UC Berkeley
Recent Work

Title

Cross-Linguistic Influence in English Argumentative Writing: A Comparative Analysis of Mandarin Chinese and Japanese L1 Learners Using ICNALE

Permalink

https://escholarship.org/uc/item/5p55k9zx

Authors

Sun, Yinwei
Lin, Iyu
Pan, Haoming

Publication Date

2026

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Cross-Linguistic Influence in English Argumentative Writing: A Comparative Analysis of Mandarin Chinese and Japanese L1 Learners Using ICNALE

Iyu Lin
UC Berkeley
iylin@berkeley.edu

Haoming Pan
UC Berkeley
haoming.p@berkeley.edu

Yinwei Sun
UC Berkeley
yinwei.sun@berkeley.edu

Abstract

This study examines how first language (L1) background shapes English argumentative writing among intermediate second language (L2) learners. Focusing on Mandarin Chinese and Japanese learners at the B1 proficiency level, we analyze essays from the ICNALE corpus written under controlled conditions. We compare the two groups across lexical diversity, syntactic complexity, discourse cohesion, and argumentation style. Our results show that Mandarin Chinese learners exhibit higher lexical diversity and, contrary to prior expectations, higher syntactic complexity, while Japanese learners rely more heavily on explicit discourse markers, particularly causal markers, and first-person singular constructions when expressing opinions. By combining interpretable linguistic features with supervised discourse classification, this study demonstrates that L1 influence persists beyond sentence-level grammar and extends to how learners structure reasoning and argumentation. These findings highlight the importance of cross-linguistically informed approaches to L2 writing feedback.

1 Introduction

Learning to write a clear and organized English argumentative essay requires not only vocabulary and grammar knowledge, but also the ability to structure reasoning and signal relationships between ideas. For learners who did not grow up in an English-speaking environment, these aspects of writing may be influenced by rhetorical and discourse habits developed in their first language (L1). Understanding how such L1 influences manifest in English writing is important for both theory and pedagogy, as it helps explain why learners from different linguistic backgrounds produce distinct textual patterns and supports the development of more responsive L2 writing feedback (Jarvis and Odlin, 2000; Connor, 1996; Kubota and Lehner, 2004).

While much previous research has compared second language (L2) English learners to native speakers, relatively few studies have examined systematic differences across L2 learner groups. Mandarin Chinese and Japanese learners provide a revealing contrast because they share a similar regional and educational context but differ typologically and rhetorically. Prior research suggests that Mandarin Chinese writers tend to rely more on additive sequencing strategies, whereas Japanese writers often employ explicit logical connectives to signal relationships between ideas (Ai and Lu, 2013; **?**). Comparing these two groups under the same writing conditions therefore offers a controlled way to observe how distinct L1 traditions shape English argumentative writing.

In this study, we analyze B1-level English argumentative essays written by Mandarin Chinese and Japanese learners drawn from the ICNALE corpus (Ishikawa, 2023). We examine lexical diversity, syntactic complexity, and discourse cohesion to explore how learners' first languages shape their English writing practices. Specifically, we address the following research questions:

- How do Mandarin Chinese and Japanese learners differ in lexical, syntactic, and discourse features when writing English argumentative essays?
- What do these differences suggest about providing L2 writing feedback that is sensitive to cross-linguistic variation?

2 Related Work

2.1 L1 Transfer in Second Language Writing

Research in second language writing has consistently shown that a learner's first language (L1) can influence L2 writing across lexical, syntactic, and discourse levels. This phenomenon is commonly described as cross-linguistic influence or L1 transfer, whereby linguistic structures and rhetorical pat-

terns developed in the L1 shape writing practices in the L2 (Jarvis and Odlin, 2000). Prior work demonstrates that such influence extends beyond grammatical error patterns to include cohesion strategies, stance expression, and the organization of argumentative structure (Connor, 1996; Kubota and Lehner, 2004).

Moreover, the degree and nature of transfer vary across learner populations depending on factors such as educational background and typological distance between the L1 and L2 (Crossley et al., 2016; Wei et al., 2020). These findings motivate studying L2 writers with different L1 backgrounds as distinct groups rather than treating L2 learners as a homogeneous population. Adopting this perspective, the present study compares Mandarin Chinese and Japanese learners of English to investigate how L1-specific rhetorical and linguistic features manifest in English argumentative writing.

2.2 Chinese and Japanese L1 Influence on English Writing

Studies of Mandarin Chinese and Japanese learners of English reveal distinct cross-linguistic influences on writing organization and discourse cohesion.

For Mandarin Chinese L1 learners, prior research has identified a tendency toward paratactic organization, in which ideas are linked sequentially rather than through deeply embedded clause structures. Corpus-based analyses indicate that Chinese learners often produce shorter T-units and fewer subordinate clauses compared to native English writers (Ai and Lu, 2013). In addition, process-oriented studies suggest that Chinese learners' planning strategies during English composition may reflect organizational conventions associated with Chinese academic writing traditions (Wang and Wen, 2002).

In contrast, research on Japanese learners of English shows a strong reliance on explicit connective expressions, particularly in sentence-initial positions, to signal logical relationships between ideas (Nakayama, 2021). While syntactic complexity generally increases with proficiency, the preference for overt connective marking appears to remain a characteristic feature of Japanese EFL writing even at higher levels (Sasaki et al., 2024).

Despite these documented tendencies, direct comparisons between Chinese and Japanese learners using the same writing task and proficiency level remain limited. Focusing on the B1 level provides an ideal testing ground for examining cross-linguistic influence, as learners at this stage possess sufficient grammatical and lexical resources to produce multi-paragraph arguments while still exhibiting observable L1-based rhetorical habits.

2.3 Native Language Identification and Interpretability

Research on Native Language Identification (NLI) has demonstrated that linguistic patterns in learner writing are systematic enough for computational models to distinguish writers from different L1 backgrounds with high accuracy (Goswami et al., 2024). However, much of the NLI literature prioritizes classification performance over interpretability, offering limited insight into why particular linguistic features differ across groups.

In contrast, learner corpus research emphasizes interpretable, feature-level analysis aimed at understanding how specific linguistic choices reflect writing development and cross-linguistic influence (Bestgen and Granger, 2014). The present study follows this interpretive tradition. Rather than attempting to predict learners' L1 backgrounds, it analyzes lexical diversity, syntactic complexity, and discourse cohesion to describe how Mandarin Chinese and Japanese learners structure English argumentative writing at the B1 proficiency level.

3 Data

The data for this study come from the International Corpus Network of Asian Learners of English (ICNALE), a learner corpus that includes written and spoken English production by learners from a range of L1 backgrounds (Ishikawa, 2023). ICNALE provides short argumentative essays produced under controlled conditions: all writers respond to the same prompt, within a fixed time limit, and without access to dictionaries or external resources. This design makes the corpus well suited for cross-linguistic comparison.

For this study, we focus on argumentative essays written by Mandarin Chinese L1 learners and Japanese L1 learners at the B1 proficiency level. Proficiency levels are based on the CEFR scale and assigned within the ICNALE corpus through standardized testing (Ishikawa, 2023). We focus on B1 learners because this stage represents an intermediate point at which learners are able to produce multi-sentence argumentative writing, while L1-based discourse and syntactic habits remain strongly observable.

All essays analyzed respond to the same argumentative prompt, which asks writers to express and justify their opinion on a public policy topic. Restricting the dataset in this way controls for variation in topic, task, and communicative intent, allowing observed differences to be interpreted as differences in writing practices rather than differences in content or genre.

Before feature extraction, all texts were tokenized and sentence-segmented. No grammatical or lexical corrections were made, as retaining original learner output is necessary for examining cross-linguistic influence (Jarvis and Odlin, 2000).

The final dataset consists of:

- Mandarin Chinese L1 learners: 232 essays (B1 level)
- Japanese L1 learners: 179 essays (B1 level)

This dataset provides a sufficient sample from each group to support reliable statistical comparison of lexical, syntactic, and discourse features.

4 Methods

4.1 Feature Selection

This study examines four dimensions of cross-linguistic influence in L2 English argumentative writing: lexical diversity, syntactic complexity, discourse cohesion, and argumentation style. Together, these dimensions capture not only surface-level linguistic variation but also deeper differences in how learners structure and justify arguments (Jarvis and Odlin, 2000; Connor, 1996; Kubota and Lehner, 2004).

4.1.1 Lexical Diversity

Lexical diversity was measured using two complementary approaches. First, we calculated the type–token ratio (TTR), defined as the number of unique word types divided by the total number of word tokens in each essay. While TTR can be sensitive to text length, the ICNALE essays are standardized to approximately 200 words, making TTR appropriate for group comparison (Ishikawa, 2023).

Second, we measured pronoun density, defined as the frequency of pronominal reference per 100 words. Pronouns were identified using spaCy part-of-speech tagging, including personal pronouns (e.g., *I, he, they*), possessive pronouns (e.g., *my, his, their*), and wh-pronouns (e.g., *who, what, whose*). Since pronouns are high-frequency function words that may be repeated frequently, pronoun density helps explain potential mechanisms underlying overall vocabulary range differences between groups(Crossley et al., 2016; Wei et al., 2020).

4.1.2 Syntactic Complexity

Syntactic complexity was evaluated using two commonly used measures: mean clause length and subordination ratio (Lu, 2010). These measures capture whether writers construct arguments through hierarchically embedded clause structures or through shorter, less syntactically complex units.

While prior research has often suggested that Japanese learners exhibit greater syntactic complexity than Chinese learners (Ai and Lu, 2013), this study adopts an exploratory stance and tests for group differences without assuming a fixed directionality.

4.1.3 Discourse Cohesion

To analyze discourse cohesion, we examined the frequency and distribution of discourse markers (DMs), which play a central role in signaling rhetorical relations between clauses and sentences. Following Lin (2023), which adopts a functional discourse-marker framework, discourse markers were grouped into eight functional categories:

- **Causal** (e.g., *so, because*)
- **Contrast** (e.g., *but, however, even though*)
- **Continuation** (e.g., *also, moreover*)
- **Elaboration** (e.g., *such as, for example*)
- **Opening and closing** (e.g., *in my opinion*)
- **Sequencing** (e.g., *first, next*)
- **Topic shifting** (e.g., *anyway, then*)
- **Summary** (e.g., *in conclusion, overall*)

However, simple lexicon-based matching of discourse markers often leads to high false-positive rates due to lexical ambiguity. For example, the word *so* may function as an intensifying adverb ("I am *so* happy") rather than as a causal connective ("*So*, I decided to leave"). To address this limitation, we developed a supervised machine learning pipeline to identify discourse markers that perform a genuine cohesive function.

We manually annotated sentences from both L1 groups according to their discourse function. Sentences containing discourse marker keywords that did not function as cohesive devices were labeled as negative examples. Using this annotated dataset, we trained a Random Forest classifier (Breiman, 2001) with TF–IDF features (Salton and Buckley,

1988) to distinguish functional discourse moves from non-functional uses.

For the analysis, we applied a hybrid strategy. First, the classifier predicted whether a sentence contained a functional discourse marker, filtering out non-cohesive uses of high-frequency lexical items. Sentences predicted as positive were then categorized into specific functional types based on their lexical realization. Group-level comparisons of discourse marker density were conducted using independent-samples t-tests, and effect sizes were reported using Cohen's d to assess the magnitude of group differences.

4.1.4 Argumentation Style

In addition to discourse cohesion, we examined how learners express opinions and stance at the sentence level. Specifically, we analyzed patterns of opinion expression, including the use of first-person singular constructions (e.g., "I think") versus impersonal or generalized constructions (e.g., "it is important to"). These features capture differences in rhetorical stance-taking strategies that may reflect L1-influenced argumentation norms (Kubota and Lehner, 2004; Wei et al., 2020).

4.2 Hypotheses

Based on earlier research and the patterns documented in the Related Work section, we tested the following hypotheses:

- **H1 (Lexical Diversity):** Japanese L1 learners will exhibit lower lexical diversity (TTR) than Chinese L1 learners, due to higher pronoun density reflecting compensation for L1 pro-drop properties (Wei et al., 2020).
- **H2 (Argumentation Style):** Japanese L1 learners will rely more heavily on first-person singular constructions to express personal opinions, while Chinese L1 learners will use fewer explicit first-person stance markers.
- **H3 (Discourse Cohesion):** Japanese L1 learners will use a higher density of explicit discourse markers, particularly causal markers, compared to Mandarin Chinese L1 learners (Nakayama, 2021; Lin, 2023).
- **H4 (Syntactic Complexity):** Mandarin Chinese and Japanese L1 learners will differ in syntactic complexity, as measured by clause length and subordination ratio, without assuming a fixed directionality.

4.3 Analysis Procedure

All texts were processed using the same feature extraction pipeline to ensure comparability across groups. Each essay was first segmented into sentences and tokenized. Lexical diversity was calculated at the document level using TTR. For syntactic complexity, we used automated parsing tools to identify clause boundaries and to compute mean clause length and subordination ratios (Lu, 2010).

For discourse cohesion, we implemented the hybrid approach described above rather than relying on raw keyword counting. Sentences were first filtered by the Random Forest classifier to identify functional discourse moves. Validated markers were then categorized into functional types (e.g., causal, elaboration). We calculated both the overall density of validated markers and the density of specific functional categories per 100 words.

After feature extraction, we compared the Mandarin and Japanese learner groups using independent-samples t-tests. Effect sizes (Cohen's d) were reported alongside significance values, as effect size provides a more interpretable indication of the magnitude of group differences. At this stage, the goal of the analysis was to identify consistent patterns and relative contrasts rather than to build predictive models or make claims about individual learner performance.

5 Analysis

We conducted analyses on lexical diversity and discourse cohesion to evaluate whether the selected feature set is sensitive to systematic differences between Mandarin Chinese and Japanese L1 English learners at the B1 proficiency level. Such feature-based comparative analyses are commonly adopted in learner corpus research to examine cross-linguistic influence beyond surface-level error patterns (Jarvis and Odlin, 2000; Wei et al., 2020). These analyses serve to confirm the feasibility of our approach before proceeding to a more comprehensive analysis.

5.1 Lexical Diversity

Lexical diversity was measured using two complementary approaches: type-token ratio (TTR), which captures overall vocabulary range, and pronoun density, which reflects the frequency of pronominal reference. Lexical diversity measures have been widely used in second language writing research to characterize vocabulary use and devel-

Measure	Chinese mean	Japanese mean	t	p	d
TTR	0.5206	0.4865	7.2599	$< .001$	0.72
Pronoun Density (%)	10.61	12.36	-6.52	$< .001$	0.65

Table 1: Lexical Diversity Measures by L1 Group

opment across learner groups (Lu, 2010; Bestgen and Granger, 2014). Examining both TTR and pronoun density allows us to assess not only differences in lexical variety between groups, but also potential mechanisms underlying those differences.

TTR was calculated as the ratio of unique word types to total word tokens for each essay. Although TTR is known to be sensitive to text length, it remains appropriate for comparison when texts are of similar length, as is the case for ICNALE argumentative essays (Lu, 2010). Pronoun density was computed using spaCy POS tagging to identify personal pronouns (e.g., *I, he, they*), possessive pronouns (e.g., *my, his, their*), and wh-pronouns (e.g., *who, what, whose*), expressed as a percentage of total words. Prior research has shown that pronominal usage can reflect L1-specific reference strategies and compensation effects in L2 writing, particularly for learners whose L1 allows subject omission (Wei et al., 2020). Table 1 reports group-level descriptive statistics and results from two-sample t-tests.

Chinese learners showed significantly higher TTR than Japanese learners (0.5206 vs 0.4865, $p < .001$, $d = 0.72$). This medium-to-large effect indicates that Chinese learners use a wider range of vocabulary, even though both groups are at the same B1 proficiency level. Japanese learners used significantly more pronouns than Chinese learners (12.36% vs 10.61%, $p < .001$, $d = 0.65$), representing a medium effect size.

These two findings support H1 and are mutually reinforcing. Japanese learners' higher pronoun use directly contributes to their lower TTR: repeating high-frequency words like "I," "we," and "they" reduces vocabulary variety. This pattern aligns with prior accounts of cross-linguistic differences in reference strategies and subject realization (Jarvis and Odlin, 2000; Wei et al., 2020). Japanese is a pro-drop language where subjects are routinely omitted when recoverable from context, whereas English requires overt subject marking. Japanese learners of English may therefore overcompensate by explicitly marking subjects with pronouns, a pattern also observed in previous learner corpus studies (Nakayama, 2021).Chinese, while allowing some subject omission, requires explicit subjects more consistently than Japanese, which may reduce pressure for over-marking in English (Wang and Wen, 2002).

Subject	Japanese mean	Chinese mean	t	p	d
I	3.11	0.83	13.68	$< .001$	1.30
We	1.98	1.78	1.00	.319	0.10
it/this/that	1.38	1.53	-1.71	.087	-0.17
they	1.58	0.86	4.65	$< .001$	0.45
people/one	1.33	1.30	0.31	.756	0.03
there	0.28	0.23	1.67	.096	0.16

Table 2: Subject Type Density by L1 Group (per 100 words)

Further analysis in Section 5.2 examines which specific pronouns drive this difference, revealing that first-person singular "I" accounts for much of the effect.

5.2 Argumentation Style

To understand what drives the pronoun density difference observed in Section 5.1, we examined how learners express personal stance in argumentative writing. Using spaCy dependency parsing, we identified grammatical subjects and categorized them by type, following prior work on stance and subject realization in L2 writing (Wei et al., 2020; Crossley et al., 2016).

Table 2 presents the density of different subject types (per 100 words) and statistical comparisons between groups.

The most striking difference was in first-person singular "I": Japanese learners used "I" as subject nearly four times more frequently than Chinese learners (3.11 vs. 0.83 per 100 words, $d = 1.30$). This very large effect accounts for much of the overall pronoun density difference observed in Section 5.1. Japanese learners also used significantly more third-person "they" subjects ($d = 0.45$), while no significant differences emerged for "we," impersonal subjects, generic subjects, or existential "there." Importantly, Chinese learners did not compensate with higher use of impersonal or generic constructions, a pattern consistent with prior observations of stance expression in Chinese EFL writing (Wang and Wen, 2002; Wei et al., 2020). The lack of significant differences for "it/this/that" ($p = .087$) and "people/one" ($p = .756$) indicates that Chinese learners simply use fewer explicit stance markers overall, rather than substituting alternative constructions.

We also analyzed specific opinion-marking patterns, comparing direct "I + verb" constructions (e.g., "I think," "I agree," "I believe") with distanc-

ing phrases (e.g., "in my opinion," "from my perspective," "personally"). Japanese learners strongly preferred direct constructions (14.26 vs. 3.42 per 1000 words, $d = 1.80$), while Chinese learners showed only a slight preference for distancing phrases (1.80 vs. 1.03 per 1000 words, $d = -0.31$). This pattern aligns with previous findings that Japanese EFL writers favor explicit personal positioning, whereas Chinese writers tend to adopt less overt stance-marking strategies (Kubota and Lehner, 2004; Connor, 1996).

To verify that "I" subjects were indeed used for opinion-marking rather than narrative or background purposes, we conducted manual annotation of 100 randomly sampled sentences beginning with "I" (50 per group). Two annotators independently classified each sentence as Opinion (e.g., "I think," "I believe," "I agree") or Non-opinion (e.g., narrative, background, meta-discourse). Inter-annotator agreement was high ($\kappa = 0.83$, 93% agreement). Using only agreed annotations, both groups showed similar functional distributions: Japanese learners used "I" for opinions 75.6% of the time ($n = 45$), compared to 70.8% for Chinese learners ($n = 48$). This confirms that the difference between groups is one of frequency rather than function, consistent with arguments in learner corpus research that L1 influence often manifests quantitatively rather than categorically (Bestgen and Granger, 2014)

These findings support H2. Japanese learners rely heavily on first-person singular constructions to express personal stance, while Chinese learners use fewer explicit first-person markers without substituting impersonal or generic alternatives. This pattern may reflect different rhetorical traditions in L1 writing (Connor, 1996; Kubota and Lehner, 2004). The high frequency of "I" in Japanese learner writing directly contributes to higher pronoun density and consequently lower lexical diversity, connecting back to H1.

5.3 Discourse Cohesion

To assess discourse cohesion more accurately, we applied the hybrid method mentioned in Section 4, which mitigates the over-estimation bias of lexicon-based matching by filtering out non-functional usage of high-frequency words. Table 3 presents the density comparisons across functional categories.

We found no statistically significant difference in the overall density of discourse markers between Mandarin Chinese and Japanese learners (t = -1.33, p = 0.185). This suggests that learners from both L1 backgrounds at the B1 proficiency level employ explicit cohesive devices with similar total frequency.

Feature	Chinese	Japanese	t	p
Overall Density	3.129	3.254	-1.329	0.185
Causal	0.608	0.860	-3.896	< .001
Contrast	0.513	0.464	1.084	0.279
Continuation	1.758	1.502	2.783	0.006
Elaboration	0.066	0.051	0.897	0.370
Opening/Closing	0.482	0.659	-3.165	0.002
Sequencing	0.184	0.133	1.758	0.080
Topic Shift	0.387	0.426	-0.979	0.328
Summary	0.000	0.005	-1.418	0.158

Table 3: Discourse marker density by functional category (per 100 words).

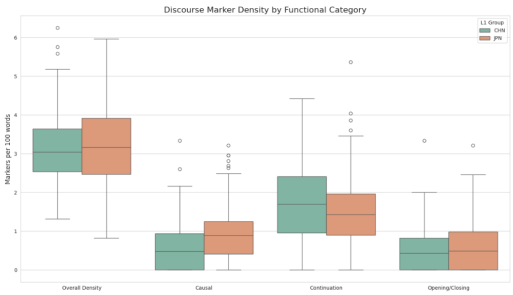


Figure 1: Discourse Marker Density by Functional Category

However, a breakdown by functional category reveals a distinct functional trade-off between the two groups. Japanese learners showed a higher density of Causal markers (e.g., because, so) (t = -3.90, p < 0.001) and Opening/Closing markers (t = -3.17, p = 0.002). This indicates a preference for explicitly signaling logical reasoning and text boundaries. In contrast, Chinese learners showed a higher density of Continuation markers (e.g., and, also, moreover) (t = 2.78, p = 0.006). As shown in Figure 1, these findings suggest differences in writing styles. Japanese learners focus on building a logical structure (like cause and effect), while Mandarin learners prefer to connect ideas by adding points one after another.

5.4 Syntactic Complexity

Syntactic complexity was measured using two complementary approaches: mean clause length (total words divided by total clauses) and subordination ratio (subordinate clauses divided by total clauses). Clauses were identified using spaCy dependency parsing, counting tokens with dependency relations indicative of clause heads (ROOT, advcl, ccomp, relcl, acl) as well as conjoined verbs. Table 4 re-

Subject Type	Japanese mean	Chinese mean	t	p	d
I	3.11	0.83	13.68	$< .001$	1.30
We	1.98	1.78	1.00	.319	0.10
it/this/that	1.38	1.53	-1.71	.087	-0.17
they	1.58	0.86	4.65	$< .001$	0.45
people/one	1.33	1.30	0.31	.756	0.03
there	0.28	0.23	1.67	.096	0.16

Table 4: Subject Type Density by L1 Group (per 100 words)

ports group-level descriptive statistics and statistical comparisons.

Chinese learners exhibited significantly higher syntactic complexity than Japanese learners on both measures. Chinese learners produced longer clauses (7.32 vs 6.69 words per clause, d = 0.80) and higher subordination ratios (47.4% vs 41.3%, d = 0.69). Both effects are medium-to-large in magnitude.

These findings support H4, confirming that the two L1 groups differ in syntactic complexity. The direction of the difference was not predicted in advance, but the results suggest that Chinese learners employ more syntactically embedded structures at the B1 level. One possible interpretation is that the similar SVO word order between Chinese and English facilitates transfer of subordination patterns, whereas Japanese learners managing the SOV-to-SVO transition may rely on simpler, shorter clauses to reduce cognitive load. This challenges common assumptions about paratactic organization in Chinese-influenced English writing.

6 Evaluation Strategy

The goal of the evaluation is to determine whether the linguistic differences observed between Mandarin Chinese and Japanese L1 writers at the B1 level are systematic and robust across the three dimensions examined: lexical diversity, syntactic complexity, and discourse cohesion. Our evaluation approach consists of two components: (1) statistical comparison of feature distributions across L1 groups (2) an interpretation-focused analysis linking the observed patterns to prior accounts of L1 influence in second language writing.

6.1 Statistical Evaluation

For each feature, we computed group-level descriptive statistics and conducted inferential tests to assess whether differences between the Mandarin and Japanese learner groups were statistically reliable.

- **Group Comparison Tests:** We used two-sample t-tests to compare mean feature values between the two groups. In cases where normality assumptions were not fully met, we verified that the results held using Mann-Whitney U tests.

- **Effect Size Estimates:** We reported Cohen's d for each comparison to characterize the magnitude of group differences. Following common interpretive guidelines, effect sizes were classified as small (|d| .20), medium (|d| .50), or large (|d| .80).

- **Multiple Comparison Considerations:** Because multiple discourse marker categories were tested, we applied a Holm–Bonferroni correction to control the familywise error rate, ensuring that the reported group differences were not due to chance.

These statistical analyses allowed us to determine whether the observed lexical, syntactic, and discourse patterns differentiated the two L1 groups beyond random variation.

6.2 Interpretation and Linguistic Analysis

In addition to statistical testing, we interpreted the observed group differences in the context of theories of cross-linguistic influence and contrastive rhetoric. Specifically, we:

- Examined which discourse marker categories showed meaningful differences and related these to known rhetorical tendencies in Mandarin and Japanese writing;

- Evaluated whether syntactic complexity patterns aligned with previously reported preferences for coordination (Chinese) and explicit connective signaling (Japanese);

- Assessed whether lexical diversity differences reflected instructional or exposure-based factors rather than structural language influence alone.

This interpretive component was central to the study's goals. Rather than predicting or classifying L1, our aim was to understand how specific linguistic features traced back to L1-influenced writing habits and what implications this had for L2 writing feedback and pedagogy.

6.3 Suitability of the Evaluation Framework

This evaluation strategy is consistent with learner corpus research that emphasizes interpretable, feature-level explanations of L1 influence (Bestgen and Granger, 2014), as opposed to Native Language Identification approaches that primarily prioritize classification accuracy over linguistic interpretation (Goswami et al., 2024). By combining statistical testing with linguistically grounded interpretation, our analysis aims to provide insights that are both empirically supported and pedagogically meaningful.

7 Conclusion

This study demonstrates that cross-linguistic influence in L2 English writing extends beyond surface-level lexical choices to encompass syntactic organization, discourse cohesion, and argumentation style (Jarvis and Odlin, 2000; Connor, 1996; Kubota and Lehner, 2004). Using ICNALE argumentative essays produced under controlled conditions (Ishikawa, 2023), we show that Mandarin Chinese and Japanese learners adopt distinct strategies for structuring arguments at the same proficiency level. Our findings align with prior learner-corpus evidence that L1 background shapes cohesion strategies and rhetorical patterning in L2 writing (**?**Wei et al., 2020; Lin, 2023). Overall, these results underscore the importance of providing L2 writing feedback that is sensitive to learners' L1 backgrounds rather than assuming a uniform developmental trajectory.

References

Haiyang Ai and Xiaofei Lu. 2013. A corpus-based comparison of syntactic complexity in nns and ns university students' writing. In Ana Díaz-Negrillo, Paul Thompson, and Nicolas Ballier, editors, *Automatic Treatment and Analysis of Learner Corpus Data*, pages 249–264. John Benjamins.

Yves Bestgen and Sylviane Granger. 2014. Quantifying the development of phraseological competence in l2 english writing: An automated approach. *Journal of Second Language Writing*, 26:28–41.

Leo Breiman. 2001. Random forests. *Machine Learning*, 45(1):5–32.

Ulla Connor. 1996. *Contrastive Rhetoric: Cross-Cultural Aspects of Second Language Writing*. Cambridge University Press.

Scott A. Crossley, Kristopher Kyle, and Danielle S. McNamara. 2016. The development and use of cohesive devices in l2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32:1–16.

Dhiman Goswami, Sharanya Thilagan, Kai North, Shervin Malmasi, and Marcos Zampieri. 2024. Native language identification in texts: A survey. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3149–3160.

Shin'ichiro Ishikawa. 2023. *The ICNALE Guide: An Introduction to a Learner Corpus Study on Asian Learners' L2 English*. Routledge.

Scott Jarvis and Terence Odlin. 2000. Morphological type, spatial reference, and language transfer. *Studies in Second Language Acquisition*, 22(4):535–555.

Ryuko Kubota and Al Lehner. 2004. Toward critical contrastive rhetoric. *Journal of Second Language Writing*, 13(1):7–27.

Jie Lin. 2023. A developmental study of chinese efl learners' use of discourse markers in l2 writings. *Journal of Linguistics and Communication Studies*, 21:18–31.

Xiaofei Lu. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4):474–496.

Shuichiro Nakayama. 2021. A corpus-based analysis of japanese EFL learners' linking adverbial use. *The Educational Review, USA*, 5(6):164–171.

Gerard Salton and Christopher Buckley. 1988. *Term-Weighting Approaches in Automatic Text Retrieval*, volume 24.

Miyuki Sasaki, Atsushi Mizumoto, and Akira Murakami. 2024. Developmental trajectories of multicompetent writers: An ecological-historical approach to l1/l2 writing abilities and l2 proficiency. *Journal of the European Second Language Association*, 8(1):114–130.

Wei Wang and Qiufang Wen. 2002. L1 use in the l2 composing process: An exploratory study of 16 chinese efl writers. *Journal of Second Language Writing*, 11(3):225–246.

Xiang Wei, Lawrence Jun Zhang, and Wen Zhang. 2020. Associations of l1-to-l2 rhetorical transfer with l2 writers' perception of l2 writing difficulty and l2 writing proficiency. *Journal of English for Academic Purposes*, 47:100907.

A Discourse Function Annotation Guidelines

Goal

To label each sentence based on the specific discourse function it performs in the argumentative text. The annotation focuses on the rhetorical purpose of the sentence (e.g., expressing cause–effect, structuring the argument) rather than the mere presence of keywords.

Label Set and Definitions

Causal

- **Definition:** Indicates cause–effect relationships, reasoning, or justification.
- **Key Markers:** because, so, therefore, as a result, since, thus.
- **Example:** "I worked part-time because I needed money."

Contrast

- **Definition:** Shows opposition, concession, counter-argument, or contrast.
- **Key Markers:** but, however, although, on the other hand, yet.
- **Example:** "The job was difficult, but I learned a lot."

Continuation

- **Definition:** Adds new information, extends a point, or lists items in a parallel structure.
- **Key Markers:** and, also, moreover, furthermore, in addition.
- **Example:** "I studied hard, and I worked part-time."

Elaboration

- **Definition:** Provides specific examples, detailed explanations, or clarification of a previous point.

Opening_Closing

- **Definition:** Signals the writer's stance, introduces the main argument, or concludes the essay.
- **Key Markers:** in my opinion, from my perspective, in conclusion, I think.
- **Example:** "In my opinion, students should work part-time."

Sequencing

- **Definition:** Indicates the chronological order of events or the logical steps in reasoning.
- **Key Markers:** first, next, then, finally, firstly, secondly.
- **Example:** "First, students learn responsibility."

Topic_Shift

- **Definition:** Explicitly signals a move to a different topic, perspective, or angle.
- **Key Markers:** anyway, now, as for.
- **Example:** "Now, I will discuss another reason."

Summary

- **Definition:** Summarizes previous points or restates the main idea for emphasis (distinct from stance-taking).
- **Key Markers:** to sum up, overall, in summary, all in all.
- **Example:** "Overall, part-time jobs are beneficial."

Annotation Principles

- **Function Over Form:** Labels are assigned based on rhetorical function in context, not keyword presence. A sentence containing "so" meaning "very" should be labeled as *None*.
- **Single Label:** Each sentence receives exactly one label. If multiple markers appear, annotators select the dominant rhetorical function.