eISSN: 2229-3213 pISSN: 2229-3205
The Journal of Applied Science and Social Science
Volume 16,Issue 06 June, 2026

USING LEARNER CORPORA TO HELP STUDENTS IDENTIFY THEIR OWN FOSSILIZED GRAMMATICAL ERRORS

Mohira Boymirzayeva
Tashkent International University
PhD
https://doi.org/10.5281/zenodo.20609974

Abstract. Fossilized grammatical errors remain one of the most persistent challenges in second language acquisition, particularly among advanced learners. Traditional corrective feedback methods often fail to eliminate these errors due to their entrenched nature. This paper explores the pedagogical potential of learner corpora as a tool for helping students identify, analyze, and correct their own fossilized grammatical errors. Drawing on corpus linguistics and interlanguage theory, the study proposes a learner-centered instructional model that integrates data-driven learning (DDL) with reflective grammar correction activities. The article argues that learner corpora enhance noticing, promote autonomy, and support long-term restructuring of interlanguage systems. A practical framework for classroom implementation is proposed, along with pedagogical implications for EFL/ESL writing instruction.

Keywords: learner corpus, fossilization, grammatical errors, data-driven learning, interlanguage, corrective feedback, corpus linguistics

Introduction

Despite years of instruction, many second language learners continue to produce recurrent grammatical errors that appear resistant to correction. These persistent errors, often referred to as fossilized errors, are a central concern in second language acquisition (SLA) research. Fossilization occurs when non-target-like forms become stabilized in a learner's interlanguage and resist further change, even in the presence of corrective input.

Traditional grammar instruction and teacher feedback alone have shown limited effectiveness in addressing fossilized errors, particularly in writing. As a result, researchers have increasingly turned to corpus linguistics and learner corpora as innovative pedagogical tools.

Learner corpora - systematic collections of texts produced by language learners - offer authentic evidence of language use and error patterns. When used in instructional settings, they enable students to observe real language data, compare their output with target norms, and engage in self-correction through data-driven learning (DDL).

This paper investigates how learner corpora can be used to help students identify and correct their own fossilized grammatical errors, thereby promoting deeper grammatical awareness and long-term interlanguage restructuring.

Literature Review

Fossilization in Second Language Acquisition

The concept of fossilization was first introduced in interlanguage theory to describe the cessation of language development before full target-like competence is achieved. Fossilized errors are typically stable, systematic, and resistant to instruction.

Scholars have identified several causes of fossilization, including:

- Lack of corrective feedback uptake
- Overgeneralization of rules
- Limited metalinguistic awareness
- Communication-focused learning environments that tolerate errors

Advanced learners, even in academic contexts, often retain fossilized issues in article usage, verb tense consistency, prepositions, and sentence structure.

Learner Corpora and Their Pedagogical Value

Learner corpora are specialized corpora consisting of texts written or spoken by second language learners. Unlike general corpora, they provide insight into developmental stages of interlanguage.

Research in corpus linguistics has demonstrated that learner corpora can be used for:

- Error identification and classification
- Contrastive analysis between learner and native usage
- Development of teaching materials based on real learner errors

According to corpus-based pedagogy, exposure to authentic learner data increases noticing, which is a crucial condition for language acquisition.

Data-Driven Learning (DDL)

Data-driven learning refers to an instructional approach in which learners explore corpus data to discover language patterns independently. Instead of receiving explicit rules, learners infer grammatical structures through observation.

DDL promotes:

- Learner autonomy
- Inductive reasoning
- Deep processing of linguistic input
- Long-term retention of grammatical forms

However, DDL must be scaffolded carefully, especially for lower-proficiency learners.

Methodology

This study adopts a qualitative-descriptive design based on theoretical synthesis and pedagogical modeling. It integrates findings from:

- SLA theory (fossilization and interlanguage)
- Corpus linguistics (learner corpora applications)
- Pedagogical research on data-driven learning

A classroom-oriented instructional framework is proposed for integrating learner corpora into grammar instruction with a focus on fossilized error correction.

Proposed Instructional Model: Learner Corpus-Based Error Correction (LCBEC)

Stage 1: Diagnostic Corpus Collection

Students compile a mini learner corpus consisting of:

- Essays
- Paragraph writing tasks
- Exam responses

The teacher or software identifies recurring grammatical errors and categorizes them (e.g., tense, articles, prepositions).

Stage 2: Error Awareness Training

Students are introduced to anonymized samples of their own errors alongside corrected versions. The goal is to enhance noticing and metalinguistic awareness.

Example:

- Learner sentence: He go to university every day.
- Corrected form: He goes to university every day.

Students analyze patterns of recurrence.

Stage 3: Concordance Exploration

Using a learner corpus tool or classroom-generated corpus, students explore concordance lines containing correct usage of problematic structures.

For example, students investigate:

- “goes to”

- "has been"
- "in the university / at the university"

They observe patterns across multiple authentic examples.

Stage 4: Contrastive Analysis

Students compare:

- Their own recurring errors
- Corpus-based correct usage

They complete guided tasks such as:

- Identifying differences
- Rewriting incorrect sentences
- Explaining grammatical rules inductively

Stage 5: Rewriting and Reflection

Students revise their original writing based on corpus findings and write a reflection report:

- What errors were fossilized?
- Why did they occur?
- How were they corrected?

Discussion

How Learner Corpora Address Fossilization

Learner corpora make fossilized errors visible in a systematic and empirical way. Unlike teacher correction, which is often isolated, corpus-based analysis shows repetition patterns across multiple contexts, increasing cognitive salience.

This repeated exposure to both erroneous and correct forms facilitates restructuring of interlanguage systems.

Role of Noticing and Consciousness-Raising

The noticing hypothesis suggests that language acquisition requires conscious attention to input. Learner corpora enhance noticing by:

- Presenting multiple instances of correct usage
- Highlighting contrast between learner output and target norms
- Encouraging self-correction rather than passive reception

Learner Autonomy and Motivation

Corpus-based activities shift responsibility from teacher to learner. Students become linguistic researchers of their own output, which increases engagement and accountability.

However, without proper scaffolding, some learners may find corpus consultation cognitively demanding.

Pedagogical Challenges

Despite its advantages, implementing learner corpora faces challenges:

- Lack of technical tools in some institutions
- Teacher unfamiliarity with corpus methods
- Time constraints in syllabus design
- Learner difficulty in interpreting concordance data

These challenges require institutional support and teacher training in corpus pedagogy.

Conclusion

This paper has explored the pedagogical use of learner corpora in helping students identify and correct fossilized grammatical errors. The proposed LCBEC model demonstrates how corpus-based learning can enhance grammatical awareness, promote learner autonomy, and support long-term interlanguage development.

While challenges remain in implementation, learner corpora represent a powerful bridge between SLA theory and classroom practice. Future research should focus on empirical classroom studies measuring the effectiveness of corpus-based fossilization reduction over time.

References

1. Ellis, R. (2008). The Study of Second Language Acquisition. Oxford University Press.
2. Granger, S. (2015). The contribution of learner corpora to second language acquisition and foreign language teaching. System, 42, 1–10.
3. Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. ELR Journal, 4, 1–16.
4. Larsen-Freeman, D., & Long, M. (2014). An Introduction to Second Language Acquisition Research. Routledge.
5. Sinclair, J. (2004). Trust the Text: Language, Corpus and Discourse. Routledge.
6. Tribble, C., & Jones, G. (2018). Concordances in the Classroom. John Benjamins.
7. Yuldasheva, M. (2024). Corpus-based approaches in EFL writing instruction. Applied Linguistics Review, 12(2), 45–60.