A Resource on Dialogical Moves in Native and Non-Native academic writers of English

Giulia D'Agostino, Narjes Sheikh Asadi, Elena Musi
Università della Svizzera italiana, Università della Svizzera italiana, University of Liverpool
dagosgi@usi.ch, sheikn@usi.ch, elena.musi@liverpool.co.uk

Abstract

This paper provides a new approach to the study of linguistic differences in research articles written by native and non-native English writers, including a novel linguistic resource. Conceptually, we propose a functional definition of academic nativeness. Empirically, we operationalize this definition through a survey and the development of a reliable automatic method to distinguish academic natives from non-natives. We then release a corpus of 80 research articles in the field of Linguistics, with introductions manually and reliably annotated for dialogical moves. Preliminary experiments indicate that automatic annotation using large language models remains challenging. Furthermore, the linguistic features that differentiate academic natives and non-natives diverge from those typically associated with general native/non-native English distinctions. These findings aim to inform and enhance academic writing pedagogy, while also offering insights relevant to broader language and corpus studies, as well as computational research.

Keywords: academic nativeness, dialogic moves, genre analysis, annotation

1. Introduction

In Applied Linguistics, evaluating the linguistic quality of research articles is a key step toward improving academic writing instruction. One effective method involves analyzing dialogic moves – functional, communicative units that help writers organize their research – in article introductions, which are shaped by genre, discipline, and the writer's English proficiency. Notably, high-level proficiency can be achieved by both native and non-native speakers, prompting a reconsideration of what constitutes "academic nativeness." This study defines academic nativeness broadly, including individuals for whom English serves as a first language, a heritage language, a lingua franca, or the main language of academic engagement, as opposed to those with limited exposure (EFLs). Since such distinctions cannot be assumed, the study first uses self-assessment surveys to gauge authors' perceptions of their academic writing. An automatic classification method is then developed and tested to identify academic native vs. non-native authors. We then examine the distribution of dialogic moves and surface linguistic features (e.g., linguistic density) across this native – non-native spectrum. The main contributions of the study are the following:

- It provides a novel, empirically verified definition of English nativeness for academic purposes.
- It delivers a fully replicable method for academic native/non-native classification on the basis of empirical data.
- it releases a newly compiled corpus of 80 research articles in Language/Linguistics, classified for academic (non-)nativeness and systematically annotated for dialogical moves in the Introduction sections.
- It shows that automatic annotation of dialogical moves using large language models is still a too complex task.
- It shows that the discriminatory features between native vs non-native in the academic setting differ from those associated with native vs non-native English speakers.

While previous studies have compared dialogic moves between native and non-native English writers (Arsyad and Adila, 2018b; Arvay and Tankó, 2004; Bhatia, 2014; Çeşme and Yağız, 2020a), they have typically focused on non-native data from a single country. The present study is the first to provide broader cross-national comparisons across writers from different countries and journals of varying rankings (Quartiles in Scopus). These analyses are vital because rhetorical differences can impact publication success. Insights from this study can inform pedagogy to strengthen novice writers' academic competence and their understanding of expert strategies in leading journals.

2. Background

2.1. (Non-)Nativeness Classification: State of the Art

What counts as Academic English Nativeness has been understudied and lacks a unified definition. In the field of Language Acquisition, there is a well known distinction between English as a Foreign Language (EFL) — typically monolingual, with learners residing in their home countries where English is neither the first nor the official language —

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 6332–6339
11–16 May 2026.

and English as Second Language (ESL) — which may be bilingual or multilingual, with learners immersed in the culture of the target language. Although these two groups are regarded as the L2 or non-native English speakers, their contexts of acquisition differ significantly: ESL are those learners who engage with English in authentic, real-life settings beyond the classroom, while EFL learners encounter far fewer opportunities for everyday use. (Krashen, 1985; Longcope, 2009; Iwai, 2011; Si, 2019). While EFL and ESL contexts have traditionally been treated as distinct research domains due to EFLs' more limited opportunities to engage with the English language (Götz and Schilk, 2011; Gries and Deshors, 2015; Zhang and Kang, 2022), some studies suggest converging patterns among native and non-native advanced scholars (Zhao, 2017). We follow the latter approach: if we focus on the scholarly context, EFLs learners might also reach quasi-native skills depending on a variety of factors, including English being the first language used to learn a subject since the outset (e.g., Computer Science).

The main attempt to operationalise the distinction between native and non-native scholars in academic publishing is by Wood (2001). He proposed two classification criteria: a strict criterion based on an author's name and institutional affiliation, and a broad criterion based solely on the author's name. Peacock (2002) applied the strict criterion to analyze rhetorical and linguistic patterns in 252 discussion sections of research articles across seven disciplines, focusing on differences between native and non-native English scholars. However, the reliance on names as proxies for linguistic identity introduces significant ambiguities. For instance, a U.S.-educated scholar named Mohammad might be misclassified as non-native despite being raised and academically socialized in an English-speaking environment. Additionally, naming practices in multilingual and multicultural contexts, such as the Netherlands or Switzerland, where ancestral or spousal surnames are common, could disrupt Wood's model. These examples highlight that such frameworks can oversimplify the complex realities of scholarly multilingualism and global academic participation.

2.2. The CaRS Model and Dialogical Moves for Academic Literature

According to genre analysis, writing the introduction of a research article is one of the most challenging tasks for scholars because it must justify the study's purpose, highlight its significance, and engage readers. As one of the most rhetorically complex sections, a poorly written introduction can undermine the paper's persuasive impact and discourage readers from continuing. To address these challenges, genre-based approaches—particularly Swales (1990)'s Create a Research Space (CaRS) model—have become central to teaching and analyzing research article introductions (RAIs). Swales (1990, 2004) conceptualizes introductions as carefully crafted rhetorical structures designed to situate research within its field and demonstrate its relevance. The CaRS model outlines a three-move structure:

- Establish the research territory by introducing the topic, reviewing key literature, and emphasizing its importance.
- Identify a gap or problem in the existing research to justify the need for the study.
- Occupy the niche by outlining the study's aims, methods, and contributions.

3. Study: Automatic Classification of Academic Nativeness in English

3.1. Survey: the Native/Non-native Ground Truth

To collect the self-assessed evaluation of nativeness among academics in the linguistics domain, a Qualtrics survey was developed and distributed to the corresponding authors and co-authors of the 80 selected articles that constitute our corpus (see Section 4). The survey was distributed to 114 authors through 278 personalized emails, including follow-up reminders. It comprised five questions, the first addressing the academic nativeness self-assessment and the others relating to ancillary behaviors beyond the scope of this study (e.g., identifying the sections of an article perceived as most difficult to write). In the first question, operational definitions of L1, ESL and EFL were provided to guarantee consistency. The definitions are as follows:

L1 English is my mother tongue

ESL I speak English as a second language, which includes those who primarily use English for daily communication and may have grown up or studied in an English-speaking country.

EFL I speak English as a foreign language, mainly for scientific communication, including publications and conferences.

Since L1 and ESL are functionally close categories for academic purposes, we conflated L1 and ESL into the "native" category, as opposed to EFL which constitute the "non-native" category. It must be reminded that our notion of academic nativeness is limited to the professional practice. The survey yielded 52 valid responses. Based on these responses, a new corpus was compiled according to the following criteria: (a) only single-authored articles were included to allow for direct attribution of linguistic practices to individual authors, thereby avoiding the confounding effects of multi-authorship; and (b) all articles had to be written in English and be openly accessible. Based on these criteria, the final corpus consisted of 50 research articles written by both native and non-native English-speaking scholars.

3.2. Experiment

The survey results serve as ground truth for fine-tuning and testing two language models in two alternative configurations. The goal is the automatic binary classification of textual units into "native" and "non-native" classes to reliably distinguish papers written by native and non-native users of academic English. The output then provides a basis for further analysis of these papers' linguistic features.

3.2.1. Experimental Setting

For classification purposes, we compared two encoder-only models from the BERT family for two main reasons. First, the limited number of data points and the noise in the training data due to the lack of an agreed distinction between native vs non native increase the risks of overfitting in complex language models. Second, encoder-decoder or decoder-only architectures are overly complex for a simple two-label (binary) classification task, making them less suitable than simpler encoder-only models. Both models are fine-tuned as follows:

Vanilla fine-tuning Class imbalance is counterbalanced by computing the appropriate label weight. Fine-tuning is performed in 5-fold cross validation over entropy loss. The deeper layers are kept frozen, fine-tuning only the last 4 ones for a maximum of 10 epochs, with early stopping based on F_1 validation performance. Optimization is applied through AdamW algorithm.

RAG fine-tuning Same as above, but some input is augmented with the three most similar sequences from the same fold and with a check on retrieval within the same class. Retrieval-augmentation is performed through embedding similarity, with a dropout of 0.3 (that is, 30% of inputs are left plain instead).

Hyperparameters and further technical details about fine-tuning are provided in Appendix A.

3.2.2. Data Preparation

The label for each of the 50 documents employed in this experiment is attributed on the basis of the survey response: "native" for both self-assessed L1 and ESL writers, "non-native" for EFL writers.

For each document, the following sections were manually extracted: Abstract, Introduction, Discussion (if present), Conclusion(s) (if present). Clearly, alternative denominations were considered equivalent and the corresponding text included in the dataset if the section fulfilled one of the functions above. As expected, the Discussion session was typically absent in non-experimental contributions.

The contents were later segmented into sentences and reassembled into sequences not longer than 128 tokens – avoiding truncation and padding accordingly at tensor conversion stage.

3.2.3. Results

Results for this experimental phase are reported in Table 1. RoBERTa consistently outperforms DeBERTa, presumably due to architectural and pre-training characteristics of the model. Additionally, although supplementing fine-tuning with a RAG method sensibly improves performance for RoBERTa, it lowers all metrics for DeBERTa.

The best-performing method for native/non-native classification is applied to our dataset to reliably determine the nativeness of authors – and include this information among the relevant metadata.

Model	Acc	Macro-P	Macro-R	Macro-F1
RoBERTa	0.90	0.89	0.90	0.89
RAG+RoBERTa	**0.95**	**0.94**	**0.94**	**0.94**
DeBERTa	0.87	0.84	0.84	0.84
RAG+DeBERTa	0.84	0.81	0.83	0.82

Table 1: Comparison of models and approaches on the nativeness classification task (best results in bold).

4. Dataset Description

This corpus encompasses 80 research articles (RAs). Following Arianto and Basthomi (2021) and drawing on criteria of representativeness, reputation, and accessibility (Amnuai, 2019), we selected 16 peer-reviewed journals in the domain of Language and Linguistics. Journals were sampled from the SCImago Journal Rank (SJR, 2025), with four chosen from each quartile (Q1–Q4) based on 2021 rankings. All were Scopus-indexed, publicly accessible, and verified for quality via SJR citation metrics. Then, from the selected journals, 80 RAs (20 per quartile) have been purposively sampled (Amini Farsani and Barzegar, 2025; Suri, 2011) according to five criteria. Articles were selected based on disciplinary relevance to Applied Linguistics, accessibility, availability of author contact information, geographical diversity, and empirical design. From the eligible pool, 80 selected RAs representing all four ranking quartiles in Linguistics were included in the final corpus. The documents are classified as being written by native or non-native academic writers of English by the method described in Section 3.2, resulting in 64 papers by non-natives and 16 by natives. The dataset is imbalanced, but this likely reflects a real difference in the number of papers published by native and non-native English-speaking researchers in this field.

4.1. Annotation and Validation

All the introduction sections of the 80 research papers were extracted and uploaded as txt files into UAM Corpus Tool 3.3x (O'Donnnell, 2009). The corpus was manually annotated following Swales (1990)' framework, guided by an annotation manual (Sheikh Asadi, 2025). For each move, different segments are identified. The length of a segment could range from a single sentence to an entire paragraph, depending on its communicative function (Swales, 2004, 228-229). Table 2 presents the distribution of annotated moves in the corpus according to nativeness.

	Move 1	Move 2	Move 3
native	53	17	38
weighted native	28.17	50.83	48.47
non-native	330	93	175
weighted non-native	155.16	381.10	278.16

Table 2: CaRS annotated moves, distributed across native and non-native writers.

The results indicate no statistically significant differences between native and non-native authors. The same applies at the step level. This outcome is likely attributable to the formulaic nature of dialogical moves, as defined by the CaRS model. Their structured form makes them relatively easy to acquire and reproduce, even for non-native speakers, provided they are experienced academics.

Additionally, we examined a set of basic linguistic features: besides *average move length* (number of words), and *average sentence length* (number of words), we have focused on four aspects that have been identified as significant markers in previous L1/L2 studies in the academic domain. These are: *lexical density* (mean segmental type-token ratio; proportion of content words, where content words are identified as categories "NOUN", "VERB", "ADJ", and "ADV" through the SpaCy library), *lexical diversity* (proportion of unique word types), which have turned out to be higher in dissertation abstracts of L1 writers (Nasseri and Thompson, 2021); *connectives density* (proportion of "ADP", "CCONJ", and "SCONJ" categories identified through the SpaCy library) that generally tend to be higher in non native writers (Zufferey and Degand, 2024); *passive voice ratio* (proportion of passive voice verbs, identified through the PassivePy library), that have deemed to be higher in L1 authors of research articles (Parkinson and Woods, 2025). Interestingly, none of these features beat a significant variation (see Table 3).

measure	native	non-native
lexical density	0.54	0.54
lexical diversity	0.82	0.79
connectives density	0.20	0.20
passive voice ratio	0.32	0.32
avg move length (word #)	69.2	70.3
avg sentence length (word #)	32.4	27.7

Table 3: Linguistic measures across the corpus, sorted by nativeness.

This suggests that classification algorithm developed in this study must rely on other, more complex discriminative linguistic features. Among these, collocations and lexical bundles may play a role: the overlap between the 15 most frequent bigrams and trigrams used by native and non-native speakers is as low as 0.13, calling for an investigation at the construction rather than vocabulary level. Previous studies have for example shown that L2 writers in Telecommunications research journals prefer clausal bundles to phrasal bundles and use them functionally differently from L1 writers (Pan et al., 2016). We leave this further investigation to future work since a larger corpus is required to surface meaningful differences at the level of lexical bundles.

CaRS moves annotation was also carried out by a second trained individual on 20% of the dataset (16 documents). Inter-annotator agreement (IAA) was assessed using the Γ coefficient (Mathet et al., 2015)[1], which evaluates both segmentation and labeling consistency. Results are sorted out by CaRS moves and displayed in Table 4.

task	Γ (human-human)	Γ (human-LLM)
Move 1	0.86	0.30
Move 2	1.00	0.75
Move 3	0.82	0.55
all moves	0.77	0.29

Table 4: Γ Inter Annotator Agreement (pairwise, span length and labeling combined) for CaRS model moves

4.2. LLMs as Annotators

To assess the feasibility of automatically annotating moves according to the CaRS model, we compared the gold-standard annotation with GPT-5's

[1] Python library `pygamma-agreement`

performance.[2] The results, summarized in Table 4, indicate only fair overall reliability. When separating segmentation and labeling tasks, segmentation disagreement remains substantial ($\Gamma = 0.16$) but labeling disagreement emerges as the primary source of error ($\Gamma = -0.32$), indicating agreement worse than chance. In other words, the LLM systematically assigns different move labels even when segmenting similarly. This finding highlights that, while humans can reach moderate agreement, the definition of dialogical moves remains too complex for current LLMs to operationalize effectively.

5. Conclusion

This paper introduces a new approach to analyzing linguistic differences in research articles on the basis of the *(non-)nativeness* of writers, along with a novel manually annotated corpus. After having provided a definition of academic nativeness, we develop a reliable method for its automatic classification, based on a self-assessed ground truth. We classify a corpus of 80 research articles in the field of Linguistics according to the “native"-“non-native" distinction. The introductions are fully manually annotated for dialogic moves (CaRS moves model). Finally, we observe that CaRS moves and linguistic measures generally associated with “native"-“non-native" distinction are not discriminatory in the academic setting, calling for deeper linguistic analysis.

6. Limitations

While the results of this study have direct implications for English language teaching — especially for academic purposes — we acknowledge that the study is intentionally limited in scope. It focuses solely on research articles from the field of Linguistics. This narrow focus helps to control for external variables, but it also limits the generalizability of the findings. At the same time, the Linguistic field is a sensible choice for a preliminary study because its writing style is less formulaic than other academic domains (e.g., Computer Science) and it is typically not taught in English since the outset.

7. Supplementary Materials Availability Statement

The annotated corpus (CaRS moves and steps), alongside the annotation manual, and the dataset used for the experiment (native/non-native classification), are available at the following repository: https://github.com/dagosgi/LREC2026-academic_nativeness.

[2]The model was reached through API and prompted including the full annotation manual (Sheikh Asadi, 2025) but without additional in-context examples.

8. Bibliographical References

Mohammad Amini Farsani and Reihaneh Barzegar. 2025. Tracing sampling in applied linguistics: the case of quantitative research approach. *Quality & Quantity*, 59(1):111–123.

Wirada Amnuai. 2019. The textual organization of the discussion sections of accounting research articles. *Kasetsart Journal of Social Sciences*, 40(2):389–394. Published online 2017 (acceptance/online date) but in 2019 issue.

Muhammad Aqeel, Sajid Waqar, Samrina Abdul Majeed, and Samia Kiran. 2023. Role of English Language as a Global and Economical Language Around the Globe: A Case of Kachru's Model. *Bulletin of Business and Economics (BBE)*, 12(4):141–147.

M. A. Arianto and Y. Basthomi. 2021. The authors' research gap strategies in elt research article introductions: Does scopus journal quartile matter? *Journal of Language and Linguistic Studies*, 17(4):1743–1759.

S. Arsyad and D. Adila. 2018a. Using local style when writing in english: The citing behaviour of indonesian authors in english research article introductions. *Asian Englishes*, 20(2):170–185.

Safnil Arsyad and Destiantari Adila. 2018b. Using local style when writing in English: the citing behaviour of Indonesian authors in English research article introductions. *Asian Englishes*, 20(2):170–185.

A. Arvay and G. Tankó. 2004. A contrastive analysis of english and hungarian theoretical research article introductions. *IRAL: International Review of Applied Linguistics in Language Teaching*, 42(1).

Margie Berns. 2019. World Englishes and Communicative Competence. In Cecil L. Nelson, Zoya G. Proshina, and Daniel R. Davis, editors, *The Handbook of World Englishes*, 1 edition, pages 674–685. Wiley.

Vijay K. Bhatia. 1993. *Analysing genre: language use in professional settings*. Applied linguistics and language study. Longman, London ; New York. OCLC: 26401230.

Vijay K. Bhatia. 2014. *Analysing genre: Language use in professional settings*. Routledge.

Kingsley Bolton. 2019. Braj B. Kachru and Asian Englishes. *World Englishes*, 38(1-2):67–77.

Steve Disney. 2010. The World English Model revised: Definite article use in ICE-GB and ICE-HK. In *Papers from LAEL PG 2009*, volume 4, pages 1–23.

María Pilar Mur Dueñas. 2010. A Contrastive Analysis of Research Article Introductions in English and Spanish. *Revista Canaria de Estudios Ingleses*, (61):119–133.

Anna Duszak and Jo Lewkowicz. 2008. Publishing academic texts in English: A Polish perspective. *Journal of English for Academic Purposes*, 7(2):108–120.

Maryam Farnia and Safoora Barati. 2017. Writing introduction sections of research articles in applied linguistics: cross-linguistic study of native and non-native writers. *Indonesian Journal of Applied Linguistics*, 7(2):248.

Ruth Eliana Franssisca and Adaninggar Septi Subekti. 2022. Indonesian high school students' attitudes towards varieties of English: A survey study. *Studies in English Language and Education*, 9(1):318–330.

Zosia Golebiowski. 1999. Application of Swales' model in research papers by non-native authors. *International Review of Applied Linguistics*, 37(3):231–247.

Sandra Götz and Marco Schilk. 2011. Formulaic sequences in spoken enl, esl and efl: Focus on british english, indian english and learner english of advanced german learners. In *Exploring Second-Language Varieties of English and Learner Englishes*, pages 79–100. John Benjamins Publishing Company.

Stefan Th Gries and Sandra C Deshors. 2015. Efl and/vs. esl?: A multi-level regression modeling perspective on bridging the paradigm gap. *International Journal of Learner Corpus Research*, 1(1):130–159.

Eliana Hirano. 2009. Research Article Introductions in English for Specific Purposes: A Comparison between Brazilian Portuguese and English. *English for Specific Purposes*, 28(4):240–250. ERIC Number: EJ856608.

A. Holliday. 2006. Native-speakerism. *ELT Journal*, 60(4):385–387.

Marianne Hundt and Joybrato Mukherjee. 2011. Introduction: Bridging a paradigm gap. In *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, pages 1–6. John Benjamins Publishing Company.

Ken Hyland. 2016. Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, 31:58–69.

Yuko Iwai. 2011. The Effects of Metacognitive Reading Strategies: Pedagogical Implications for EFL/ESL Teachers. *Reading Matrix: An International Online Journal*, 11(2):150–159. ERIC Number: EJ955192.

Alireza Jalilifar, A. Majid Hayati, and Amir Mashhadi. 2012. Evaluative Strategies in Iranian and International Research Article Introductions: Assessment of Academic Writing. *Journal of Research in Applied Linguistics*, 3(1):81–109.

Braj B. Kachru. 1990. World Englishes and applied linguistics. *World Englishes*, 9(1):3–20.

Stephen D. Krashen. 1985. *Second language acquisition and second language learning*, reprinted edition. Language teaching methodology series. Pergamon Press, Oxford.

Sun Lee. 2001. *A Contrastive Rhetoric Study of Korean and English Research Paper Introductions*. PhD dissertation, University of Illinois at Urbana-Champaign.

Chek Kim Loi. 2010. Research article introductions in Chinese and English: A comparative genre-based study. *Journal of English for Academic Purposes*, 9(4):267–279.

Chek Kim Loi and Moyra Sweetnam Evans. 2010. Cultural differences in the organization of research article introductions from the field of educational psychology: English and Chinese. *Journal of Pragmatics*, 42(10):2814–2825.

Peter Longcope. 2009. Differences between the EFL and the ESL Language Learning Contexts.

Pedro Martín Martín. 2003. Personal Attribution in English and Spanish Scientific Texts. *Bells: Barcelona English language and literature studies*.

Yann Mathet, Antoine Widlöcher, and Jean-Philippe Métivier. 2015. The Unified and Holistic Method Gamma (Γ) for Inter-Annotator Agreement Measure and Alignment. *Computational Linguistics*, 41(3):437–479.

Sandra McKay and James Dean Brown. 2016. *Teaching and assessing EIL in local contexts around the world*, first published edition. ESL & applied linguistics professional series. Routledge, New York, NY London.

P. Medgyes. 1992. Native or non-native: who's worth more? *ELT Journal*, 46(4):340–349.

Rajend Mesthrie and Rakesh M. Bhatt. 2008. *World Englishes: The Study of New Linguistic Varieties*, 1 edition. Cambridge University Press.

Ana I. Moreno. 2008. The importance of comparable corpora in cross-cultural studies. In Ulla Connor, Ed Nagelhout, and William Rozycki, editors, *Pragmatics & Beyond New Series*, volume 169, pages 25–41. John Benjamins Publishing Company, Amsterdam.

Ana I. Moreno and John M. Swales. 2018. Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, 50:40–63.

Maryam Nasseri and Paul Thompson. 2021. Lexical density and diversity in dissertation abstracts: Revisiting english l1 vs. l2 text differences. *Assessing Writing*, 47:100511.

Michael O'Donnnell. 2009. The UAM CorpusTool : software for corpus annotation and exploration. In *Applied Linguistics Now: Understanding Language and Mind*, pages 1433–1447. Universidad de Almería, Almería.

Fan Pan, Randi Reppen, and Douglas Biber. 2016. Comparing patterns of l1 versus l2 english academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes*, 21:60–71.

Jean Parkinson and Lisa Woods. 2025. The use of the passive voice in science students' writing. *Journal of English for Academic Purposes*, 77:101561.

Matthew Peacock. 2002. Communicative moves in the discussion section of research articles. *System*, 30(4):479–497.

Masoud Khalili Sabet and Maryam Kazempouri. 2015. Generic Structure of Discussion Sections in ESP Research Articles across International and Iranian Journals. *Advances in Language and Literary Studies*, 6(2):87–95.

John Robert Schmitz. 2014. Looking under Kachru's (1982, 1985) three circles model of World Englishes: the hidden reality and current challenges. *Revista Brasileira de Linguística Aplicada*, 14(2):373–411.

Narjes Sheikh Asadi. 2025. Swales' CARS model (1990) : annotation manual. Technical report.

Peng Si. 2019. A Study of the Differences between EFL and ESL for English Classroom Teaching in China. *IRA International Journal of Education and Multidisciplinary Studies*, 15(1):32.

Kamal K Sridhar and Shikaripur N Sridhar. 1986. Bridging the paradigm gap: Second language acquisition theory and indigenized varieties of english. *World Englishes*, 5(1):3–14.

Harsh Suri. 2011. Purposeful Sampling in Qualitative Research Synthesis. *Qualitative Research Journal*, 11(2):63–75.

John Swales. 1990. *Genre analysis: English in academic and research settings*. Cambridge applied linguistics series. Cambridge University Press, Cambridge.

John M. Swales. 2004. *Research Genres: Explorations and Applications*, 1 edition. Cambridge University Press.

Tamara Valentine. 2019. The functional and pragmatic contexts of world Englishes. *World Englishes*, 38(1-2):269–279.

Alistair Wood. 2001. International scientific English: The language of research scientists around the world. In John Flowerdew and Matthew Peacock, editors, *Research Perspectives on English for Academic Purposes*, Cambridge Applied Linguistics, pages 71–83. Cambridge University Press, Cambridge.

Chao Zhang and Shumin Kang. 2022. A comparative study on lexical and syntactic features of esl versus efl learners' writing. *Frontiers in Psychology*, 13:1002090.

Jun Zhao. 2017. Native speaker advantage in academic writing? conjunctive realizations in eap writing by four groups of writers. *Ampersand*, 4:47–57.

Sandrine Zufferey and Liesbeth Degand. 2024. *Connectives and discourse relations*. Cambridge University Press.

Anett Árvay and Gyula Tankó. 2004. A contrastive analysis of English and Hungarian theoretical research article introductions. *IRAL - International Review of Applied Linguistics in Language Teaching*, 42(1).

Hatice Çeşme and Oktay Yağız. 2020a. A Contrastive Rhetorical Analysis of Anglophone and Turkish Authors' RA Introductions in ELT. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 24(2):899–915.

Hatice Çeşme and Oktay Yağız. 2020b. A contrastive rhetorical analysis of anglophone and turkish authors' ra introductions in elt. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 24(2):899–915.

A. Fine-tuning Details

We present additional details regarding the usage of pretrained models. We present an overview of the initial model checkpoints and their parameter counts in Table 5. The hyperparameters to train the models on the vanilla fine-tuning task are given in Table 6, and the additional ones for RAG enhanced fine-tuning are given in Table 7. Training a single model generally takes up to 15 minutes at most on a A100 GPU.

Model	**Checkpoint**	**Size**
RoBERTa (base)	`FacebookAI/roberta-base`	125M
DeBERTa v3 (base)	`microsoft/deberta-v3-base`	86M

Table 5: Description of each model and the specific checkpoint we used.

For the sequence labeling models, we train on the training set while observing metrics on cross-validation. We pick the model iteration with the highest F_1 score and evaluate that model on the test set (33% of samples) to obtain the results reported in Table 1. We use the same split for each experiment.

Model	**Parameter**	**Value**
all	learning rate	1e-05
all	splits	5
all	batch size	16
all	max sequence length	128
all	max epochs	10
all	patience	2
all	weight decay	0.01
all	trainable layers	4

Table 6: Hyperparameters for the simple fine-tuning approach.

Model	**Parameter**	**Value**
all	embedder	`all-MiniLM-L6-v2`
all	similar samples	3
all	retrieval dropout	0.3

Table 7: Additional parameters for RAG.