

PAPER

CHUNK-ORIENTED INSTRUCTION IN TEACHING CHINESE AS A FOREIGN LANGUAGE: AN AI-ENHANCED AND CORPUS-INFORMED TRANSLATION APPROACH

Xursanova Farida Farhod qizi^{1,*}

¹Shahrisabz Davlat Pedagogika Instituti

* qobulova.g@urdu.uz

Abstract

Recent advances in cognitive linguistics and artificial intelligence have reshaped perspectives on foreign language pedagogy. Chunk-oriented instruction, which focuses on formulaic multi-word units, has proven effective in second language acquisition; however, its application in Teaching Chinese as a Foreign Language (TCFL) remains insufficiently developed. This study reinterprets chunk theory through the lenses of corpus linguistics, translation studies, and AI-assisted language learning. By examining recurrent Chinese language chunks and their translational behavior, the study argues that chunk-based pedagogy significantly improves learners' fluency, pragmatic accuracy, and translation quality. The findings indicate that integrating chunk theory with corpus-driven analysis and AI-supported translation tools offers a sustainable and innovative model for TCFL.

Key words: chunk-based instruction; TCFL; corpus-informed teaching; AI-assisted translation; formulaic sequences

INTRODUCTION

The growing influence of artificial intelligence and data-driven methodologies has accelerated changes in foreign language education. Traditional grammar-focused approaches are increasingly

challenged by pedagogical models that emphasize language use, cognition, and contextualized learning. Within this paradigm shift, chunk theory has emerged as a powerful explanatory framework, highlighting the role of prefabricated linguistic units in language processing.

Compiled on: February 14, 2026.

Copyright: ©2026 by the authors. Submitted to *Advances in Science and Humanities* for possible open access publication under the terms and conditions of the [Creative Commons Attribution \(CC BY\) 4.0 license](https://creativecommons.org/licenses/by/4.0/).

Although chunk theory has been widely adopted in English language teaching, its systematic integration into TCFL remains limited. Learners of Chinese often encounter difficulties not because of individual lexical items, but due to inappropriate collocations, unnatural sentence patterns, and culturally embedded expressions—issues that become particularly evident in translation tasks.

This paper proposes a re-conceptualization of chunk-based instruction in TCFL by incorporating insights from corpus linguistics and AI-assisted translation. It aims to demonstrate that chunk-oriented pedagogy offers a cognitively efficient and pedagogically effective solution for both language learning and translation training.

METHODOLOGY AND THEORETICAL ORIENTATION

This research adopts a qualitative and interdisciplinary approach, drawing upon theories of chunk-based learning, cognitive linguistics, and translation studies. The analytical framework is grounded in the premise that language competence develops through repeated exposure to meaningful, recurrent linguistic patterns rather than isolated grammatical rules. Representative Chinese language chunks were selected based on their frequency and communicative relevance as reported in grammar references and corpus-based research. These chunks were examined with respect to their structural flexibility, pragmatic function, and translational equivalence. Additionally, this study explores the role of AI-assisted translation systems from a conceptual perspective, focusing on how chunk-level processing contributes to improved translation coherence and semantic accuracy. The analysis does not rely on experimental data but instead synthesizes theoretical insights with pedagogical observations. Corpus-Informed Characteristics of Chinese Chunks Corpus-based observations reveal that many commonly used Chinese expressions function as stable units rather than freely generated combinations. These units can be broadly categorized into fixed expressions, grammatical constructions, and semi-flexible sentence frames. Corpus analysis indicates that high-frequency Chinese chunks often function as

stable units in discourse. These include:

- Fixed lexical chunks (e.g., idioms, set phrases);
- Grammatical chunks (e.g., “因为……所以……”, “一……就……”);
- Sentence-frame chunks that allow limited substitution.

Corpus evidence shows that these chunks occur more frequently than randomly generated word combinations, confirming their status as formulaic language units.

Certain Chinese chunks originate from highly grammaticalized lexical items whose meanings extend beyond their literal sense. Expressions such as 总算, 简直, 始终, and 万一 function as holistic units that encode stance, emphasis, or conditionality. Corpus evidence suggests that such items rarely operate independently and are typically processed as fixed lexical units.

In addition, short fixed expressions including evaluative and affective phrases (e.g., expressions equivalent to to be honest or to feel both amused and helpless) exhibit strong contextual dependency. Idiomatic chunks formed through metaphorical mapping further illustrate the lexicalized nature of Chinese formulaic language. These expressions convey condensed cultural meaning and resist word-level substitution, which makes them particularly challenging for foreign learners and translation systems.

Such chunks demonstrate high frequency and strong contextual constraints, supporting the view that they serve as core building blocks in Chinese discourse. Their recurrent nature makes them particularly suitable for instructional focus in TCFL.

Implications for Translation Practice

From a translation perspective, chunk-oriented competence enables learners to move beyond literal word-for-word rendering. Translating at the chunk level allows for better preservation of pragmatic intent and cultural nuance, especially in Chinese–English and Chinese–Uzbek translation contexts. Learners trained in chunk-based processing show greater awareness of functional equivalence, resulting in translations that are more natural and target-language oriented.

AI-Supported Chunk Processing

AI-based translation models increasingly benefit from chunk-sensitive input. When language data are processed as meaningful units rather than isolated tokens, translation outputs demonstrate improved fluency and coherence. This observation underscores the pedagogical value of aligning human language instruction with machine processing mechanisms, thereby reinforcing the relevance of chunk theory in AI-enhanced TCFL environments.

DISCUSSION

The integration of chunk theory with corpus linguistics and artificial intelligence reflects a broader transition toward cognitively and technologically informed language teaching. Chunk-oriented instruction bridges the gap between human language acquisition and automated language processing, offering benefits for both learners and AI systems.

In TCFL contexts, this approach mitigates persistent challenges related to syntactic rigidity and cultural mismatch. Corpus-informed chunk selection further strengthens instructional validity by grounding teaching content in authentic language use.

The findings suggest that chunk theory serves as a bridge between human cognition and machine language processing, making it a valuable framework for both teaching and translation.

Chunk-based teaching also contributes to enhancing learners' confidence and motivation. Compared with rule-based learning, chunk learning enables learners to produce meaningful utterances at an early stage, which gives them a sense of achievement.

One of the most significant advantages of chunk-based teaching is its positive impact on learners' communicative competence. Since chunks are prefabricated units stored in long-term memory, learners can retrieve and use them quickly during communication, reducing cognitive load and improving fluency.

For learners of Chinese, mastering chunks helps overcome difficulties caused by word order, grammatical constraints, and cultural differences. As learners accumulate a large number of chunks, they gradually develop a more native-like sense of language use, leading to greater accuracy and naturalness in expression.

As learners become more capable of expressing themselves fluently using chunks, their anxiety decreases and their willingness to communicate increases. This positive learning experience further motivates learners to engage actively in Chinese language learning.

CONCLUSION

This study demonstrates that chunk-oriented instruction, when enriched by corpus analysis and AI-assisted translation, constitutes a robust and forward-looking model for Teaching Chinese as a Foreign Language. By emphasizing recurrent linguistic patterns, this approach enhances fluency, translation competence, and intercultural sensitivity. Future research should pursue empirical validation through learner corpora and experimental studies, as well as explore the development of intelligent chunk-based teaching platforms. Such efforts will contribute to the continued modernization of TCFL pedagogy.

References

1. Lewis, M. (1993). *The Lexical Approach*. Hove: Language Teaching Publications.
2. Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
3. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
4. Wray, A. (2006). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.