Learning Challenges of Prepositions in English: A Corpus-based Study of "on" and Pedagogical Strategies for Chinese EFL Learners

Yuntian Yang

University of Adelaide, Adelaide, Australia
karvdtop@gmail.com

Abstract. Prepositions constitute a fundamental yet challenging area in English acquisition, especially for learners whose first language has a markedly different system. This study investigates the use of the English preposition on among Chinese EFL learners, drawing on corpus-based methods to identify common errors, their underlying causes, and effective pedagogical interventions. A mixed-methods approach was applied, combining quantitative error analysis with qualitative interpretation. A sample of 200 instances of on was randomly extracted from the Chinese and English Learner Language Corpus and compared with an equal number of tokens from native-speaker corpora. Results show that selection errors are the most frequent, followed by omission and addition errors, while order errors are rare. These errors can be traced to negative transfer from Mandarin Chinese, overgeneralization of simplified textbook rules, and insufficient acquisition of collocational patterns. The findings highlight the role of semantic differentiation, interlanguage development, and idiomatic competence in mastering prepositions. Pedagogically, the study proposes data-driven learning, explicit contrastive tasks, and collocation notebooks to enhance learner awareness and accuracy. This research contributes to second language acquisition theory and offers practical strategies for improving EFL instruction, with implications for broader preposition teaching beyond on.

Keywords: Preposition, corpus linguistics, Chinese EFL learners, error analysis, pedagogy

1. Introduction

Prepositions are frequent yet error-prone in English, as they encode spatial, temporal, and abstract relations. Mastery is crucial for communicative competence but difficult for second language learners, especially when the L1 diverges from English. For Chinese EFL learners, on is a common source of error. It appears in expressions like on the table and on the weekend, but these do not align with the Chinese preposition “在” (zài), which functions broadly as a locative and temporal marker. Overgeneralization of zài often causes negative transfer, where on is misused in contexts requiring in or at. Studies confirm that prepositions are among the top error categories in Chinese EFL writing, with in, on, and at particularly problematic [1]. Tahaineh found Arab learners showed comparable tendencies, illustrating Odlin’s Language Transfer Theory, where L1 structures shape L2

use [2, 3]. Interlanguage Theory also explains such errors, viewing the learner's system as a construct combining L1 and L2 features [4-6]. Within this system, errors arise from transitional competence, overgeneralized textbook rules (e.g., "on + day") and limited authentic input.

Corpus linguistics offers tools to analyze such errors. Sinclair's Idiom Principle argues much language is retrieved as prefabricated chunks [7]. For on, collaborations such as on time, on duty, on purpose, and on the other hand are frequent, but learners who fail to acquire them apply on mechanically [8]. Learner corpora, such as the Chinese and English Learner Language Corpus (CELL Corpus), allow comparison with native usage, quantify errors, and reveal pedagogical patterns [9].

This study asks: What errors do Chinese EFL learners make with on? What factors cause them? How can corpus-based pedagogy improve mastery? The findings aim to enrich SLA theory and suggest strategies to enhance accuracy and autonomy.

2. Methodology

This research employs a mixed-methods design that combines quantitative corpus analysis with qualitative error categorization. The approach follows the principles of Contrastive Interlanguage Analysis, comparing learner usage patterns with those of native speakers [10].

The primary L2 dataset is the Chinese and English Learner Language Corpus (CELL Corpus), which contains authentic written texts produced by Chinese EFL learners across proficiency levels [11]. The L1 benchmark is drawn from the CQPWeb platform, which provides access to multiple native-speaker corpora.

Two search patterns were employed to capture spatial and temporal uses of on:

Spatial: [word="on"] [pos="NN"] (e.g., on the desk, on the floor)

Temporal: [lemma="on"] [] "Monday" (e.g., on Monday morning)

A total of 200 instances of on were extracted from the learner corpus, along with 200 from the native-speaker corpus for comparison. The selection followed a random sampling method, ensuring that examples were evenly distributed across different learner texts and proficiency levels. This approach avoids bias toward particular topics or writing styles and allows for more representative error analysis.

Errors were categorized according to Corder's taxonomy: omission, addition, constitution, order, and selection errors [12]. This classification enables both the quantification of error types and the identification of their linguistic causes.

The analysis was guided by Language Transfer Theory to explain L1 interference; Interlanguage Theory to account for learner developmental stages [3, 4]; Idiom Principle to address collocational patterns [7]; Sociocultural Theory to inform pedagogical recommendations [13].

3. Analysis

The corpus analysis revealed clear patterns in the misuse of on by Chinese EFL learners. Across the 200 learner tokens examined, a total of 34 errors were identified, with selection errors comprising the majority (61.8%), followed by omission errors (20.6%), addition errors (14.7%), and a minimal occurrence of order errors (2.9%) (See Table 1). Constitution errors were not found. The dominance of selection errors is consistent with previous studies on prepositional misuse [1, 2].

Table 1. Error distribution for "on" in the CELL corpus

Error Type	Frequency	Percentage	Example	Correct Form
Selection Error	21	61.8%	He put the book on the bag	in the bag
Omission Error	7	20.6%	I sat chair for two hours	I sat on the chair
Addition Error	5	14.7%	We arrived on at school	We arrived at school
Order Error	1	2.9%	Tell me what on you put it	Tell me what you put it on
Constitution Error	0	0%	---	---

3.1. Selection errors

Selection errors are the most pervasive and are largely the product of negative transfer from Mandarin Chinese. In Mandarin, the preposition "在" (zài) is used in a wide range of locative and temporal contexts, which in English require different prepositions. This results in overextension of on into contexts where in or at is required.

Spatial selection errors:

• He lives on Beijing. → He lives in Beijing.
• The meeting is held on the room. → The meeting is held in the room.
• The flowers grow on the garden. → The flowers grow in the garden.
• The key is on my pocket. → The key is in my pocket.
• Students study on the library. → Students study in the library.

Temporal selection errors:

• on the night → at night
• on this year's summer holiday → during this year's summer holiday
• on the 1990s → in the 1990s
• on the winter → in winter
• on every morning → every morning

From the perspective of Insufficient Semantic Differentiation, learners have not internalized the subtle but crucial semantic distinctions among in, on, and at [14]. This is compounded by overgeneralization of textbook rules like "on + day" to inappropriate contexts, leading to non-idiomatic usage.

Pedagogical implication: For selection errors, targeted corpus-based collocation lists can be effective. Teachers can create frequency-ranked lists of correct on collocations from native corpora and contrast them with common learner errors, prompting students to discover patterns through data-driven learning.

3.2. Omission errors

Omission errors occur when on is required but omitted, often resulting in incomplete or ambiguous expressions. These errors are particularly frequent in spatial contexts. Spatial omission examples:

• There is a picture the wall. → There is a picture on the wall.
• The plate is the table. → The plate is on the table.
• The pen is the floor. → The pen is on the floor.
• He sat the bench for hours. → He sat on the bench.

Temporal omission examples:

• I will see you Monday morning. → I will see you on Monday morning.

• The meeting will be held Tuesday. → The meeting will be held on Tuesday.
• She arrived Christmas Day. → She arrived on Christmas Day.

These errors suggest that some learners may perceive prepositions as optional, influenced by the syntactic flexibility of Chinese, where locative and temporal relations can be expressed without an explicit preposition. From an Interlanguage Theory perspective, such omissions may reflect an early developmental stage where functional morphemes are underrepresented in learner output [5]. To address omission errors, awareness-raising tasks can be employed, where learners are presented with paired sentences (with and without prepositions) and asked to judge acceptability based on corpus evidence.

3.3. Addition errors

Addition errors involve the unnecessary insertion of on where it is not required, often due to overgeneralization or hypercorrection.

Examples:
• We arrived on at school. → We arrived at school.
• He discussed on about the problem. → He discussed the problem.
• She depends on to her parents for support. → She depends on her parents for support.
• He married on with his high school sweetheart. → He married his high school sweetheart.

These patterns suggest that learners sometimes conflate separate learned rules or apply on mechanically to verbs that do not require it, possibly influenced by the teaching of phrasal verbs (on as in carry on, depend on). Such verb-preposition mapping difficulties are well-documented in learner corpus research [8].

The recommended pedagogical implication is teachers can use verb-preposition mapping charts from corpora, showing which verbs require on, which require other prepositions, and which require none, to help learners refine their selection.

3.4. Order errors

Order errors are rare but worth noting because they affect sentence processing.

Example: Tell me what on you put it. → Tell me what you put it on.

Such errors may stem from direct translation from Chinese word order or from limited practice with syntactic flexibility in English questions and relative clauses.

For order errors, drill exercises involving wh-questions and preposition stranding can improve fluency and reduce processing delays.

3.5. Error trends and proficiency considerations

Although this study did not explicitly segment learners by proficiency level, preliminary observation suggests that omission errors are more common among lower-proficiency learners, while addition errors appear more in intermediate learners attempting to produce more complex sentences. Selection errors persist across proficiency levels, suggesting a deeper semantic issue that may require extended exposure and deliberate practice to resolve. A breakdown of error types by spatial vs. temporal usage reveals an interesting trend:

Table 2. Error types by semantic domain

Domain	Selection	Omission	Addition	Order
Spatial	15	5	4	1
Temporal	6	2	1	0

As shown in Table 2, spatial contexts account for the majority of errors, indicating that learners may find spatial uses of on more challenging than temporal uses. This aligns with findings in SLA research that spatial relations are more variable across languages than temporal relations, leading to greater cross-linguistic interference.

The analysis of Chinese EFL learners' use of the preposition on reveals a clear hierarchy of error types---selection errors dominate, followed by omission, addition, and minimal order errors. These patterns offer rich insights when interpreted through established SLA frameworks, particularly Language Transfer Theory, Interlanguage Theory, and the Idiom Principle.

From a Language Transfer perspective, the overgeneralization of on into contexts where in or at is required can be directly traced to the semantics of the Mandarin preposition "在" (zài). In Mandarin, zài functions as an all-purpose locative and temporal marker, covering a semantic range far broader than any single English preposition. When learners apply zài-based mapping rules to English, the result is frequent selection errors such as He lives on Beijing or on the night. This mapping strategy is not inherently illogical; rather, it reflects a cognitive economy where learners rely on familiar L1 categories to process L2 input and produce L2 output [5, 6].

Interlanguage Theory helps explain why such errors persist across proficiency levels. Selection errors are not merely the residue of beginner inaccuracy; they often become fossilized features of a learner's interlanguage, particularly when input lacks sufficient corrective feedback. For example, a learner who habitually writes in Monday instead of on Monday may never encounter enough corrective evidence in classroom or reading contexts to trigger restructuring. This persistence underscores the need for targeted, high-frequency exposure to correct forms in meaningful contexts. The Idiom Principle adds another layer to this explanation [7]. Native-speaker usage of on is heavily collocational. Expressions like on duty, on purpose, and on the other hand are not compositional in meaning and cannot be reliably generated from general prepositional rules. Learners who focus exclusively on rule-based instruction (e.g., "on + day of week") without sufficient exposure to these fixed expressions will lack the phraseological competence to deploy on idiomatically [8]. The addition errors in this study, such as discuss on about, likely reflect learners' attempts to "patch" their sentences with prepositions that feel appropriate based on partial analogies to learned patterns.

A further dimension to these findings is provided by Cognitive Linguistics, particularly the notion of image schemas [15]. The spatial sense of on involves a "support" schema---one entity rests on the surface of another. Temporal uses of on are metaphorical extensions of this schema to points in time (e.g., on Monday as a "surface" in the temporal domain). Chinese zài does not encode this support schema, instead functioning as a more general "existence at location/time" marker. The absence of a support-based conceptual distinction in the L1 may make it harder for learners to internalize the narrower range of on.

From a pedagogical standpoint, the evidence points to three complementary interventions. First, data-driven learning (DDL) using corpora can help learners notice the gap between their usage and native norms. By searching for on in a native-speaker corpus and comparing results with their own writing, learners can see high-frequency collocations in authentic contexts, reinforcing the correct semantic range of the preposition. Second, contrastive analysis tasks can make the L1--L2

differences explicit, particularly around high-interference items like on/in/at. Learners could be asked to translate short texts containing "在" into English, then use corpus tools to check prepositional accuracy. Third, collocation notebooks---where students record and review fixed expressions with on---can help build phraseological competence, which is essential for reducing both selection and addition errors.

Finally, the analysis also has implications for syllabus design. Traditional grammar-focused syllabi often treat prepositions as minor items within larger units on nouns or verbs, resulting in insufficient focused practice. Given the persistence and communicative impact of prepositional errors, a lexico-grammatical syllabus that integrates prepositions into thematic vocabulary units (e.g., on the bus, on time, on the list) could yield better long-term accuracy. This approach would combine the benefits of structural awareness with repeated lexical reinforcement, aligning with both SLA theory and empirical evidence from corpus linguistics [9].

4. Limitation

This study provides valuable insights into the challenges Chinese EFL learners face with the preposition on, but several limitations should be noted. First, the dataset comes exclusively from written academic texts in the CELL Corpus. Written language allows more planning and editing, so the errors identified may not fully represent spontaneous spoken usage. Future research should incorporate spoken corpora and compare results across modalities.

Second, the learner corpus used represents a relatively homogeneous group: mainly university-level Chinese learners with intermediate to advanced proficiency. This limits the generalizability to other populations, such as secondary school students, adult migrants, or learners from different L1 backgrounds. Cross-linguistic comparisons would help clarify which patterns are specific to Chinese learners and which are more universal in L2 English acquisition [9]. Third, error categorization relies on the researcher's judgment, especially when distinguishing selection and addition errors in ambiguous cases. Although Corder's taxonomy provides a clear framework, subtle semantic and pragmatic nuances may lead to misclassification [12]. The reliability of such classifications could be improved in future work through inter-rater agreement measures. Fourth, corpus extraction used automatic part-of-speech tagging to identify candidate tokens of on. Tagging errors, common in learner corpora, may produce false positives or negatives. Manual validation mitigates this somewhat but cannot eliminate the risk entirely.

Finally, focusing on a single preposition allows detailed analysis but narrows the scope. Learner difficulties with on may relate to other prepositions like in or at. Future research could examine multiple prepositions or adopt a longitudinal design to track how mastery of one affects others.

5. Conclusion

This study examined how Chinese EFL learners use the preposition on and identified the major patterns of error. Analysis of learner corpus data revealed that selection errors dominate, followed by omission and addition errors, with order errors being rare. These findings indicate that difficulties with on stem from negative transfer from Mandarin, incomplete differentiation among English prepositions, and limited acquisition of idiomatic collocations. The results highlight that mastering prepositions involves more than memorizing rules; it requires restructuring learners' conceptual categories and developing sensitivity to authentic usage. Learners need to internalize both semantic distinctions and collocational patterns in order to achieve accuracy and fluency. From a pedagogical perspective, the study proposes three complementary strategies. First, data-driven learning allows

students to observe authentic usage directly from corpora. Second, contrastive analysis tasks make L1–L2 differences explicit, helping learners avoid negative transfer. Third, keeping collocation notebooks encourages systematic review of fixed expressions with on. Together, these approaches can strengthen both awareness and autonomy.

Although the study focused on a single preposition, the methodology and insights are applicable to other prepositions and broader areas of lexico-grammar. Future research should expand to spoken data, include multiple prepositions, and explore longitudinal changes. By refining how prepositions are taught, educators can help learners reduce persistent errors and enhance communicative competence in English.

References

[1] Zhan, H. (2015) A corpus-based study of prepositional errors in Chinese EFL learners' writing. English Language Teaching, 8(5): 142–151. https: //doi.org/10.5539/elt.v8n5p142

[2] Tahaineh, Y. (2010) Arab EFL university students' errors in the use of prepositions. Modern Journal of Applied Linguistics, 1(1): 76–112.

[3] Odlin, T. (1989) Language transfer: Cross-linguistic influence in language learning. Cambridge University Press.

[4] Selinker, L. (1972) Interlanguage. International Review of Applied Linguistics in Language Teaching, 10(3): 209–231. https: //doi.org/10.1515/iral.1972.10.1-4.209

[5] Ellis, R. (2008) The Study of Second Language Acquisition (2nd ed.). Oxford University Press.

[6] Gass, S. M., & Selinker, L. (2008) Second Language Acquisition: An Introductory Course (3rd ed.). Routledge.

[7] Sinclair, J. (1991) Corpus, concordance, collocation. Oxford University Press.

[8] Laufer, B., & Waldman, T. (2011) Verb-noun collocations in second language writing: A corpus analysis of learners' English. Language Learning, 61(2): 647–672.

[9] Gilquin, G., & Granger, S. (2011) From EFL to ESL: Evidence from the International Corpus of Learner English. In Exploring the Lexis-Grammar Interface (pp. 72–92). John Benjamins.

[10] Granger, S. (1996) From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), Languages in contrast: Text-based cross-linguistic studies (pp. 37–51). Lund University Press.

[11] Leung, C. H., Chow, B. W.-Y., & Ge, G. (2022) Chinese and English Learner Language Corpus (CELL Corpus) [Data set]. Hong Kong Polytechnic University.

[12] Corder, S. P. (1967) The significance of learners' errors. International Review of Applied Linguistics in Language Teaching, 5(4): 161–170. https: //doi.org/10.1515/iral.1967.5.1-4.161

[13] Vygotsky, L. S. (1978) Mind in society: The development of higher psychological processes. Harvard University Press.

[14] Jarvis, S., & Pavlenko, A. (2008) Crosslinguistic influence in language and cognition. Routledge.

[15] Tyler, A., & Evans, V. (2003) The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning, and Cognition. Cambridge University Press.