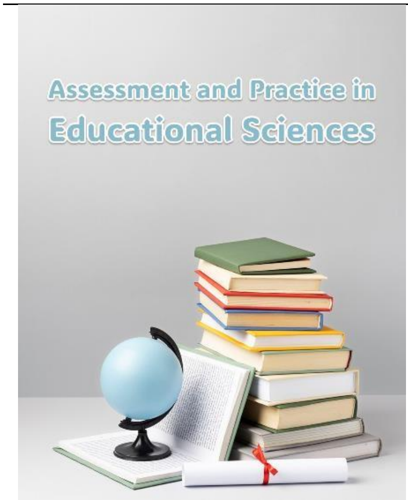
Assessment and Practice in Educational Sciences



Cohesion-based Errors and Intelligibility of Persian-English Machine Translation Output

1. Sadegh. Alizadeh Tehrani : Department of English Translation Studies, Faculty of Persian literature and Foreign Languages, University of Allameh Tabataba'i, Tehran, Iran (Email: sadeghtehrani1017@gmail.com)

Article type:
Original Research

Article history:
Received 29 October 2025
Revised 12 February 2026
Accepted 16 February 2026
Initial Publish 24 June 2026
Published online 01 July 2026

How to cite this article:
Alizadeh Tehrani, S. (2026). Cohesion-based Errors and Intelligibility of Persian-English Machine Translation Output. *Assessment and Practice in Educational Sciences,* 4(4), 1-12. https://doi.org/10.61838/japes.231

ABSTRACT

Basically, nowadays, such translation-based computer technologies as machine translation systems are employed so as to increase the speed and achieve far higher quality in translation due to the high volume of translation projects and limited time periods. The main goal of this research is to identify the cohesion- based errors committed in machine translation, to evaluate their impact upon the chains of the cohesive devices and also upon the intelligibility of machine translated outputs. In this research, the data was collected from three different machine translation systems, followed by 30 master students of TEFL as the participants in this research were employed in order to back translate a selected quantity of machine translated texts so as to evaluate the quality and the level of intelligibility. Based on the obtained results, it was determined that Google machine translation system assigned the first position due to its better quality and the less proportion of cohesion-based errors committed. On the other hand, Abadis, with a relatively small distinction compared to Google's performance in terms of the quality and the statistic of committed errors, allocated the second position, which confirmed its far better performance than Bing. In addition, regarding the cohesion-based errors affecting the intelligibility of the machine translated texts, it could be argued that among the six types of cohesion-based errors identified, the missing-word error and the non-translated word error were taken into account with the highest and the lowest impact upon the participants' intelligibility of the machine translated texts.

Keywords: Machine translation, Translation engines, Translation errors, Intelligibility, Technology

Introduction

The accelerating globalization of knowledge production, commerce, education, and digital communication has intensified the demand for rapid and reliable cross-linguistic transfer. In response to this demand, machine translation (MT) has evolved from a peripheral computational experiment into a central technological infrastructure embedded in everyday communication practices. Early MT systems were predominantly rule-based, relying on hand-crafted grammatical rules and bilingual lexicons to generate target-language output, yet their scalability and adaptability were limited (1). Over time, the field transitioned toward statistical paradigms, which modeled translation as a probabilistic mapping between aligned corpora, and subsequently toward neural machine translation (NMT), which leverages deep learning architectures to capture complex linguistic patterns (2). The introduction of large-scale neural frameworks, exemplified by Google's neural MT system, marked a turning point by significantly reducing lexical and syntactic errors compared to phrase-based systems (3). More recent evaluations of Google

Translate across multiple language pairs confirm that its performance has steadily improved between 2012 and 2025, particularly in high-resource languages (4).

Despite these advancements, translation quality remains uneven, particularly in linguistically distant language pairs such as Persian and English. Structural asymmetries, including differences in word order, pro-drop phenomena, morphological richness, and discourse organization, create persistent challenges for MT systems (5). Although neural models have improved fluency at the sentence level, they frequently struggle to maintain discourse coherence across sentences, leading to inconsistencies in lexical choice, pronoun reference, and logical connectors (6). These limitations are exacerbated in contexts where MT output is consumed without extensive human post-editing, such as in educational or informal communication settings (7). As MT becomes increasingly integrated into EFL learning environments, students' reliance on automated systems raises questions not only about accuracy but also about textual intelligibility and cohesion (8).

A central dimension of translation quality is intelligibility, defined as the extent to which a target text can be understood by its intended audience. Intelligibility is not reducible to lexical correctness alone; rather, it is intrinsically linked to the cohesive fabric that binds textual elements into a meaningful whole (9). The theoretical foundation for analyzing textual cohesion was established by Halliday and Hasan, who conceptualized cohesion as the network of grammatical and lexical ties—reference, substitution, ellipsis, conjunction, and lexical relations—that create textual texture (10). When these ties are disrupted, the reader's ability to construct coherence is compromised, even if individual sentences appear grammatically acceptable. In MT output, cohesion-based errors may manifest as missing referents, inconsistent lexical chains, inappropriate conjunctions, or faulty morphological agreement, all of which weaken the semantic continuity of discourse (11).

Empirical studies have repeatedly demonstrated that cohesion-related deficiencies are a major source of incomprehensibility in machine-translated texts. Research on Polish–English MT output revealed that breakdowns in reference chains and lexical repetition significantly reduced comprehensibility, even when local sentence accuracy was relatively high (12). Similarly, investigations into Swedish–English MT found that errors in cohesive devices were directly associated with reader confusion and misinterpretation (11). In the Persian–English context, comparative analyses of Google and Bing translators identified systematic lexical and structural inconsistencies, particularly in specialized academic texts (13). These findings align with broader evaluations comparing human and machine translation, which indicate that MT often achieves surface-level fluency while failing to replicate the cohesive density characteristic of human-produced discourse (14).

The persistence of cohesion-based errors reflects deeper architectural constraints in MT systems. Although neural models encode contextual information within distributed representations, they typically process input in limited segments, thereby underrepresenting discourse-level dependencies (15). Pronoun resolution remains particularly problematic, as MT systems may not accurately track antecedents across sentence boundaries (16). In Persian, where subject pronouns are frequently omitted and encoded morphologically within verb inflections, translation into English requires explicit reconstruction of reference chains; failure to perform this reconstruction results in missing-word or ambiguous-reference errors. Such disruptions undermine not only grammatical cohesion but also semantic coherence.

In addition to reference, lexical cohesion constitutes a critical dimension of textual integrity. Consistent repetition, synonymy, hyponymy, and antonymy create semantic networks that guide reader interpretation (10). MT systems sometimes generate inconsistent lexical choices for repeated source terms, thereby fragmenting lexical chains and increasing cognitive load. Bowker's corpus-based approach to translation evaluation underscores the importance of systematically identifying such patterns through empirical analysis rather than relying solely on impressionistic judgments (17). Corpus-informed methodologies enable researchers to quantify recurring error types and correlate them with measurable declines in intelligibility.

Another relevant dimension concerns translation errors that extend beyond cohesion yet indirectly affect it. Studies of machine-translated proficiency tests have demonstrated that missing-word errors and inaccurate lexical substitutions disproportionately impair comprehension compared to minor morphological deviations (18). Similarly, research on unsupervised representation learning in speech and language modeling suggests that contextual encoding remains imperfect when semantic dependencies are long-range or implicit (19). These findings indicate that while neural architectures capture statistical regularities, they do not inherently guarantee discourse-level coherence.

The growing prevalence of MT in professional and educational domains has intensified interest in post-editing and quality evaluation frameworks. Abdi's evaluation of the Abadis translator highlights the necessity of systematic quality assessment, particularly in relation to post-editing efficiency (20). Pshenichnikov identifies post-editing as an indispensable professional task, emphasizing that even advanced MT outputs require human intervention to correct semantic and cohesive inconsistencies (21). However, reliance on post-editing presupposes that errors can be reliably detected and categorized, underscoring the need for fine-grained taxonomies of cohesion-based errors.

Methodologically, the complexity of evaluating both quantitative error frequency and qualitative intelligibility suggests the suitability of mixed-methods research designs. Creswell's reconceptualization of mixed-method frameworks underscores their capacity to integrate statistical measurement with interpretive analysis (22). Such integration is particularly appropriate in MT research, where numerical error counts must be contextualized within discourse-level interpretations of coherence.

Within the Persian–English translation domain, relatively limited research has systematically examined cohesion-based errors as a distinct analytical category. While Mirzaeian explored editing techniques in MT-informed academic writing (5), and Saffari compared system performance across academic fields (13), comprehensive empirical linkage between specific cohesion errors and intelligibility outcomes remains underdeveloped. Furthermore, as MT becomes embedded in EFL contexts, learners' strategic use of these tools introduces additional variables related to machine translation literacy (8). Deng's investigation of Chinese undergraduates demonstrates that users' perceptions of MT reliability may not align with objective quality measures, particularly regarding discourse coherence (7).

Taken together, the literature reveals three converging insights. First, MT technology has achieved remarkable progress at the sentence level, particularly through neural architectures (2, 3). Second, cohesion and intelligibility remain vulnerable dimensions, especially in linguistically distant language pairs and discourse-level contexts (6, 16). Third, systematic, corpus-based, mixed-method evaluations are necessary to identify how specific error types disrupt cohesive chains and impede comprehension (17, 22).

In light of these theoretical and empirical considerations, the present study seeks to conduct a comprehensive analysis of cohesion-based errors in Persian–English machine translation outputs generated by major online MT systems, to examine how these errors disrupt cohesive chains as conceptualized by Halliday and Hasan, and to empirically assess their differential impact on reader intelligibility.

Methods and Materials

This study employs a mixed-methods convergent parallel design (22) to comprehensively investigate the relationship between cohesion-based errors and intelligibility in Persian-English machine translation (MT). Both qualitative and quantitative data strands were collected concurrently but analyzed independently, with integration occurring during the final interpretation phase. This approach allows for a nuanced exploration of both the frequency of specific error types and their practical impact on understanding.

A purposive sample of 30 MA graduates in English Language Teaching from Mazandaran State University participated in the study. Comprising 18 females and 12 males, all participants possessed high-intermediate to advanced English proficiency, with many working as English instructors. This specific profile ensured participants had the necessary linguistic competence to reliably assess the intelligibility of translated texts. Participants were recruited via social media and email, focusing on individuals capable of evaluating comprehensibility rather than translation techniques, aligning with the study's core objective.

The primary instrument was a researcher-constructed translation intelligibility test. It featured 24 sentences selected from a corpus of MT outputs, each exemplifying one of five predefined cohesion-based error types. Participants performed back-translation of these English sentences into Persian. The source material consisted of three non-literary, general-topic news articles (totaling ~2500 words) from the IMNA News Agency, covering themes like public health and social issues. These texts were processed through three MT systems: Google Translate, Bing Translator, and the domestic Abadis Translator, enabling a comparative analysis of error profiles across platforms.

Data collection proceeded in three stages. First, the Persian source texts were fed into the three MT systems to generate nine English outputs. Second, a qualitative analysis identified and categorized cohesion errors (reference, conjunction, lexical cohesion) within these outputs using Halliday and Hasan's (1976) model. Third, the intelligibility test was administered: participants back-translated the selected error-containing sentences, and their renditions were compared to the original Persian. This process directly linked specific error types to measurable outcomes in comprehension.

Analysis combined quantitative and qualitative methods. Descriptive statistics (frequencies, percentages) summarized the distribution of error types across MT systems and participant back-translation accuracy. Concurrently, qualitative content analysis, guided by Halliday and Hasan’s (1976) framework, provided a detailed classification of error manifestations and their disruptive effects on cohesive chains (10). The two analytical strands were finally merged to interpret how quantitative error patterns correlate with qualitative breakdowns in coherence, offering a holistic answer to how cohesion errors impact intelligibility.

Findings and Results

Findings for Research Question 1: Performance of MT Systems in Light of Cohesion-Based Errors

The first research question sought to evaluate how the three MT systems—Google Translate, Bing Translator, and Abadis Translator—perform when exposed to Persian source texts prone to cohesion-based errors in translation. To address this, three Persian texts on distinct topics ("Adolescents' Health," "Unemployment," and "Drug Addiction") were processed through each system. The resulting English outputs were analyzed based on Halliday and Hasan's (1976) cohesion framework, leading to the identification of errors into six categories: Missing Word, Extra Word, Word Order, Incorrect Word, Incorrect Word Form, and Non-translated Word. A qualitative review of the outputs revealed the nature of these errors. For instance, Table 1. provides excerpts from Google Translate's output for Text 1, illustrating various error types such as the omission of key terms (*Missing Word*), erroneous additions (*Extra Word*), and syntactical disarray (*Word Order*).

Table 1. Excerpts of Cohesion-based Errors by Google Translate in Text 1 (Adolescents' Health)

Source Text (ST)	Machine Translation (TT)	Error Type
مدیر بخش بهبود تغذیه خدمات بهداشتی درمانی دانشگاه علوم پزشکی اصفهان	Director of Nutrition Improvement Department of Isfahan University of Medical Sciences.	Missing Word
سبک زندگی بیتحرک و عادات غذایی غلط یکی از علل مهم ...چاقی	Sedentary and wrong lifestyle and eating habits are one of the important causes of obesity...	Word Order
...باید گفت که اضافه وزن یک نگرانی جدی پزشکی است	It should be said that overweight is a serious medical concern that can strongly affect...	Incorrect Word
دلایل گرسنگی و سوءتغذیه نوجوانان	Reasons for starvation and su nutrition in adolescents	Non-translated Word

(Note: This is an illustrative excerpt. The full analysis generated 391 error instances.)

To quantify performance, the frequency of each error type per system was calculated. The aggregated results, presented in Tables 2, 3, and 4, reveal the distinct error profiles of each system.

Table 2. Frequency and Percentage of Cohesion-Based Errors by Google Translate

Error Type	Total	Percentage	Text 1	Text 2	Text 3
Missing Word	25	27.47%	7.69%	9.88%	9.88%
Incorrect Word	18	19.78%	6.59%	7.69%	5.49%
Extra Word	19	20.87%	5.49%	9.88%	5.49%
Word Order	14	15.38%	5.49%	6.59%	3.29%
Incorrect Word Form	9	9.89%	2.19%	6.59%	1.09%
Non-translated Word	6	6.59%	2.19%	2.19%	2.19%
Total	91	100%	29.64%	42.82%	27.43%

Table 3. Frequency and Percentage of Cohesion-Based Errors by Bing Translator

Error Type	Total	Percentage	Text 1	Text 2	Text 3
Word Order	46	28.04%	8.53%	14.02%	5.48%
Missing Word	36	21.95%	9.14%	5.48%	7.31%
Incorrect Word	30	18.29%	4.87%	9.14%	4.26%
Extra Word	24	14.63%	6.09%	4.26%	4.26%
Incorrect Word Form	16	9.75%	3.04%	4.87%	1.82%
Non-translated Word	12	7.31%	1.82%	2.43%	3.04%
Total	164	100%	33.49%	40.20%	26.17%

Table 4. Frequency and Percentage of Cohesion-Based Errors by Abadis Translator

Error Type	Total	Percentage	Text 1	Text 2	Text 3
Missing Word	53	36.30%	10.27%	15.06%	10.95%
Extra Word	25	17.12%	4.10%	6.84%	6.16%
Word Order	23	15.75%	4.79%	6.16%	4.79%
Incorrect Word	18	12.32%	4.79%	4.10%	3.42%
Non-translated Word	15	10.27%	2.73%	4.79%	2.73%
Incorrect Word Form	12	8.21%	2.05%	4.78%	1.36%
Total	146	100%	28.73%	41.73%	29.41%

The data shows that Google Translate committed the fewest total errors (91), with its most prevalent issue being Missing Words (27.47%). Bing Translator produced the highest error count (164), dominated by Word Order errors (28.04%). Abadis Translator (146 total errors) struggled most acutely with Missing Words (36.30%).

A cross-system comparison of total errors per text consistently ranked Google Translate as the best-performing system, committing the fewest errors in all three texts. Finally, an aggregate analysis of all 391 errors across the nine outputs identified Missing Word errors as the most frequent challenge (114 instances, 29.15%), while Non-translated Word errors were the least common (33 instances, 8.43%).

In summary, in response to RQ1, the systems performed with varying degrees of success. Google Translate demonstrated superior overall performance with the lowest error frequency. However, all systems exhibited significant and systematic weaknesses, particularly in preserving complete lexical content (Missing Word) and maintaining correct syntactic structure (Word Order), confirming their vulnerability to cohesion-based errors when translating Persian texts.

Findings for Research Question 2: Impact of Cohesion Errors on Chains of Cohesive Devices

The second research question investigated the extent to which various cohesion-based error types disrupt the chains of cohesive devices within the machine-translated English outputs. To analyze this, the cohesive devices in the target texts were identified and categorized based on Halliday and Hasan's (1976) model into main tiers (Semantic, Reference, Conjunction) and their sub-tiers (e.g., Synonym, Personal, Additive). The frequency of these devices was then quantified to establish a baseline for cohesive "chains" (Table 5. provides an excerpt of such devices identified in Google's output for Text 1).

Table 5. Excerpt of Cohesive Devices Identified in Google Translate's Output (Text 1)

Sub-tier Category	Example from Target Text
Semantic: Synonym	…in order to maintain the health and well-being of teenagers...
Semantic: Hypernym/Hyponym	…the importance of healthy nutrition (hyponym) to prevent obesity (hyponym)...
Conjunction: Additive	In addition, people undergo many physical changes...
Conjunction: Causal	Therefore, this sedentary lifestyle is one of the important causes...
Reference: Personal	…the habits… They can even play a more effective role...
Reference: Comparative	…a more serious impact… compared to other cases.

The subsequent quantitative analysis aimed to determine if and how cohesion errors broke these established chains. The total quantities and percentages of cohesive devices used across all systems for each text are presented in Tables 6, 7 and 8.

Table 6. Frequency and Percentage of Cohesive Devices in Text 1 Outputs

Sub-tier	Google	Bing	Abadis	Avg. %
Semantic				
Synonym	25 (47.16%)	25 (51.8%)	17 (32.7%)	43.88%
Hyponym	14 (26.41%)	12 (24.1%)	24 (47.15%)	32.22%
Hypernym	14 (26.41%)	12 (24.1%)	10 (20.15%)	23.55%
Conjunction				
Additive	25 (54.4%)	24 (58.4%)	23 (61.4%)	58.06%
Temporal	12 (25.8%)	9 (21.9%)	8 (20.6%)	22.77%
Causal	9 (19.8%)	8 (19.7%)	5 (12.8%)	17.43%
Adversative	0 (0%)	0 (0%)	2 (5.2%)	1.73%
Reference				
Personal	21 (70.5%)	24 (44.5%)	23 (63.6%)	59.53%
Comparative	6 (19.3%)	19 (34.3%)	6 (16.3%)	23.30%
Demonstrative	4 (10.2%)	11 (21.2%)	7 (20.1%)	17.16%

Table 7. Frequency and Percentage of Cohesive Devices in Text 2 Outputs

Sub-tier	Google	Bing	Abadis	Avg. %
Semantic				
Synonym	25 (53.1%)	25 (54.8%)	22 (34.3%)	47.40%
Antonym	9 (20.96%)	10 (23.2%)	21 (31.7%)	25.29%
Hyponym	7 (16.4%)	4 (9.5%)	11 (17%)	14.30%
Hypernym	4 (9.54%)	4 (9.5%)	11 (17%)	12.01%
Conjunction				
Additive	17 (77%)	21 (69.7%)	24 (55.8%)	67.50%
Adversative	3 (15.3%)	5 (17.2%)	11 (26.2%)	19.57%
Temporal	2 (7.7%)	3 (9%)	5 (12.7%)	9.80%
Causal	0 (0%)	2 (4.1%)	3 (5.3%)	3.13%
Reference				
Personal	24 (44.2%)	26 (47.6%)	23 (64.3%)	52.03%
Demonstrative	17 (31.4%)	20 (38.4%)	3 (9.7%)	26.50%
Comparative	13 (24.4%)	8 (14%)	9 (26.2%)	21.53%

Table 8. Frequency and Percentage of Cohesive Devices in Text 3 Outputs

Sub-tier	Google	Bing	Abadis	Avg. %
Semantic				
Synonym	22 (32.2%)	22 (34.2%)	23 (36%)	34.13%
Antonym	21 (30.8%)	16 (25%)	18 (29%)	28.27%
Hyponym	15 (22.4%)	16 (24.4%)	14 (22%)	22.93%
Hypernym	10 (14.6%)	11 (16.4%)	8 (13%)	14.66%
Conjunction				
Additive	23 (65.3%)	19 (73.4%)	24 (61.5%)	66.73%
Temporal	8 (22.8%)	3 (10.1%)	9 (23%)	18.63%
Adversative	5 (14.5%)	4 (16.5%)	6 (15.5%)	15.50%
Causal	0 (0%)	0 (0%)	0 (0%)	0.00%
Reference				
Demonstrative	37 (49.5%)	25 (45.5%)	24 (40.5%)	45.16%
Personal	19 (25.5%)	15 (27%)	22 (36.5%)	29.66%
Comparative	18 (25%)	15 (27.5%)	19 (23%)	25.16%

The impact of cohesion errors was then traced by mapping specific error types onto disruptions in these cohesive chains. The analysis revealed a direct and detrimental relationship:

In Text 1, Missing Word errors in outputs from Google and Abadis broke the chains of Personal reference (e.g., ambiguous pronoun reference). Word Order errors in Bing's output disrupted Causal conjunction chains, weakening logical connectivity.

In Text 2, Missing Word errors were again the primary cause of broken chains, affecting Additive and Personal devices in Google's output and Personal devices in Bing's. In Abadis's translation, Missing Word and Word Order errors broke chains of Personal reference and Hyponym semantic relations, respectively.

In Text 3, the pattern continued: Missing Word errors in Google's output broke Additive conjunction chains. Word Order errors in Bing's output disrupted Personal reference chains. Abadis's output showed broken chains in Personal and Demonstrative references due to Missing Word errors.

Illustrative Example (Text 1, Google Translate):

Source: افراد در عنفوان نوجوانی دچار تغییرات فیزیکی زیادی میشوند. بنابراین، برخورداری از یک رژیم غذایی مناسب... حیاتی است، زیرا بدن به تنوعی از ریزمغذیها... نیاز دارد.

Translation: People experience numerous physical changes during adolescence. Proper and efficient diet should be paid more attention to as a vital factor, because they need to receive a variety of micronutrients... Therefore...

Analysis: A Missing Word error (omission of a clear subject like "teenagers" or "they" linked to "برخورداری") breaks the Personal reference chain. The referent for "they" becomes ambiguous, and the subsequent "Therefore" (Causal conjunction) loses its logical anchor, damaging the cohesive chain.

In summary, in direct response to RQ2, the chains of cohesive devices were significantly impacted by cohesion-based errors. Missing Word and Word Order errors were the most disruptive, consistently breaking chains across all systems, particularly affecting Personal and Demonstrative reference devices and Additive and Causal conjunctions. This breakdown in the surface-level ties that create texture in a text directly undermines its coherence.

Findings for Research Question 3: Impact of Cohesion Errors on Readers' Intelligibility

The third research question sought to assess the direct impact of the six identified cohesion-based error types on the intelligibility of the machine-translated English outputs for human readers. To measure this, a back-translation test was administered. Thirty advanced-level TEFL M.A. students were asked to translate 24 selected English sentences (extracted from the nine MT outputs) back into Persian. These sentences contained five instances of each error type, resulting in 150 test items per error category (30 participants× 5 sentences). The intelligibility was operationalized as the ability to correctly back-translate the meaning of the original Persian source text. The percentages of correct and incorrect back-translations for each error type are presented in Table 9.

Table 9. Impact of Cohesion Error Types on Intelligibility (Back-Translation Test)

Error Type	Correct Back-Translation	Incorrect Back-Translation
Missing Word	6 (4%)	144 (96%)
Extra Word	20 (13.33%)	130 (86.66%)
Incorrect Word	23 (15.33%)	127 (84.66%)
Word Order	99 (66%)	51 (34%)
Incorrect Word Form	129 (86%)	21 (14%)
Non-translated Word	135 (90%)	15 (10%)

Note: N = 150 test items per error type.

The results reveal a stark hierarchy in the disruptive potential of different error types on intelligibility:

Most Disruptive: Missing Word errors had the most severe impact, with a 96% incorrect back-translation rate. This indicates that the omission of critical lexical content catastrophically hinders comprehension, as readers cannot recover the missing meaning from context.

Highly Disruptive: Extra Word and Incorrect Word errors also severely compromised intelligibility, with incorrect back-translation rates of 86.66% and 84.66%, respectively. The introduction of spurious content or semantically inaccurate terms significantly distorts the intended message.

Moderately Disruptive: Word Order errors showed a moderate impact (34% incorrect). While syntactic disarray poses a challenge, readers could often infer the intended meaning through lexical cues.

Least Disruptive: Incorrect Word Form (14% incorrect) and Non-translated Word (10% incorrect) errors had the smallest impact. For advanced learners, grammatical form errors (e.g., wrong part of speech) and untranslated terms (often left as conspicuous markers like "su") were more likely to be identified as "noise" or guessed correctly from the surrounding context, without completely blocking comprehension of the propositional content.

In summary, in direct response to RQ3, cohesion-based errors have a profound but varied impact on intelligibility. Semantic and content-level errors (Missing Word, Incorrect Word) are far more detrimental to comprehension than formal or syntactic-level errors (Word Order, Incorrect Word Form). The Missing Word error is the single most critical flaw, rendering texts largely unintelligible.

Discussion and Conclusion

The findings of this study provide a detailed empirical account of how cohesion-based errors manifest in Persian–English machine translation (MT) outputs and how these errors directly influence textual intelligibility. In response to the first research question, the comparative analysis of Google Translate, Bing Translator, and Abadis Translator demonstrated clear performance differentiation, with Google producing the lowest total number of cohesion-based errors, followed by Abadis and then Bing. This ranking is consistent with broader evaluations of MT progress indicating that large-scale neural systems trained on expansive corpora generally outperform smaller or less resourced systems (2). The superior performance of Google may be interpreted in light of the neural architecture introduced in its NMT framework, which was explicitly designed to reduce structural and lexical inaccuracies compared to earlier phrase-based systems (3). Longitudinal analyses of Google Translate's development across language pairs also confirm sustained improvement in fluency and lexical selection over time (4).

However, despite this relative superiority, the results indicate that even the best-performing system remains vulnerable to cohesion-related deficiencies. Missing-word errors were the most frequent across all systems, suggesting that content omission is a persistent weakness in Persian–English MT. This pattern aligns with observations that neural systems, while fluent, may occasionally under-generate lexical elements, particularly when implicit grammatical information in the source language must be made explicit in the target language (6). Persian's pro-drop characteristics and morphological encoding of subject agreement create conditions in which explicit English pronouns must be reconstructed; failure to do so results in broken reference chains. Similar cohesion-related challenges have been documented in other language pairs, where MT systems struggle to maintain cross-sentential referential continuity (16).

The prominence of missing-word errors also corresponds with earlier research on cohesion and comprehensibility. Studies examining Swedish–English MT found that omitted lexical items frequently disrupted textual texture, particularly within reference and conjunction chains (11). Likewise, analyses of Polish–English MT demonstrated that omissions within lexical or referential networks were strongly associated with reduced reader comprehension (12). The present findings reinforce this pattern within the Persian–English domain, indicating that omission errors fracture cohesive chains at both semantic and grammatical levels.

The second research question explored how cohesion-based errors affect the chains of cohesive devices. The mapping of error types onto Halliday and Hasan's cohesion model revealed that missing-word and word-order errors most consistently disrupted personal and demonstrative reference, as well as additive and causal conjunction chains (10). These disruptions weakened the logical progression of discourse and increased cognitive processing demands on readers. Such results corroborate discourse-oriented critiques of mainstream MT architectures, which argue that sentence-level optimization fails to adequately model document-level dependencies (15). When cohesive devices are mistranslated or omitted, the resulting text may appear grammatically intact at the sentence level but lack global coherence.

The observed disruption of lexical chains further supports corpus-based perspectives on translation evaluation. Bowker emphasizes that systematic analysis of recurring lexical patterns can reveal weaknesses invisible to surface fluency assessments (17). In this study, inconsistent lexical choices and inappropriate substitutions fragmented semantic relations such as synonymy and hyponymy, undermining semantic continuity. This phenomenon mirrors findings in comparative analyses of Google and human translation across scientific domains, where MT frequently demonstrated weaker lexical cohesion despite acceptable grammatical accuracy (14).

The third research question addressed the measurable impact of cohesion-based errors on intelligibility through a back-translation test. The hierarchy of impact revealed that missing-word errors were the most detrimental, followed by extra-word and incorrect-word errors, whereas incorrect-word-form and non-translated-word errors exerted comparatively limited influence on comprehension. This stratification aligns with findings from machine-translated proficiency test research, where lexical omission and inaccurate substitution were identified as primary contributors to incomprehensibility (18). Similarly, experimental work on language modeling indicates that disruptions in semantic continuity exert stronger effects on comprehension than purely morphological deviations (19).

The relatively limited impact of incorrect-word-form errors suggests that advanced readers can often reconstruct intended meaning despite minor grammatical inaccuracies. This observation resonates with research on intelligibility in human and machine translation, which found that surface grammatical errors do not necessarily impede understanding if semantic relations remain intact (9). Conversely, lexical omission removes semantic anchors entirely, leaving readers without sufficient cues to infer meaning. This distinction highlights the central role of semantic cohesion in sustaining intelligibility.

The findings also have implications for the pedagogical use of MT. Research on MT literacy among EFL learners demonstrates that students frequently rely on machine outputs without critically evaluating coherence or cohesion (8). Deng's investigation into Chinese undergraduates similarly reports that learners often overestimate the reliability of MT systems, particularly regarding discourse-level accuracy (7). The present results suggest that without explicit awareness of cohesion-based vulnerabilities, learners may accept outputs that are locally fluent yet globally incoherent.

From a professional perspective, the necessity of post-editing remains evident. Abdi's evaluation of Abadis Translator underscores that MT outputs require systematic post-editing to address lexical and structural inaccuracies (20). Pshenichnikov further identifies cohesion-related inconsistencies as key challenges in post-editing practice (21). The empirical demonstration that specific error types disproportionately impair intelligibility reinforces the importance of targeted post-editing strategies focused on reference resolution, lexical consistency, and connective accuracy.

Methodologically, the mixed-methods approach adopted in this study enabled integration of quantitative error frequencies with qualitative discourse analysis. Such methodological triangulation corresponds with contemporary frameworks advocating convergent designs for complex linguistic phenomena (22). By combining statistical distribution with back-translation intelligibility measures, the study moves beyond generic accuracy metrics toward a discourse-sensitive evaluation model.

Overall, the results confirm that while neural MT systems have achieved substantial progress, cohesion-based errors remain a critical bottleneck in Persian–English translation quality. The disruption of cohesive chains, particularly through missing lexical content and mismanaged reference, directly compromises intelligibility. These findings extend prior research by explicitly linking error typology to measurable comprehension outcomes, thereby contributing a discourse-oriented dimension to MT evaluation.

Despite its contributions, this study has several limitations. The corpus was restricted to three non-literary news texts, which may not represent the full range of Persian discourse genres. Literary, technical, or highly specialized texts could exhibit different cohesion patterns and error distributions. The participant pool consisted exclusively of advanced TEFL graduates, whose linguistic proficiency may have mitigated certain intelligibility effects; less proficient readers might experience greater comprehension breakdown. Additionally, the back-translation method, while effective in operationalizing intelligibility, measures reconstructive accuracy rather than real-time comprehension processes. Finally, the evaluation focused on three specific MT systems at a particular point in time; rapid updates to neural models may alter performance profiles.

Future research should expand the textual corpus to include diverse genres such as legal documents, academic writing, literary prose, and informal digital communication to examine whether cohesion-based error patterns vary across discourse types. Comparative studies across additional language pairs involving Persian would clarify whether observed vulnerabilities are language-specific or system-wide. Experimental designs incorporating eye-tracking or real-time comprehension measures could provide deeper insight into cognitive processing of cohesion errors. Moreover, longitudinal studies evaluating updated MT versions would determine whether improvements in neural modeling reduce omission and reference-related deficiencies. Investigations integrating user MT literacy training with cohesion-focused awareness may also reveal whether informed users can compensate for discourse-level weaknesses.

Practically, translation professionals and educators should prioritize cohesion-focused post-editing strategies, with particular attention to resolving pronoun references, restoring omitted lexical items, and ensuring consistent lexical chains. Training programs for translators and EFL learners should incorporate modules on identifying cohesion-based errors in MT output. Developers of Persian–English MT systems should enhance document-level modeling capabilities and expand corpora to better capture discourse dependencies. Finally, translation agencies selecting MT tools should evaluate systems not solely on fluency metrics but also on their ability to preserve cohesive integrity and intelligibility.

Acknowledgments

We would like to express our appreciation and gratitude to all those who helped us carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

All ethical principles were adhered in conducting and writing this article.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

1. Forcada ML. Machine translation today In - Handbook of Translation Studies: Volume 1 (pp. 215-223). John Benjamins Publishing Company; 2012.
2. Wang H, Wu H, He Z, Huang L, Church KW. Progress in machine translation. Engineering. 2022;18:143-53. doi: 10.1016/j.eng.2021.03.023.
3. Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, et al., editors. Google's neural machine translation system: Bridging the gap between human and machine translation2016.
4. Al-Jarf R. Google Translate then and now: Translations from five languages into English and Arabic (2012-2025). Journal of Computer Science and Technology Studies. 2025;7(12):413-27. doi: 10.32996/jcsts.2025.7.12.50.
5. Mirzaeian VR. The effect of editing techniques on machine translation-informed academic foreign language writing. The EuroCALL Review. 2021;29(2):33-43. doi: 10.4995/eurocall.2021.12930.
6. Hardmeier C, editor On statistical machine translation and translation theory In - Proceedings of the Second Workshop on Discourse in Machine Translation (pp. 168-172)2015.
7. Deng Y. Machine Translation in EFL Learning: Chinese Undergraduates' Use, Perceptions, and Machine Translation Literacy. J Electrical Systems. 2024;20(7s):204-16. doi: 10.52783/jes.3263.
8. Yuasa M, Takeuchi O. Strategic use of machine translation: A case study of Japanese EFL university students. AILA Review. 2024;37(2):215-40. doi: 10.1075/aila.24020.yua.
9. Yusof NM, Darus S, Ab Aziz MJ. Evaluating Intelligibility in Human Translation and Machine Translation. 3L, Language, Linguistics, Literature. 2017;23(4). doi: 10.17576/3L-2017-2304-19.
10. Halliday MAK, Hasan R. Cohesion in English: Longman; 1976.
11. Askarieh S, editor Cohesion and Comprehensibility in Swedish-English Machine Translated Texts2014.

12. Weiss S. Cohesion and comprehensibility in Polish-English machine translated texts: Unpublished Master Thesis, Linkoping University, Sweden; 2011.

13. Saffari M, Sajjadi S, Mohammadi M. Evaluation of machine translation (Google Translate vs. Bing Translator) from English into Persian across academic fields. Modern Journal of Language Teaching Methods. 2017;7(8):429-42.

14. Aslerasouli P, Abbasian GR. Comparison of Google Online Translation and Human Translation with regard to Soft & Hard science Texts. Journal of Applied Linguistics and Language Research. 2015;2(3):169-84.

15. Bawden R. Going beyond the sentence: Contextual machine translation of dialogue: Université Paris Saclay (COmUE); 2018.

16. Guillou LK. Incorporating pronoun function into statistical machine translation. 2016.

17. Bowker L. Towards a methodology for a corpus-based approach to translation evaluation. Meta. 2001;46(2):345-64. doi: 10.7202/002135ar.

18. Matsuzaki T, Fujita A, Todo N, Arai NH, editors. Translation errors and incomprehensibility: A case study using machine-translated second language proficiency tests In - Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 2771-2776)2016.

19. Chorowski J, Weiss RJ, Bengio S, Van Den Oord A. Unsupervised speech representation learning using wavenet autoencoders. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2019;27(12):2041-53. doi: 10.1109/TASLP.2019.2938863.

20. Abdi H. Machine Translation Quality Evaluation and Post-Editing Efficiency: The Case of Abadis Translator. k@ ta: A Biannual Publication on the Study of Language and Literature. 2025;27(1):18-33. doi: 10.9744/kata.27.1.18-33.

21. Pshenichnikov D. Key Challenges and Professional Tasks in the Field of Machine Translation Post-Editing. Universal Library of Languages and Literatures. 2024;1(1).

22. Creswell JW, Plano Clark VL. Revisiting mixed methods research designs twenty years later In - Handbook of mixed methods research designs (pp. 21-36). 2023.