

Corpus-Based Study of L1 Transfer in English Vocabulary Learning

Jingxuan Wang

College of International Studies, Shenzhen University, Shenzhen, Guangdong, 518060, China

ABSTRACT

Language transfer (L1 transfer) has long been a classic topic in second language acquisition research, and the influence of Mandarin Chinese on English learning is particularly prominent at the lexical level. Guided by corpus linguistics and second language acquisition theories, this study analyzes manifestations, types, and causes of negative transfer from Chinese in Chinese learners' English vocabulary use through contrasted analysis of English and Chinese corpora and evaluates the efficacy of corpus techniques in identifying and correcting transfer errors. First, we should review relevant theories of language transfer and interlanguage, including Lado's Contrastive Analysis Hypothesis, Jarvis and Pavlenko's Conceptual Transfer Hypothesis, and Cook's perspective on L1 usage as well as applications of corpora in vocabulary teaching. Next, combining empirical research design and learner language samples, we analyze instances of negative transfer caused by Chinese thinking patterns in meaning selection and collocational habits and explore their cognitive and linguistic underpinnings. Finally, we discuss teaching strategies for diagnostic analysis and corrective feedback using resources such as COCA and the Cambridge English Corpus. The findings show a high frequency of Chinese transfer errors in vocabulary collocation and semantic usage among Chinese learners and demonstrate that data-driven Learning (DDL) and corpus-assisted instruction effectively heighten learners' awareness of lexical conventions and enable teachers to design targeted activities, thereby mitigating interference from negative transfer. This research offers new perspectives for L2 vocabulary teaching and suggests greater utilization of large-scale corpora to enhance pedagogical practice.

KEYWORDS

Corpus linguistics; L1 transfer; English vocabulary learning; Second language acquisition

1 Introduction

The deepening of globalization has reinforced English as the lingua franca of international communication. For Chinese learners, proficiency in English has become a critical competitive asset for academic advancement and career development. However, L1 cognitive frameworks rooted in Mandarin Chinese lead learners to unconsciously transfer Chinese structures, thought patterns, or cultural habits into their English production, which is known as a phenomenon called language transfer. Transfer errors can occur at lexical, syntactic, and pragmatic levels. Negative transfer at the lexical level, characterized by its high frequency and subtlety, poses a major barrier to authentic English expression. Traditional English instruction often relies on teachers' experiential correction methods, lacking systematic empirical support.

Advances in corpus linguistics now provide novel tools and perspectives for studying language transfer. Large-scale electronic corpora enable systematic examination of actual target-language usage, offering credible references for L2 vocabulary teaching. Major corpora such as the Corpus of Contemporary American English (COCA) on English-Corpora.org and the Cambridge English Corpus, provided by Cambridge University Press, includes rich modern English data. COCA comprises over one billion words of American English since 1990 and is widely used in linguistic research^[1]. The Cambridge English Corpus covers global English variants to understand usage across contexts. Contrasting English and Chinese corpora reveals not only surface errors from L1 transfer but also allows in-depth semantic and cognitive analysis. Studies comparing native-speaker corpora and Chinese learner corpora confirm that Chinese thinking patterns negatively influence English collocations^[2]. Consequently, integrating corpus methods with SLA theories enriches theoretical studies into transfer phenomena and provides data-driven support for vocabulary pedagogy. This study focuses on negative L1 transfer in Chinese learners' English vocabulary use and evaluates corpus techniques in identifying and correcting transfer misuse.

2 Literature review

Language transfer research originated in the mid-20th century, evolving from behaviorism and contrastive analysis to cognitive semantics. Lado posited that learners tend to transfer L1 structures and meanings into L2, advocating that teachers prevent negative transfer by contrasting linguistic differences^[3]. This contrastive approach laid the foundation for L1 transfer studies and guided language teaching principles. Subsequently, Selinker's interlanguage theory proposed that learners develop a language system intermediate between L1 and L2, containing systematic errors including transfer effects^[4]. The multi-factorial influence model further emphasized that L1 is only one of many factors affecting L2 acquisition.

In the 21st century, transfer research has shifted emphasis to cognitive mechanisms and motivations. Jarvis and Pavlenko's Conceptual Transfer Hypothesis argue that differing lexicalization patterns for identical concepts can lead to transfer errors, highlighting cross-linguistic thinking modes^[5]. This view aligns with the Sapir-Whorf hypothesis, which points out language shapes thought and perception. Odlin contended that conceptual transfer reflects linguistic

relativity: distinct conceptual systems provoke transfer phenomena^[6]. Chinese scholars Li and Liu further explained that mastering one language entails acquiring its conceptual system, which naturally influences the use of another^[7].

Cook observed that even highly proficient L2 users process information less efficiently in the L2 than in the L1, advocating controlled L1 use in the classroom to facilitate comprehension, memory, and reasoning^[8]. In vocabulary acquisition, L1 can aid in semantic mapping and associative inference but may also trigger misuse—for instance, the Chinese terms “世界” (world) and “宇宙” (universe) are often conflated in Chinese, leading to confusion if translated directly^[8]. Thus, semantic equivalence cannot be assumed; attention to semantic fields and collocational constraints is essential.

Corpus linguistics emphasizes inductive observation from authentic data, contrasting with traditional subjective hypotheses. Corpus-driven and corpus-assisted learning methods have gained popularity in vocabulary instruction, providing frequency data, collocational patterns, and semantic prosody. Corpora such as COCA and the British National Corpus (BNC) allow learners to query usage across registers. Yu highlighted that corpus functions—frequency statistics, morphological searches, and concordance analysis—support deep learning of word forms, meanings, and collocations^[9]. Empirical classroom studies confirm that tasks involving corpus queries significantly enhance learners’ lexical awareness and motivation. For example, through COCA searches, learners observe authentic collocations and reduce erroneous inferences based on intuition or L1 transfer.

Empirical research on lexical transfer among Chinese learners has also proliferated. Chang’s corpus analysis of 52 Chinese university essays revealed high frequencies of collocational errors primarily attributable to L1 transfer^[2]. The study combined quantitative error coding with qualitative corpus examples to demonstrate the centrality of L1 influence. Zeng compared COCA with the Chinese Learner English Corpus (CLEC) to examine differences in quick/fast usage, finding significant L1-related collocational discrepancies^[10]. These findings confirm that corpus methods effectively reveal cognitive influences of L1 on L2 vocabulary.

However, few domestic studies apply corpus techniques specifically to analyze Chinese transfer, and pedagogical recommendations tend to be superficial—emphasizing language awareness, immersion, and practice. This study addresses these gaps by systematically tracing negative L1 transfer in learner production and proposing targeted corrective strategies, thereby informing future corpus-assisted pedagogy in English instruction.

3 Analysis: diagnosing and intervening in lexical transfer from a corpus perspective

3.1 Research design and corpus techniques

This study employs a combination of corpus comparison and error analysis. A learner corpus of Chinese English writing or translation is compiled or derived from existing resources (e.g., CLEC). COCA serves as a native-speaker reference. Target lexical items and collocations are analyzed via keyword searches and collocation tools. Frequencies, contexts, and co-occurrence patterns in learner and reference corpora are contrasted to identify potential L1 transfer. For instance, degree and intensity modifiers (e.g., very, quite, strong) are queried to examine whether learners overgeneralize Chinese gradable terms into inappropriate English usage. A pre- and post-test design assesses corpus-based instructional effects: students write an essay, receive corpus-driven collocation practice and discovery tasks, and subsequently produce a revised text. Error rate changes validate the efficacy of corpus-assisted learning.

3.2 Types of lexical transfer in learners’ output

Learner lexical misuse manifests in semantic selection, collocational choice, and fixed expressions. In semantic selection, Chinese-English concept differences lead to direct translation errors (unconscious transfer). For example, learners may write “He was tired very” to mirror the Chinese emphasis “他太累了,” which is unidiomatic in English. Collocational errors are prevalent due to Chinese collocational flexibility lacking strict English constraints, resulting in “Chinglish collocations,” such as “strong rain” for “heavy rain” or “torrential rain.” Synonym misuse also stems from L1 transfer: “fast” and “quick” are conflated without sensitivity to semantic and collocational nuances^[10]. Students often lack sensitivity to the concept of collocation and frequently apply Chinese usage patterns instead. For example, although directly translating “打电话” as “make a call” is not grammatically incorrect, native English typically prefers “make a phone call” or varies based on context. In some cases, learners even produce hybrid expressions by mixing English and Chinese syntactic logic. For instance, the sentence “We should pay more attention to environmental protection” may be misinterpreted in Chinese as “关注环境保护更重要,” which is then awkwardly retranslated into English as “Pay attention to protect the environment more.” In idiomatic and fixed expressions lacking Chinese equivalents, direct structural borrowing occurs—translating “谈天说地” as “talk sky and earth” instead of idiomatic English alternatives like “chat about everything.” These phenomena reveal that Chinese transfer extends beyond literal mistranslation to underlying conceptual mapping differences.

3.3 Causes of Chinese transfer

Negative L1 transfer arises from linguistic and cognitive factors. Linguistically, Chinese and English differ in lexical

semantics, collocational rules, and morphological marking. Chinese polysemy and broad semantic boundaries contrast with English's finer lexical distinctions—e.g., Chinese “家人” covers ‘parents,’ ‘siblings,’ and ‘relatives’ in English. A direct translation often overlooks subtle semantic distinctions. Moreover, the high degree of collocational flexibility in Chinese and the lack of morphological variation lead learners to organize English expressions based on their native language logic. This conceptual transfer is evident when Chinese cognitive frameworks influence second language L2 expression^[11]. Cognitively, learning strategies that rely on piecemeal memorization rather than holistic collocational perception predispose learners to word-for-word translation. Culturally, intrinsic Chinese concepts like “面子” (‘face’) lack direct English correspondents, prompting awkward literal renderings. Thus, L1 transfer emerges from the interplay of conceptual schemas, cognitive strategies, and cultural constructs.

3.4 Corpus-based identification and correction

Corpus tools offer robust means to detect and remediate transfer errors. By querying native corpora, teachers can verify whether learner collocations align with authentic usage. For “strong rain,” COCA concordance searches reveal high-frequency collocates “heavy rain” and “torrential rain,” not “strong rain.” Similarly, Zeng’s analysis of quick/fast collocational patterns in COCA highlights contextual preferences, enabling learners to inductively deduce appropriate usage^[10]. Corpora also supply abundant example sentences for learners to internalize collocational norms. Learners gradually revise misunderstandings caused by L1 transfer by analyzing how vocabulary is used across different contexts.

Pedagogically, instructors can design corpus-based tasks to heighten sensitivity to transfer errors. A middle school vocabulary instruction case used COCA-driven activities to enhance awareness of polysemy and collocational rules, fostering autonomous inquiry and motivation^[12]. We recommend a diagnostic–corpus intervention model: The initial assessment identifies learner transfer patterns; subsequent COCA and Cambridge Corpus queries allow students to analyze and summarize authentic collocations. This data-driven learning approach aligns with the “assessment for learning” principle in the Chinese High School English Curriculum Standards, facilitating individualized correction.

Furthermore, teachers should also develop corpus resources targeting frequent L1 transfer issues, such as collocation comparison lists and semantic prosody reports, to inform instruction. Group-based corpus projects can cultivate learner corpus literacy^[12]. In sum, corpus techniques significantly aid error identification and evidence-based correction, supporting both research and pedagogy in lexical transfer.

4 Conclusion

L1 transfer, particularly negative transfer from the mother tongue in L2 vocabulary learning, remains a focal concern in SLA. This study reviewed foundational theories and empirical findings, demonstrating various manifestations of Chinese transfer in learner vocabulary use: semantic misselection, collocational misuse, and structural deviations. Causes lie in divergent conceptual systems, collocational habits, and learning strategies. Comparative corpus analysis using COCA and Cambridge English Corpus provides objective usage data, uncovering error patterns and guiding corrective instruction. Evidence suggests that corpus-driven vocabulary teaching enhances learners’ contextual awareness and counteracts L1 thinking inertia.

Future instruction should prioritize corpus resources: teachers must systematically demonstrate corpus querying and analysis functions, guiding learners in semantic differentiation and collocation discovery. Integrating corpus tasks into regular assessment normalizes evidence-based inquiry. Through corpus-assisted learning, learners construct vocabulary meaning in authentic contexts, gradually overcoming negative L1 transfer and improving lexical accuracy and fluency.

References

- [1] Mark Davies. (2020) English-Corpora.org User Guide. <https://www.english-corpora.org/pdf/english-corpora-zh.pdf>.
- [2] Chang, Y. X. (2019) Lexical Transfer Phenomena in Chinese EFL Learners' Collocation Use. *Teaching and Research of International English Exams*, 1: 121–129.
- [3] Lado, R. (1957) *Linguistics across Cultures: Applied Linguistics for Language Teachers*. University of Michigan Press, Michigan.
- [4] Selinker, L. (1972) Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10: 209–232.
- [5] Jarvis S., Pavlenko A. (2008) *Crosslinguistic Influence in Language and Cognition*. Routledge, New York.
- [6] Odlin, T. (1989) *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge University Press, New York.
- [7] Li X.J., Liu Y.B. (2013) From Comparative Analysis to Conceptual Transfer: The Evolution of Theoretical Research on Language Transfer. *Journal of Northeast Normal University (Philosophy and Social Sciences)*, 1: 101–104.
- [8] Cook, V. (2001) Using the First Language in the Classroom. *The Canadian Modern Language Review*, 57: 402–423.
- [9] Fund for Assessment Research in Foreign Language Education. (2020) Lesson Design: Corpus-Assisted Vocabulary Instruction Based on Diagnostic Assessment. <https://far.fltrp.com/c/2020-08-18/525565.shtml>.
- [10] Zeng, N. (2024) A Corpus-Based Comparative Study on English Synonyms—The Case Study of “Quick” and “Fast”. *Modern Linguistics*, 12: 1023–1029.
- [11] Yu H., Li L.J. (2017) A Review of Studies on Conceptual Transfer. *Modern Linguistics*, 5: 166–174.
- [12] Zhang J.Y., Li Y.H. (2025) A Corpus-Assisted Approach to Vocabulary Instruction in Junior Secondary EFL Classrooms. *International Journal of Education*, 7: 116–120.