

Analysis of the use of phonology-based Mongolian orthographic rules in university students' writing

Lkhamjav Myadag^{1*} , Zugderdemid Nyamjargal²

¹Zokhiomj University, Ulaanbaatar, Mongolia

²Nalaikh Polytechnic College, Ulaanbaatar, Mongolia

*Corresponding author: Lkhamjav Myadag, myadag779@gmail.com

CITATION

Lkhamjav Myadag, Zugderdemid Nyamjargal. Analysis of the use of phonology-based Mongolian orthographic rules in university students' writing. International Journal of Social Science and Humanities Research-MIYR, 2026, 6(3), 49~57.

<https://doi.org/10.53468/mifyr.2026.6.3.49>

ARTICLE INFO

Received: 29 August 2026
Revised: 02 September 2026
Accepted: 15 September 2026
Available online: 30 September 2026

COPYRIGHT



Copyright © 2026 by author(s).
International Journal of Social
Science and Humanities Research-
MIYR is published by Misheel
institute for young researcher's NGO.
This work is licensed under the
Creative Commons Attribution (CC
BY 4.0) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract- Orthographic competence is a fundamental component of academic writing quality and plays a crucial role in maintaining the standardized norms of the Mongolian literary language. Among the subsystems of Mongolian orthography, phonology-based spelling rules account for a substantial proportion of writing errors. However, corpus-based studies examining the actual use of these rules in university students' written texts remain limited. The present study aimed to identify the characteristics and common patterns of phonology-based orthographic rule usage in students' writing through corpus and statistical analyses. A corpus consisting of 10,000 words was compiled from authentic student writing. Orthographic errors were identified using Microsoft Word Review Count and AntConc, while descriptive statistical analyses, including frequency, percentage, mean, and standard deviation, were performed using IBM SPSS Statistics 29.0. The analysis revealed 3,761 misspelled words, representing 37.6% of the corpus. Of these, 1,829 errors (48.6%) were classified as phonology-based orthographic errors. The most frequent error categories were the rule of elided reduced vowels (7.9%), the distinction between “-ii” and “-y” suffixes (5.1%), the non-elided reduced vowel rule (4.0%), the rule governing the transformation of the soft sign (ᠡ) into “i” (3.5%), and vowel-sonorant consonant correspondence (3.1%). Descriptive statistics further indicated that foreign-word spelling, distinguishing vowels, non-elided reduced vowels, and vowel harmony exhibited the highest levels of orthographic difficulty. These findings provide empirical evidence for improving the teaching methodology of Mongolian orthography and contribute to the development of evidence-based academic writing instruction.

Keywords: Mongolian orthography, Phonology, Reduced vowels, Soft sign, Corpus linguistics, Academic writing, SPSS

1. INTRODUCTION

The Mongolian literary language represents a standardized linguistic system governed by orthographic norms that ensure consistency, clarity, and mutual intelligibility in written communication [1]. Orthographic rules are founded upon phonological, morphological, grammatical, and semantic principles, making accurate spelling an essential indicator of academic literacy rather than merely linguistic knowledge [7].

Recent educational reforms have emphasized the development of academic writing skills in higher education. Nevertheless, phonology-related spelling errors continue to occur

frequently in university students' written texts. Errors involving reduced vowels, distinguishing vowels, vowel harmony, and the soft sign (ᠪ) are particularly prevalent because these rules require learners to reconcile differences between pronunciation and orthographic representation [2].

Previous studies have primarily investigated orthographic competence among primary and secondary school students, focusing on spelling achievement and error classification. Comparatively few studies have employed authentic student writing corpora to examine the practical use of phonology-based orthographic rules in higher education through quantitative statistical analysis. Consequently, empirical evidence describing university students' actual orthographic performance remains insufficient.

Recent research in the Mongolian educational context has also emphasized the importance of examining students' actual performance in learning activities, rather than relying solely on their reported interest or perceived competence [10]. This perspective further supports the need to examine learners' actual performance through evidence derived from authentic educational contexts.

Therefore, this study investigates the structural characteristics, frequency, and application level of phonology-based orthographic rules by analyzing a corpus of authentic student writing. The study integrates corpus linguistics with descriptive statistical analysis to provide evidence-based implications for improving Mongolian orthography instruction.

1.1. Research objective

To identify the characteristics and common error patterns in the use of phonology-based Mongolian orthographic rules in university students' writing through corpus-based statistical analysis.

- ① To classify orthographic errors identified in students' written texts.
- ② To determine the structure and frequency of phonology-based spelling errors.
- ③ To evaluate the application level of individual orthographic rules using descriptive statistics in SPSS.
- ④ To propose pedagogical implications for improving Mongolian orthography instruction based on the findings.

1.2. Materials and Methods

This study employed a quantitative document analysis design integrated with corpus linguistics to investigate the use of phonology-based Mongolian orthographic rules in university students' writing. Document analysis enables the systematic examination of authentic written texts, while corpus analysis provides reliable evidence regarding the frequency and distribution of linguistic phenomena [3]. The analysis of learners' written errors provides a basis for identifying recurring patterns of language use and distinguishing systematic difficulties from isolated deviations [8].

The research was conducted in four consecutive stages:

- ① Compilation of a student writing corpus;
- ② Identification of orthographic errors;
- ③ Classification of phonology-based spelling errors;
- ④ Statistical analysis using IBM SPSS Statistics 29.0 [4].

This integrated approach allowed both linguistic and statistical examination of authentic orthographic usage.

This study was approved by the Research Ethics Committee of Zokhiomj University prior to data collection. The research involved the analysis of university students' written assignments for academic research purposes. All participants provided informed consent, and participation was voluntary. Personal identifiers were removed before corpus construction, and all data were anonymized to ensure confidentiality. The study was conducted in accordance with the ethical principles for research involving human participants.

Ethical Approval: Research Ethics Committee, Zokhiomj University.

2. RESEARCH MATERIALS

The research materials consisted of authentic academic texts produced by university students during regular coursework, including essays, reports, and free-writing assignments. The texts were digitized and combined into a single linguistic corpus without modifying their original lexical or grammatical forms.

The final corpus contained 10,000 running words. Duplicate word forms were retained in order to preserve natural language use and accurately represent the frequency of orthographic errors.

A total of 3,761 misspelled words were identified, of which 1,829 cases were classified as phonology-based orthographic errors. Consequently, the present study focuses exclusively on this phonological category of spelling errors.

Table 1. Characteristics of the Research Corpus

Variable	Value
Corpus size	10,000 words
Misspelled words	3,761
Error rate	37.6%
Phonology-based errors	1,829
Proportion of total errors	48.6%

Source: Author's corpus analysis (2026).

Orthographic errors were identified through a two-stage an

Stage 1. Microsoft Word Review Count

The Review function in Microsoft Word was used to detect preliminary spelling errors in each document. All identified misspelled words were extracted and compiled into an initial error list.

Stage 2. AntConc Corpus Analysis

The preliminary error list was subsequently analyzed using AntConc Version 4.2.0 [3]. The software was employed to:

- calculate word-frequency distributions;
- identify recurring error patterns;
- distinguish lexical variants sharing the same root;
- verify the classification of phonology-based orthographic errors.

Using AntConc increased the reliability of frequency analysis by examining orthographic patterns across the entire corpus rather than within individual documents.

For statistical analysis, the level of orthographic performance for each phonological rule was coded as an ordinal variable. Error frequency was categorized into five ordered levels.

Table 2. Coding Scheme for SPSS Analysis

Code	Error category	Measurement
0	No errors	Ordinal
1	1–3 errors	Ordinal
2	4–6 errors	Ordinal
3	More than 6 errors	Ordinal
4	Multiple errors	Ordinal

Source: Author's corpus analysis (2026).

Analysis of the use of phonology-based Mongolian orthographic rules in university students' writing

Because the number of observations differed across individual orthographic rules, statistical analyses were conducted separately for each rule rather than assuming equal sample sizes.

All quantitative analyses were performed using IBM SPSS Statistics Version 29.0 [4]. Descriptive statistics were employed to evaluate the application level of each phonology-based orthographic rule.

Table 3. Statistical Indicators Used in the Study

Indicator	Purpose
Frequency (n)	Number of observed errors
Percentage (%)	Relative proportion of errors
Mean (M)	Average error level
Standard Deviation (SD)	Variability of orthographic performance

Source: Author's corpus analysis (2026).

Higher mean values indicate greater difficulty in applying a particular orthographic rule, whereas standard deviation reflects the dispersion of performance across observed writing samples.

The study utilized authentic student writing exclusively for academic research purposes. All personal identifiers were removed prior to analysis, and the findings are presented only in aggregated statistical form. Therefore, the confidentiality and anonymity of participants were fully protected throughout the research process.

To evaluate the application level of individual phonology-based orthographic rules, spelling performance was classified into five ordinal categories: no errors, 1–3 errors, 4–6 errors, more than 6 errors, and multiple errors. Descriptive statistics were calculated using IBM SPSS Statistics 29.0.

3. RESULT ANALYSIS

3.1 Distribution of Orthographic Performance

Table 6. Distribution of Students' Performance Across Phonological Orthographic Rules

Orthographic rule	No errors	1–3	4–6	>6	Multiple
Masculine, feminine and neutral vowels	70	52	52	29	12
Long and diphthong vowels	76	63	42	25	9
Vowel harmony	39	59	63	25	16
Distinguishing -ii / -y	77	53	40	19	9
Vowel–sonorant consonants	49	55	55	25	15
Sonorant consonants	41	66	50	31	13
Distinguishing vowels	30	59	68	24	15
Spelling of j, ch, sh	73	58	41	23	8
Double and triple consonants	43	51	65	30	10
Distinguishing b / v	61	55	48	22	12
Elided reduced vowels	40	62	65	19	17
Non-elided reduced vowels	36	56	66	30	12
Separating signs (ᠪ, ᠬ)	50	52	62	22	13
Geminate consonants	41	52	76	25	8
Soft sign transformation	60	51	62	16	11
Soft sign unchanged	51	52	58	25	14
Foreign-word spelling	34	51	66	32	18

Source: Author's corpus analysis (2026).

Note. The frequencies represent the observed distribution of spelling performance for each orthographic rule. The number of observations varies across individual rules.

Table 6 indicates considerable variation in students' orthographic performance across phonological rules. The highest number of error-free performances was observed in the distinction between “-ii” and “-y” (77 cases), followed by long and diphthong vowels (76 cases) and the spelling of j, ch, sh (73 cases). These findings suggest that these orthographic rules are relatively well established in students' writing.

Conversely, the categories containing 4–6 errors were dominated by geminate consonants (76 cases), distinguishing vowels (68 cases), non-elided reduced vowels (66 cases), foreign-word spelling (66 cases), and double and triple consonants (65 cases). This pattern indicates that these phonological rules represent the greatest challenges in authentic written language use.

The ordinal error categories were further analyzed using descriptive statistics to determine the relative difficulty of each orthographic rule. Mean values represent the average error level, while standard deviation reflects the variability of spelling performance.

Table 7. Descriptive Statistics for Phonology-Based Orthographic Rules

Orthographic rule	Mean	SD
Foreign-word spelling	1.75	1.18
Distinguishing vowels	1.67	1.11
Non-elided reduced vowels	1.63	1.12
Vowel harmony	1.60	1.16
Double and triple consonants	1.56	1.13
Elided reduced vowels	1.56	1.16
Sonorant consonants	1.55	1.17
Geminate consonants	1.54	1.07
Vowel-sonorant consonants	1.51	1.21
Soft sign unchanged	1.50	1.20
Separating signs (ᠪ, ᠶ)	1.48	1.17
Masculine, feminine and neutral vowels	1.35	1.22
Distinguishing b / v	1.34	1.20
Soft sign transformation	1.33	1.15
Long and diphthong vowels	1.20	1.16
Spelling of j, ch, sh	1.19	1.16
Distinguishing -ii / -y	1.14	1.17

Source: Author's corpus analysis (2026).

Note. Higher mean values indicate greater difficulty in applying the corresponding orthographic rule.

The descriptive statistics reveal that foreign-word spelling demonstrated the highest level of orthographic difficulty ($M = 1.75$, $SD = 1.18$). This finding suggests that students experience considerable difficulty applying phonological spelling conventions to loanwords and foreign lexical items.

Similarly, distinguishing vowels ($M = 1.67$), non-elided reduced vowels ($M = 1.63$), and vowel harmony ($M = 1.60$) exhibited relatively high mean scores, indicating that these vowel-related phonological rules constitute the most problematic areas of Mongolian orthography.

In contrast, the distinction between “-ii” and “-y” ($M = 1.14$), the spelling of *j*, *ch*, *sh* ($M = 1.19$), and long and diphthong vowels ($M = 1.20$) showed the lowest mean error levels. These results suggest that students apply these orthographic conventions with comparatively greater consistency.

The descriptive statistical analysis demonstrates that phonology-based orthographic performance is not uniformly distributed across spelling rules. Foreign-word spelling, vowel distinction, reduced vowels, and vowel harmony represent the most difficult areas, whereas suffix distinction and long-vowel spelling show relatively stable orthographic performance. These findings provide the statistical foundation for the discussion of pedagogical implications presented in the following section.

4. DISCUSSION

The present study examined the application of phonology-based Mongolian orthographic rules through a corpus of authentic university students' writing. The findings demonstrate that phonological spelling errors constitute 48.6% of all orthographic errors, indicating that phonology remains the most problematic subsystem of Mongolian orthography in academic writing. This finding supports the argument that orthographic competence depends not only on grammatical knowledge but also on learners' ability to coordinate phonological representation with written language conventions [2].

The highest frequency of errors occurred in the elided reduced vowel rule, followed by the non-elided reduced vowel rule. Reduced vowels represent one of the most complex features of Mongolian orthography because their written form is determined by the interaction between pronunciation, syllable structure, and morphological composition rather than surface phonetic realization [1].

The corpus findings suggest that many students rely primarily on pronunciation when producing written texts, resulting in frequent omission or incorrect preservation of reduced vowels. This tendency is consistent with Sanjaa's [2] description of reduced vowels as the least stable component of Mongolian spelling acquisition. Therefore, instructional practice should emphasize morphological analysis alongside phonological awareness instead of memorization of isolated spelling rules.

Descriptive statistics revealed relatively high mean error scores for distinguishing vowels ($M = 1.67$) and vowel harmony ($M = 1.60$). These results indicate that students experience persistent difficulty applying vowel-related orthographic principles during authentic writing.

Vowel harmony requires writers to recognize internal phonological relationships among syllables, whereas distinguishing vowels depends upon lexical and morphological knowledge [1]. Consequently, these rules require analytical linguistic processing rather than simple rule recall. Similar patterns have been reported by Davaasuren [9], who observed that vowel harmony errors remain frequent even after learners have mastered fundamental spelling conventions. The present study extends those findings by demonstrating that the same difficulties persist at the university level.

The transformation of the soft sign (*ᠨ*) and the spelling of double and triple consonants also exhibited substantial error frequencies. Unlike reduced vowels, these rules involve the relationship between phonological alternation and consonant representation. Students frequently applied pronunciation-based spelling strategies, particularly when consonant clusters occurred across morpheme boundaries.

These observations suggest that phonological instruction should incorporate authentic error analysis and corpus-based revision activities. Corpus linguistics enables learners to observe recurring spelling patterns within real texts, thereby promoting evidence-based orthographic awareness [3].

The integration of Word Review Count, AntConc, and SPSS provides a methodological framework for evaluating orthographic performance beyond traditional classroom assessment. Rather than identifying isolated mistakes, corpus-based analysis allows educators to determine systematic patterns of spelling difficulty and prioritize instructional content accordingly.

Based on the present findings, reduced vowels, distinguishing vowels, soft-sign transformation, and foreign-word spelling should receive greater emphasis within university-level academic writing courses. Developing learning materials from authentic student errors may improve both phonological awareness and practical orthographic competence.

Overall, the findings indicate that phonology-based spelling errors are primarily associated with the interaction between pronunciation, morphology, and orthographic representation. Consequently, improving Mongolian orthographic competence requires pedagogical approaches grounded in corpus evidence, phonological analysis, and authentic writing practice rather than rule memorization alone.

5. CONCLUSION

This study investigated the use of phonology-based Mongolian orthographic rules by analyzing a 10,000-word corpus compiled from authentic university students' writing. Through corpus analysis and descriptive statistics, the structural characteristics and application levels of phonological spelling errors were identified.

The findings lead to four principal conclusions.

First, phonology-based orthographic errors accounted for 1,829 cases (48.6%) of the 3,761 misspelled words, demonstrating that phonological spelling constitutes the largest category of orthographic errors in students' academic writing.

Second, the most frequent error patterns were associated with elided reduced vowels, the distinction between “-ii” and “-y”, non-elided reduced vowels, and soft-sign transformation. These rules represent the most problematic areas of Mongolian orthography because they require writers to coordinate pronunciation, syllable structure, and morphological composition.

Third, descriptive statistical analysis revealed that foreign-word spelling, distinguishing vowels, non-elided reduced vowels, and vowel harmony exhibited the highest mean error scores, indicating greater orthographic difficulty than other phonological rules.

Finally, the integration of corpus linguistics, AntConc, Microsoft Word Review Count, and SPSS proved to be an effective methodological framework for evaluating authentic orthographic performance. The results provide empirical evidence for improving Mongolian orthography instruction and academic writing pedagogy in higher education.

5.1 Limitations

Several limitations should be acknowledged.

First, the study focused exclusively on phonology-based orthographic rules and did not examine errors related to morphology, syntax, or stylistics.

Second, the statistical analyses were conducted using aggregated frequency data for individual orthographic rules. Consequently, the findings describe overall patterns of orthographic performance rather than individual learner differences.

Third, the corpus consisted solely of academic writing produced in university coursework. Therefore, the findings should not be generalized directly to other genres of written Mongolian without further investigation.

5.2 Pedagogical Implications

Based on the findings, several recommendations are proposed for Mongolian language education.

- ① Greater instructional emphasis should be placed on reduced vowels, distinguishing vowels, and soft-sign rules, as these constitute the most frequent sources of spelling difficulty.

- ② Orthographic instruction should incorporate corpus-based error analysis rather than relying exclusively on rule memorization.
- ③ Digital linguistic tools such as AntConc and Microsoft Word Review may be integrated into formative assessment to identify systematic spelling patterns in students' writing.
- ④ Future research should expand the corpus to include morphological, grammatical, and stylistic errors in order to develop a comprehensive model of Mongolian academic writing proficiency.

REFERENCES

- [1] Ministry of Education and Science of Mongolia. (2018). Mongolian Orthographic Rules. Ulaanbaatar, Mongolia.
- [2] Sanjaa, J. (2008). Mongol khelnii avia zui [Phonology of the Mongolian Language]. Ulaanbaatar, Mongolia.
- [3] Anthony, L. (2022). AntConc (Version 4.2.0) [Computer software]. Waseda University. <https://www.laurenceanthony.net/software/antconc/>
- [4] IBM Corp. (2023). IBM SPSS Statistics for Windows (Version 29.0). <https://www.ibm.com/products/spss-statistics>
- [5] Kasyanenko, K. K. (2002). Sovremenniy mongol'skii yazyk: Uchebnoe posobie [Modern Mongolian Language: Textbook]. Moscow, Russia.
- [6] Polivanov, E. D. (2006). Fonetika i orfografiya [Phonetics and Orthography]. Moscow, Russia.
- [7] Beshenkova, E., & Ivanova, O. (2018). Aspektnoe opisanie russkoi orfografii [Aspectual Description of Russian Orthography]. Moscow, Russia.
- [8] Corder, S. P. (1981). Error Analysis and Interlanguage. Oxford University Press.
- [9] Davaasuren, N. (2021). Integrated methodological study, experiment, and results of Mongolian Cyrillic orthographic rules [Doctoral dissertation]. Ulaanbaatar, Mongolia.
- [10] Baatarjav, M., Gombo, B., & TAONUOEN. (2025). Evaluation of 8th grade students' engagement in literature lessons in secondary education. International Journal of Social Science and Humanities Research-MIYR, 5(4), 15–24. <https://doi.org/10.53468/mifyr.2025.5.4.15>

AUTHOR'S INTRODUCTION

1. First Author and Corresponding Author

	Lkhamjav Myadag	myadag779@gmail.com
	L. Myadag is a lecturer and researcher at Zokhiomj University, Ulaanbaatar, Mongolia. Her research interests include Mongolian orthography, phonology, corpus linguistics, and academic writing. She conducted the conceptualization, corpus analysis, statistical analysis, and manuscript preparation for this study.	

2. Co- Author

	Zugderdemid Nyamjargal	z.nyamjargal0728@gmail.com
	Z. Nyamjargal is a lecturer at Nalaikh Polytechnic College and a doctoral candidate at Graduate University of Mongolia. His research focuses on Educational Studies, particularly language education and educational research. He contributed to the research design, data interpretation, and manuscript review.	