this comment, users can obtain sentence-level feedback, which evaluates each sentence for move/step intentions, so users can determine whether their intended meaning is communicated adequately. Users then have quick access to other modules to revisit prior learning.

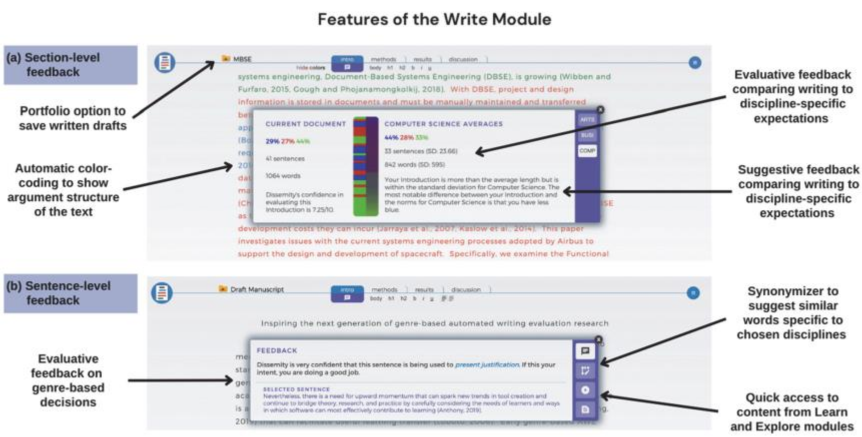

Figure 7: Dissemity Write module with automatic color coding and section/sentence- level feedback.

4.2 Implications for Dissemity Within a Genre-Based Pedagogical Framework

The main aim of GBI is to raise students' rhetorical consciousness through analysis of lexico-grammatical features of a genre, and to develop skills for becoming more aware of socio-rhetorical contexts and organizations (Cheng, 2021). Despite some debate about whether the goal of GBI is genre acquisition (see Tardy et al., 2020, for a discussion), the aim with Dissemity is to centralize genre awareness. This section outlines how to integrate Dissemity into a genre-based pedagogical framework to develop "skillful and rhetorically aware learners of genres" (Cheng, 2021, p. 35).

4.2.1 Discovery-Based Orientation

As a first stage of GBI, discovery-based orientation heightens learners' awareness of the generic features and rhetorical parameters in a text, and supports the transfer of rhetorical consciousness to future writing tasks (Cheng, 2008). Dissemity's Discover module facilitates this process. Learners can upload model research articles or choose from the open-source corpus. In a classroom- or lab-based setting, discussions about the genre, text, rhetorical situation, discourse-level move/step features, and general language features (e.g., style, cohesion) should take place (Swales & Feak, 2012) by examining model articles. Questions informed by Cheng (2008) enable learners to write reflections on their observations. For example: "What was the author(s) trying to do with the first two sentences in the text?" and "What are the words, phrases, or sentences that the author used to achieve this purpose?" These questions should guide learners in uncovering a set of heuristics for analyzing genres, rather than imposing rules for genre construction.

4.2.2 Genre Knowledge Activation

Given that the research article genre has been thoroughly explored, there are existing frameworks (e.g., the CARS framework) that can be used to raise learners' rhetorical consciousness. Dissemity's Learn module introduces the schema in the Appendix and guides learners to complete analytical tasks that support their understanding of macro-level issues (Tardy, 2017) and the function of lexico-grammatical features (Cheng, 2021). Learners should be encouraged to take notes and quizzes to deepen their understanding of the rhetorical context (who the audience is and what audience expectations are), the discourse (how text is organized), and the language (Paltridge, 2019).

4.2.3 Genre Exploration

A general understanding of the research article framework can be deepened through additional analytical tasks in the Analyze module. Here, learners analyze each sentence in their model corpus, annotate each sentence with a move/step category, and receive feedback. If a teacher is involved in the learning process, they can provide additional feedback about the socio-rhetorical organization and context. Reflecting on prompting questions draws learners' attention to argument structures and to the lexico-grammatical features underpinning the rhetorical organization of the text.

4.2.4 Lexico-Grammatical Feature Identification

Data-driven learning tasks (Boulton & Vyatkina, 2021) within the Explore module can provide the nexus between rhetorical patterns of writing and the lexico-grammatical features that enable communicative functioning. This focus on language can help learners connect lexico-grammatical features with rhetorical functions by helping them explore frequent, salient, and unique linguistic features within rhetorical moves and steps. Teachers can comment on learner-archived features and provide reflective questions, such as "What verb tenses do you notice?" or "How is hedging used?" or "What evaluative statements can you find that help to achieve this category's purpose?" Teachers can also use the Explore module to develop activities and materials for in-class activities.

4.2.5 Application and Assessment

Dissemity's Write module can be used to apply learned and practiced tenets of GBI to the writing of a research article. Learners should be encouraged to outline their rhetorical organization with the move/step framework in mind. They can then begin drafting. In addition to genre-based AWE feedback, teachers can comment in real time on any aspect of the learner's writing, but most important with regard to GBI are comments on the effectiveness in representing the target genre (e.g., cohesion, coherence, convention). Teachers can ask learners to complete a "reflective cover letter" (Tardy et al., 2023, p. 78), where learners can evaluate and critique their progress.

5. Conclusion

Automated analysis of rhetorical structures in scientific research articles has the potential to foster learners' awareness of disciplinary-specific genre conventions. With this aim in mind, we presented a genre-based learning tool with automated feedback capabilities to boost learners' rhetorical consciousness regarding how to enter communities of scientific writers. We demonstrated how our network classifier leverages context awareness to perform accurately with high precision and recall, as compared to other approaches (Cotos & Pendar, 2016; Fiacco et al., 2019). Furthermore, we highlighted the potential for the neural network output to offer additive system affordances, including multi-disciplinary comparisons and phrase frame analysis of functional units within moves/steps, all of which are a considerable departure from existing scientific writing technologies. As noted, the feedback is suggestive and evaluative, and as development continues, corrective feedback will be explored. Our future work will also expand Dissemity feedback affordances, including support for multimodal visualization of feedback, expanded genre support, and opportunities for adaptive learning.

Acknowledgments

This work was supported by the National Science Foundation (grant numbers 2044642, 1912226); the Oklahoma State University Vice President for Research Office and College of Arts and Sciences Special Research Grant; and the Oklahoma State University Technology Business Development Program. A special thanks to the *CALICO Journal* editors and anonymous reviewers for their constructive comments during the review process.

About the Authors

Stephanie Link is an Associate Professor of Applied Linguistics at Oklahoma State University. Her research focuses on automated writing evaluation tools and intelligent tutoring systems for second language writing and written scientific communication. She is the Editor of the *Advances in CALL Research and Practice* book series and Book Review Editor for the *English for Specific Purposes Journal*. Her work has been published in notable journals, such as *CALICO Journal*, *Computer Assisted Language Learning*, and *Language Learning & Technology*. Her latest project (Dissemity, funded through the National Science Foundation) integrates genre-based pedagogy and artificial intelligence to help developing writers disseminate scientific results.

Robert Redmon is a postdoctoral researcher at Oklahoma State University, working on the development of Dissemity, a genre-based writing instruction platform with AI-driven automated writing evaluation features. His research interests are in corpus linguistics, discourse analysis, and natural language processing. He served as entrepreneurial lead for a recent grant through the National Science Foundation Innovation Corps National Program to bring Dissemity into the commercial market.

Yaser Shamsi is a PhD student in TESOL and Applied Linguistics at Oklahoma State University. Currently serving as a graduate research associate, he is involved in developing Dissemity, an online writing tool designed to provide feedback using AI-driven technology. His research interests involve second language acquisition, automated writing evaluation, and genre analysis.

Martin Hagan is Professor Emeritus of Electrical and Computer Engineering at Oklahoma State University, Stillwater, where he has taught and conducted research in the areas of statistical modeling, neural networks, and dynamic systems since 1986. He is the author, with H. Demuth and M. Beale, of the textbook *Neural*

Network Design (Boston: PWS, 1994, 2nd ed. 2014). He was also a co-author of the *Neural Network Toolbox* (now *Deep Learning Toolbox*) for MATLAB until 2015.

References

Anthony, L., & Lashkia, G. V. (2003). Mover: A machine learning tool to assist in the reading and writing of technical papers. *IEEE Transactions on Professional Communication*, *46*(3), 185–193. https://doi.org/10.1109/TPC.2003.816789

Bahktin, M. M. (1986). *Speech genres and other late essays*. Austin: University of Texas Press.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Harlow: Longman.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, *25*(3), 66–89. http://hdl.handle.net/10125/73450

Burstein, J., Elliot, N., & Molloy, H. (2016). Informing automated writing evaluation using the lens of genre: Two studies. *CALICO Journal*, *33*(1), 117–141. https://doi.org/10.1558/cj.v33i1.26374

Cheng, A. (2008). Analyzing genre exemplars in preparation for writing: The case of an L2 graduate student in the ESP genre-based instructional framework of academic literacy. *Applied linguistics*, *29*(1), 50–71. https://doi.org/10.1093/applin/amm021

Cheng, A. (2011). Language features as the pathways to genre: Students' attention to non-prototypical features and its implications. *Journal of Second Language Writing*, *20*(1), 69–82. https://doi.org/10.1016/j.jslw.2010.12.002

Cheng, A. (2021). The place of language in the theoretical tenets, textbooks, and classroom practices in the ESP genre-based approach to teaching writing. *English for Specific Purposes*, *64*, 26–36. https://doi.org/10.1016/j.esp.2021.07.001

Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). *Electra: Pre-training text encoders as discriminators rather than generators*. arXiv:2003.10555. https://doi.org/10.48550/arXiv.2003.10555

Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. London: Palgrave Macmillan. https://doi.org/10.1057/9781137333377

Cotos, E., & Huffman, S. (2013). Learner fit in scaling up automated writing evaluation. *International Journal of Computer-Assisted Language Learning and Teaching*, *3*(3), 77–98. https://doi.org/10.4018/ijcallt.2013070105

Cotos, E., & Pendar, N. (2016). Discourse classification into rhetorical functions for AWE feedback. *CALICO Journal*, *33*(1), 92–116. https://doi.org/10.1558/cj.v33i1.27047

Cotos, E., Link, S., & Huffman, S. (2016). Studying disciplinary corpora to teach the craft of Discussion. *Writing and Pedagogy*, *8*(1), 33–64. https://doi.org/10.1558/wap.v8i1.27661

Devitt, A. J. (2004). *Writing genres*. Carbondale: Southern Illinois University Press.

Dizon, G., & Gayed, J. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *JALT CALL Journal*, *17*(2), 74–9. https://doi.org/10.29140/jaltcall.v17n2.336

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge: Cambridge University Press.

Feng, H.-H., & Chukharev-Hudilainen, E. (2022). Genre-based AWE system for engineering graduate writing: Development and evaluation. *Language Learning & Technology*, *26*(2), 58–77. https://doi.org/10125/73479

Fiacco, J., Cotos, E., & Rose, C. (2019). Towards enabling feedback on rhetorical structure with neural sequence models. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 310–319). New York: ACM Press. https://doi.org/10.1145/3303772.3303808

Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, *48*(1), 99–116. https://doi.org/10.1017/S0261444813000037

Freedman, A. (1993). Show and tell? The role of explicit teaching in the learning of new genres. *Research in the Teaching of English*, *27*(3), 222–251.

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, *3*, 100055. https://doi.org/10.1016/j.caeai.2022.100055

Habibie, P., & Starfield, S. (2023). AI-mediated English for research publication purposes: Are we there yet? *Journal of English for Research Publication Purposes*, *4*(1), 1–4. https://doi.org/10.1075/jerpp.00013.hab

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Howard, J., & Ruder, S. (2018). *Universal language model fine-tuning for text classification*. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (vol. 1: Long papers, pp. 328–339). Melbourne: Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-1031

Huang, J. C. (2014). Learning to write for publication in English through genre-based pedagogy: A case in Taiwan. *System*, *45*, 175–186. https://doi.org/10.1016/j.system.2014.05.010

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, *16*(3), 148–164. https://doi.org/10.1016/j.jslw.2007.07.005

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly*, *30*(4), 693–722. https://doi.org/10.2307/3587930

Johns, A. M. (Ed.) (2002). *Genre in the classroom: Multiple perspectives*. Mahwah: Lawrence Erlbaum. https://doi.org/10.4324/9781410604262

Kershaw, D., & Koeling, R. (2020). *Elsevier CC-BY corpus (V3)* [Dataset]. Elsevier Data Repository. https://doi.org/10.17632/zm33cdndxs.3

Knight, S., Shibani, A., Abel, S., Gibson, A., Ryan, P., Sutton, N., Wight, R., Lucas, C., Sándor, A., Kitto, K., Liu, M., Mogarkar, R. V., & Buckingham Shum, S. (2020). AcaWriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research*, *12*(1), 141–186. https://doi.org/10.17239/jowr-2020.12.01.06

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605–634. https://doi.org/10.1080/09588221.2020.1743323

McCarthy, K. S., Roscoe, R. D., Allen, L. K., Likens, A. D., & McNamara, D. S. (2022). Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing*, *52*, 100608. https://doi.org/10.1016/j.asw.2022.100608

Paltridge, B. (2019). Focusing on language in second language writing classrooms: Rethinking the approach. *Journal of Second Language Writing*, *46*, 100680. https://doi.org/10.1016/j.jslw.2019.100680

Park, L. A., & Read, J. (2019). A blended metric for multi-label optimisation and evaluation. In M. Berlingerio, F. Bonchi, T. Gärtner, N. Hurley, & G. Ifrim (Eds.), *Machine learning and knowledge discovery in databases* (pp. 719–734). ECML PKDD 2018. Lecture Notes in Computer Science, vol. 11051. Cham: Springer. https://doi.org/10.1007/978-3-030-10925-7_44

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, *37*(1), 8–25. https://doi.org/10.1080/01443410.2015.1136407

Saeli, H., Rahmati, P., & Koltovskaia, S. (2023). Learner engagement with written corrective feedback: The case of automated writing evaluation. *Journal of Response to Writing*, *9*(2), 1–39. https://doi.org/10.32038/frsl.2023.01.08

Singh, M. K. M. (2014). Challenges in academic reading and overcoming strategies in taught master programmes: A case study of international graduate students in Malaysia. *Higher Education Studies*, *4*(4), 76–88. https://doi.org/10.5539/hes.v4n4p76

Swales, J. M. (1981). *Aspects of article introductions*. Birmingham: University of Aston.

Swales, J. M. (1990). *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524827

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills* (3rd ed.). Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.2173936

Tardy, C. M. (2006). Researching first and second language genre learning: A comparative review and a look ahead. *Journal of Second Language Writing*, *15*(2), 79–101. https://doi.org/10.1016/j.jslw.2006.04.003

Tardy, C. M. (2009). *Building genre knowledge*. West Lafayette: Parlor Press.

Tardy, C. M. (2017). The challenge of genre in the academic writing classroom: Implications for L2 writing teacher education. In J. Bitchener, N. Storch, & R. Wette (Eds.), *Teaching writing for academic purposes to multilingual students: Instructional approaches* (pp. 69–83). New York: Routledge. https://doi.org/10.4324/9781315269665-5

Tardy, C. M., Sommer-Farias, B., & Gevers, J. (2020). Teaching and researching genre knowledge: Toward an enhanced theoretical framework. *Written Communication*, *37*(3), 287–321. https://doi.org/10.1177/0741088320916554

Tardy, C. M., Caplan, N. A., & Johns, A. M. (2023). *Genre explained: Frequently asked questions and answers about genre-based instruction*. Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.11714330

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, *46*(2), 327–369. https://doi.org/10.1111/j.1467-1770.1996.tb01238.x

Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing*, *8*(2), 111–122. https://doi.org/10.1016/S1060-3743(99)80124-6

Williams, J. M., & Colomb, G. G. (1993). The case for explicit teaching: Why what you don't know won't help you. *Research in the Teaching of English*, *27*(3), 252–264. https://www.jstor.org/stable/40171226

Yasuda, S. (2011). Genre-based tasks in foreign language writing: Developing writers' genre awareness, linguistic knowledge, and writing competence. *Journal of Second Language Writing*, *20*(2), 111–133. https://doi.org/10.1016/j.jslw.2011.03.001

Zhai, N., & Ma, X. (2022). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*, *61*(4), 875–900. https://doi.org/10.1177/07356331221127300

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*, 100439. https://doi.org/10.1016/j.asw.2019.100439

Appendix: Introduction Move/Step Coding Scheme

Move	Step	Simplified step description	Example linguistic patterns
Establish territory	Claim centrality	• To affirm that the research topic is central in the field by highlighting the prominence, importance, and interest in the topic	The increasing interest in ...; ... play a key/an important role in ...; ... are essential
	Provide background	• To overview the targeted knowledge space and empirical and/or theoretical background to the study by presenting generally known information on the topic or by referring to and/or synthesizing previous research	It is logical to accept that ...; There are models that have been developed to ...; Research shows ...
Identify niche	Problematize research	• To evidence a gap in the targeted research or domain of practice that needs to be filled and/or conditions/difficulties that require attention • To raise general questions based on the existing body of knowledge and/or based on the identified gap or problem • To put forth general hypotheses about possible future findings or implications based on the existing body of knowledge and/or the specified gap, problem, and/or questions	However, ... no work has been reported on ...; Very few studies investigated ...; ... does not reflect ...; ... appears to be limited by ...; ... raise the question of how ...; Given the ..., why ...?; ... is expected that ...;
	Present justification	• To emphasize and justify the need to address the specified gap, problem, questions, and/or hypotheses that constitute the niche	... is hence needed ...; Therefore, ... is needed to ...; ... are necessary ...; It is important to ...;

Move description	Step	Simplified step	Example linguistic patterns
Address niche	Announce purpose	• To declare the main purpose/s of the study	The aim of the present paper is to ...; The main purpose was to ...
	Highlight study specifics	• To introduce main features of the study (e.g., method/approach, actions taken, strategy chosen, or principal results of the study) • To present the research questions and/or hypotheses about findings relevant to the research objectives/ questions of the study	This research will focus on ...; A comparison of ... is presented ...; ... will illustrate how ...; In the present study ... was investigated.
	State contribution	• To articulate the value of the current study	The results of this basic investigation can help to ...; This paper extends and deepens ...
	Outline the study	• To preview the structure of paper and/or content	The remainder of this research is divided into five sections.; Section 1 describes ...; Section 2 provides ...

[a] Step descriptions are adapted from Cotos et al. (2016), who further the move/step constructs established by Swales (1990).