

Error Analysis of English Translations of Persian Sport Sciences Thesis Abstracts

Narjes Asgari Vartooni 

MA Student, Translation Studies, Department of English Language and Literature, University of Isfahan, Isfahan, Iran

Mahmoud Afrouz* 

Associate Professor, Department of English Language and Literature, University of Isfahan, Isfahan, Iran

Vazgen Minasian 

Professor, Department of Exercise Physiology, Faculty of Sports Sciences, University of Isfahan, Isfahan, Iran

Abstract

Informative texts, including those in sport sciences, demand precise factual communication yet pose translation challenges due to specialised jargon, culturally bound idioms, and context-dependent nuances. This study aimed to identify the most recurrent errors in human translations of thesis abstracts in sport sciences using the Multidimensional Quality Metrics (MQM) framework, focusing on a period prior to the emergence of Neural Machine Translation (NMT) and artificial intelligence (AI). A sample of 15 English abstracts was randomly selected from Persian-language sport sciences theses from that period. Errors were classified according to MQM dimensions, and two independent raters confirmed the identification and documentation. The analysis revealed a fluctuating error pattern between 2005 and 2010, characterised by an initial decline followed by a subsequent rise, indicating temporary variability in translation quality. These findings provide empirical evidence of recurrent translation challenges in pre-NMT and pre-AI academic sport sciences texts, with implications for translation scholars, academic editors, and thesis supervisors.

Keywords: Abstract translation, error analysis, multidimensional quality metrics, sport sciences

*Corresponding author: afrouz300@gmail.com

Cite this article: Asgari Vartooni, N., Afrouz, M., & Minasian, V. (2026). Error analysis of English translations of Persian sport sciences thesis abstracts. *Translation and Interpreting Research*, 3(9), 1-17.
<https://doi.org/10.22054/tir.2026.21086>



© 2026 – Allameh Tabataba'i University Press

Print ISSN: 3092-6599

Online ISSN: 3060-7019

Publisher: Allameh Tabataba'i University Press

Published: 16.04.2026

Accepted: 16.04.2026

Received: 07.03.2026

1. Introduction

The translation of informative texts, which prioritizes the accurate and clear communication of facts and knowledge, offers a unique set of challenges and requirements. Within this broad category, the translation of sports texts holds a significant and distinctive position. Unlike technical or scientific texts, sports language presents a dynamic fusion of specialized terminology (e.g., rules, techniques, and strategies), cultural references, and emotional expressiveness, particularly in match reports, commentaries, and athlete interviews. This linguistic blend in sports texts makes translation a challenging endeavor, as it not only requires the accurate translation of terms but also the capacity to convey the tone, pace, and cultural nuances of the source text.

The use of sports translation is multi-faceted and wide-ranging. On a global level, sports translation plays a pivotal role in the intercultural transfer of sports coaching methods, sports science studies, and sports-related knowledge, and consequently makes an important contribution to the development and professionalization of sports worldwide. For the media and fans, sports translation is also crucial because it provides authentic access to sports events and sports-related stories from around the world. Effective sports translation strikes a delicate balance between two competing requirements: it must be faithful to the content of the source text, accurately conveying technical detail with unerring precision, while also being sensitive to the stylistic and affective content of the text, adapting it to suit the target audience. This requires a careful selection of equivalent terms and an awareness of the culture of sport within both source- and target-language-speaking communities to avoid misunderstandings and make the translated text resonate as intended.

Since academic knowledge is conveyed through abstracts, the necessity of accurately translating these abstracts for the complete transfer of concepts and to prevent their inadequacy is quite evident. Also, although translation studies predominantly focuses on the translation of books, journal articles, and other textual forms, it has not investigated the critical area of thesis abstracts enough. This gap may lead to consequences such as ambiguity caused by poor translations of studies, leading to misinterpretations that can damage the international standing and reputation of both authors and academic institutions on a global scale. Abstracts containing translation errors prevent researchers from effectively communicating a summary of their studies to English-speaking audiences.

Examining translation errors in student abstracts offers useful information. These errors are frequently the result of persistent linguistic weaknesses, for instance, in grammar, technical terminology, or syntactic organization. Identifying such areas can aid students in refining their language competence, thereby not only enhancing the quality of their present submissions but also preparing them for upcoming academic writing assignments and publication opportunities.

The overall objective of the study is to identify these errors and inform Physical Education students about areas for improvement in abstract translation, thereby enhancing their overall competence in academic language related to the discipline. The study is guided by the following research questions:

1. What are the most recurrent errors identified in human-generated translations compared with AI-generated translations?
2. How does the quality of AI-generated translations compare with that of human-generated translations based on MQM criteria?

2. Literature Review

House (2001) defines translation as “an operation where the meaning of linguistic units is to be kept equivalent across languages” (p. 243). Various frameworks have been proposed for translation quality

assessment (TQA), among which House's (2001, 2015) functional-pragmatic model is one of the prominent approaches. This model evaluates translation quality by examining the extent to which a translation conveys the meaning, style, and function of the source text, with particular attention to textual and cultural factors. Another widely used framework is the Multidimensional Quality Metrics (MQM) model, a comprehensive framework that employs a hierarchical system of error categories to facilitate the systematic and detailed evaluation of translation quality (Lommel et al., 2014).

As MQM constitutes the primary framework adopted in the present study, its structure and relevant error categories are discussed in detail in the Methodology section. The following subsections review relevant error typologies and previous studies to provide the conceptual and empirical background for the analysis.

As Popovic (2018, p. 132) asserts, for any translation practice or approach, "a set of error categories (i.e. an error typology or taxonomy) should be clearly defined beforehand". The development of explicit error typologies has therefore been an important step toward systematic translation quality assessment (TQA). Vilar et al. (2006) report one of the first documented attempts to employ explicit error analysis and a corresponding taxonomy. Their categorization, which draws on an error typology previously used to improve rule-based machine translation systems (Llitió et al., 2005), adopts a hierarchical structure consisting of three levels. At the first level, Vilar et al.'s (2006) typology identifies six major categories: missing words, word order, incorrect words, extra words, unknown words, and punctuation. The category of missing words is divided at the second level into content words and filler words. Word-order errors are classified according to the linguistic level at which they occur, namely word level and phrase level. At the third level, both word-level and phrase-level word-order errors are further distinguished as either local-range or long-range errors. Incorrect-word errors are divided into errors of sense and form. Errors of sense include wrong lexical choice and incorrect disambiguation, whereas form constitutes a separate category. The typology also identifies extra words as a distinct error category. Unknown-word errors are further divided into unknown stems and unseen forms. Finally, punctuation is treated as a separate error category. Thus, the framework by Vilar et al. (2006) distinguishes not only broad error types but also their linguistic level and, where applicable, their scope or specific manifestation.

Farrús et al. (2010, as cited in Popovic, 2018, p. 134) subsequently employed a more concise five-category error typology in their comparison of bidirectional Spanish-Catalan statistical machine translation systems. Their classification consists of morphological errors, lexical errors, orthographic errors, syntactic errors, and semantic errors. Their findings indicate that lexical and semantic errors exert the greatest influence on perceived translation quality. This five-category scheme was later adapted by Comelles et al. (2012) as the basis for a linguistically informed Automatic Evaluation Metric (AEM).

Another relatively basic error typology was employed by Federico et al. (2014, as cited in Popovic, 2018, p. 134). Their framework, implemented through MT-EQuAl2 (Girardi et al., 2014), comprises morphological errors, lexical choice, additions, omissions, casing and punctuation, reordering errors, and cases involving too many errors. In their evaluation procedure, annotators used MT-EQuAl2 both to highlight erroneous words within individual segments and to assign an overall quality score to each segment. This approach therefore combines the identification of specific localized errors with an overall assessment of translation quality.

Costa et al. (2015, as cited in Popovic, 2018, p. 137) proposed a more elaborate hierarchical typology for comparing four different English-to-Portuguese translation systems. Their classification was marginally tailored to Romance languages and comprises several major categories, including orthography, lexis and grammar, semantic errors, discourse, and style. Within orthography, they distinguish punctuation, capitalization, and spelling errors. The lexis and grammar category includes

omission, word-class errors, addition, untranslated items, misselection, verb-related errors, agreement, contraction, and misordering. Semantic errors encompass confusion of senses, wrong choice, collocational errors, and idiomatic errors. At the discourse level, the typology identifies discourse-related problems, while the style category includes variety and cases in which an item should not be translated. This hierarchical organization consequently captures both broad linguistic domains and more specific manifestations of translation errors. Their findings demonstrate that lexical and semantic errors have the strongest influence on sentence-level ranking.

Although there are some studies on the assessment of translation quality of academic texts in Iran, there are still some gaps, especially concerning the TQA of sports-related text translations from Persian into English. Previous investigations have been limited in both scope and methodology. For instance, Candra (2005) examined the linguistic errors of broadcasters whose English was their second language on Indonesia's national television network. Only pronunciation errors among the categories of linguistic errors were analyzed in Candra's study. Hasani (2005) analyzed linguistic errors in some TV programs broadcast by the Islamic Republic of Iran Broadcasting. He identified some types of errors: pronunciation, semantic, grammatical, and prosodic errors, relying on a purely qualitative approach with minimal examples. Akbari and Kheirabadi (2015) explored the linguistic errors detected in three Persian-language websites: "Tabnak", "Asr Iran", and "Alef", finding that all errors fell into three categories: lexical, grammatical, and punctuation errors. The most frequent punctuation errors were typos and the deletion of punctuation marks.

Parvizi et al. (2016) studied the errors students made when translating advertising slogans from Persian to English and vice versa. The participants were MA students in English translation from three Azad universities (Tehran, Shiraz University of Science and Research, and Bandar Abbas). They found that students encountered more difficulties and committed more errors when translating from their native language (Persian) into a foreign language (English). They concluded that interference from the native language in lexical and grammatical features was the major source of errors when translating from Persian to English.

Ghasemi and Hashemian (2016) explored translation errors from English into Persian and vice versa. Based on their findings, semantic-lexical errors had the highest occurrence among the other error types. The findings of this study were confirmed by Laksana and Mursita Putri (2018).

Mugambi et al. (2016) investigated translation errors in Kenyan media based on a corpus of 37 news items. Drawing on Corder's (1967) model, they classified the identified errors into five categories: semantic, lexical, pragmatic, syntactic, and grammatical errors. The study adopted a purely qualitative approach and provided only a limited number of examples to illustrate the identified errors. Consequently, the absence of quantitative analysis constitutes a notable limitation, particularly given the study's focus on the distribution and occurrence of different error types.

Mahdavi (2017) examined translation problems in academic writing by analyzing 15 Persian translations of thesis abstracts. The study compared English abstracts of MA theses written by students at Shahid Bahonar University with their corresponding Persian translations, with the aim of identifying the most and least problematic intertextual factors in the translation of thesis abstracts. The findings indicated that most translation problems were associated with content and sentence structure, whereas topic, text structure, and nonverbal elements presented relatively few identifiable problems.

Zandi (2020) investigated the prevalence of different categories of translation errors in the abstracts of theses and research articles written by Iranian PhD students in psychology. The study aimed to identify the most and least frequent types of translation errors and to compare the translation quality of thesis and article abstracts. The findings showed that omission, addition, faux sens, and contresens were the most frequent error types across both groups of abstracts. The study also reported a significant difference in translation quality between thesis and article abstracts, with statistical analysis

indicating higher overall translation quality in article abstracts than in thesis abstracts. In contrast, errors involving lexical items and inappropriate linguistic variation were found to be the least frequent.

Afrouz (2022) investigated the types of errors committed by news announcers on two radio networks over an eight-month period. After identifying and classifying the errors, he developed a model for their analysis and proposed strategies for addressing the underlying difficulties that resulted in linguistic errors. Although the study provided a relatively comprehensive classification of error types, its framework was developed solely on the basis of Persian-language texts, limiting its applicability to cross-linguistic translation assessment.

Mitselos (2024) examined the effectiveness of Neural Machine Translation (NMT) in the specialized domain of sports journalism. The study provided a detailed classification of the errors encountered by NMT when translating international sports-related press materials from English into Greek for a Greek-speaking audience. The findings indicated that NMT encountered particular difficulties in translating specialized terminology, syntactic and grammatical structures commonly used in journalistic texts, and idiomatic and metaphorical expressions. By systematically documenting these error types, this study highlighted the limitations of NMT in sports journalism and provided a basis for evaluating potential improvements in automated sports-text translation.

3. Methodology

The study adopted a descriptive, corpus-based research design comprising two translation phases: human-generated translation and AI-generated translation. Fifteen abstracts were randomly selected from the 1380s in the Iranian calendar (2005–2010), a period when the human translators of these abstracts did not have access to NMT or AI-based translation tools.

The data collection process followed a multi-stage procedure to ensure the accuracy and completeness of the dataset. First, each abstract was individually scanned to create a digital image database. The scanned images were then processed using Optical Character Recognition (OCR) software to convert them into editable Microsoft Word documents. The resulting Persian texts were carefully verified against the original scans before being entered into an artificial intelligence system for Persian-to-English translation.

For the AI-generated translation phase, DeepSeek was used to translate all abstracts. To ensure consistency across the AI-generated translations, the same prompt was applied to all abstracts within a 24-hour period. In the final stage, the human-generated and AI-generated Persian-to-English translations were systematically analyzed to identify translation errors and assess translation quality. This procedure resulted in the compilation of a complete and consistent dataset for subsequent analysis.

To ensure the reliability of the error classification process, two raters, both MA students in Translation Studies with expertise and experience in Persian-English translation, were invited to independently review the data. Both raters were familiar with the MQM framework and were asked to verify the researchers' identification and documentation of translation errors. Their assessments confirmed the identified errors, and no disagreements were reported between the raters and the researcher.

The Multidimensional Quality Metrics (MQM) framework was used to analyze and evaluate the translations in the present study. MQM is a hierarchical, theory-based framework designed for the systematic assessment of translation quality. The present study employed the latest available version at the time of analysis, *2024_03-07-MQMFull_Master-Official*. The top-level MQM error dimensions used in the analysis were *terminology*, *accuracy*, *linguistic conventions*, *style*, *locale conventions*, *audience appropriateness*, and *design and markup*.

Terminology: This dimension covers errors that occur when a term does not conform to the accepted or normative terminology of the relevant subject field or when the term used in the target text does not constitute the appropriate equivalent of the corresponding source term.

Accuracy: This dimension encompasses errors that arise when the target text does not accurately convey the propositional content of the source text. This dimension reflects the principle that preserving the meaning of the source text is a fundamental component of translation quality.

Linguistic Conventions: Formerly termed “Fluency”, this dimension addresses errors related to the linguistic well-formedness of the target text, including problems with grammaticality, idiomaticity, and mechanical correctness. It covers errors that are clearly incorrect according to established linguistic norms, rather than differences that primarily concern matters of style or authorial preference.

Style: This dimension addresses errors in texts that are grammatically acceptable but inappropriate because they deviate from relevant style guidelines or employ language that is unsuitable for the intended communicative context.

Locale Conventions: The Locale Conventions dimension concerns errors that arise when a translation fails to comply with locale-specific requirements governing the content or formatting of particular data elements. It reflects the need for translated texts to conform not only to general linguistic norms but also to the conventions of the target locale.

Audience Appropriateness: Formerly termed “Verity”, this dimension addresses errors resulting from content that is invalid or inappropriate for the target locale or intended audience. Unlike linguistic errors, such problems are determined largely by extralinguistic considerations, including cultural values and audience expectations.

Design and Markup: This dimension covers errors related to the physical presentation and technical formatting of the translation product. These include problems involving formatting, markup, the integration of textual and graphical elements, and the overall layout and visual presentation of the translated content.

4. Results and Discussion

The section provides a detailed analysis of the research findings, commencing with a presentation of the results. This is followed by a discussion that attempts to offer an in-depth interpretation of the most salient instances extracted from the corpus.

Data analysis involved classifying the identified translation errors according to the MQM criteria. Each error was categorized by type and examined in terms of its frequency of occurrence. For comparative analysis, the identified errors were organized into two categories: human-generated translations and AI-generated translations.

4.1. Classification of Errors in Human Translation

Table 1 presents the frequencies of the top-level MQM error dimensions identified in the abstracts translated by human translators during the 1380s (2005–2010).

Table 1. Frequency of Top-Level Error Dimensions in Each Year

Top-Level Error Dimensions \ Years	1384 (2005)	1385 (2006)	1386 (2007)	1387 (2008)	1388 (2009)	1389 (2010)
Terminology	10				1	1
Accuracy	7	4	5	3	20	23
Linguistic conventions	5	8	10	7	35	32
Style		1	2		5	7
Locale conventions						
Audience appropriateness						1
Design and markup						

The following charts illustrate the percentage distribution of the seven top-level MQM error dimensions identified within the specified time frame.

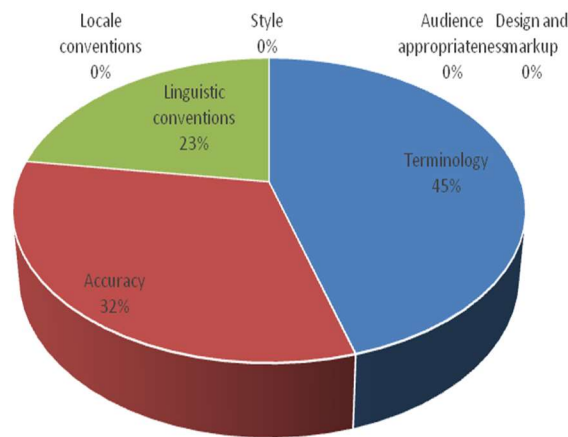


Figure 1. Top-Level Error Dimensions in 1384 (2005)

As shown in Figure 1, terminology accounts for nearly half of the identified errors (45%), making it the most frequent error category. Accuracy represents approximately one-third of the errors (32%), followed by linguistic conventions (23%). The remaining four categories—audience appropriateness, style, locale conventions, and design and markup—record no errors (0%), indicating that no instances were identified in these categories.

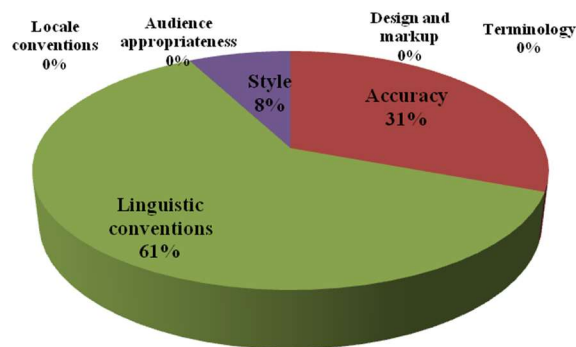


Figure 2. Top-Level Error Dimensions in 1385 (2006)

As shown in Figure 2, “linguistic conventions” accounts for 61% of the total identified errors, making it the most frequent error category. Accuracy follows with 31%, while style accounts for the remaining 8%. The four other categories—audience appropriateness, locale conventions, design and markup, and terminology—record no errors (0%), indicating that no instances were identified in these categories.

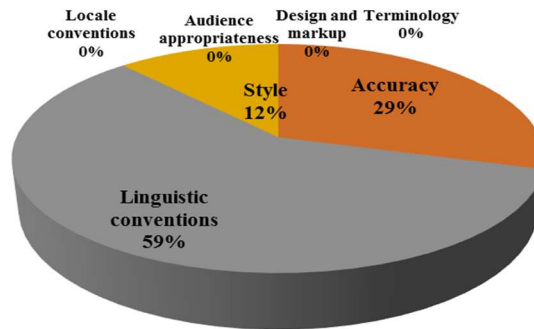


Figure 3. Top-Level Error Dimensions in 1386 (2007)

As shown in Figure 3, “linguistic conventions” accounts for 59% of the total identified errors, making it the most frequent error category. Accuracy follows with 29%, while style accounts for the remaining 12%. The four other categories—audience appropriateness, locale conventions, design and markup, and terminology—record no errors (0%), indicating that no instances were identified in these categories.

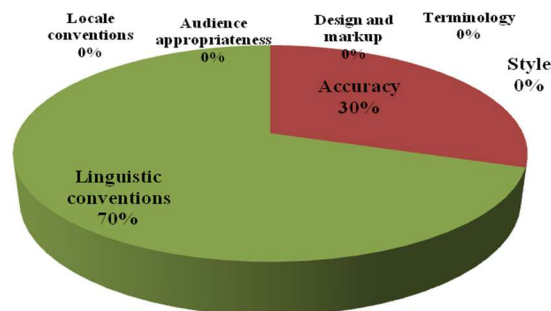


Figure 4. Top-Level Error Dimensions in 1387 (2008)

As shown in Figure 4, “linguistic conventions” accounts for 70% of the total identified errors, while accuracy accounts for the remaining 30%. No errors were identified in any of the other MQM categories.

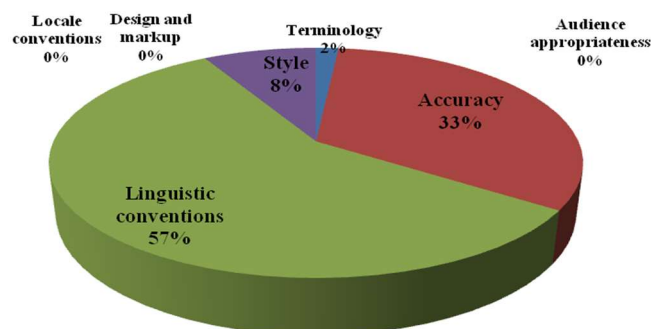


Figure 5. Top-Level Error Dimensions in 1388 (2009)

As shown in Figure 5, “linguistic conventions” accounts for the largest proportion of the identified errors (57%), followed by accuracy (33%) and style (8%). Terminology accounts for only 2% of the errors. No errors were identified in the remaining three categories—audience appropriateness, locale conventions, and design and markup.

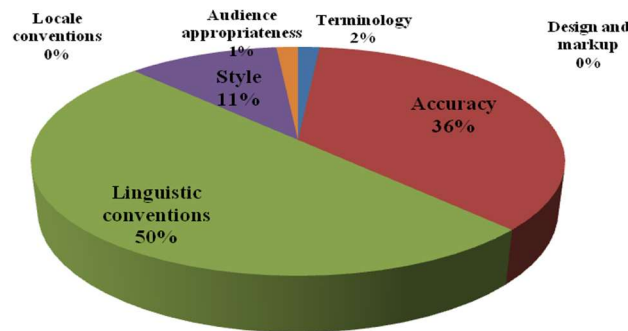


Figure 6. Top-Level Error Dimensions in 1389 (2010)

As shown in Figure 6, “linguistic conventions” accounts for half of the identified errors (50%), followed by accuracy (36%). Style accounts for 11% of the errors, while terminology and audience appropriateness together account for the remaining 3%. No errors were identified in the other two categories.

Table 2 presents the total frequency of the top-level MQM error dimensions identified for each year from 1384 to 1389 (2005–2010).

Table 2. Total Frequency of Top-Level Error Dimensions in Each Year

Years	1384 (2005)	1385 (2006)	1386 (2007)	1387 (2008)	1388 (2009)	1389 (2010)	TOTAL
Frequency	22	13	17	10	61	64	187
Percentage	12%	7%	9%	5%	33%	34%	100%

The line chart presented in Figure 7 illustrates the overall trend in the occurrence of errors across the years.

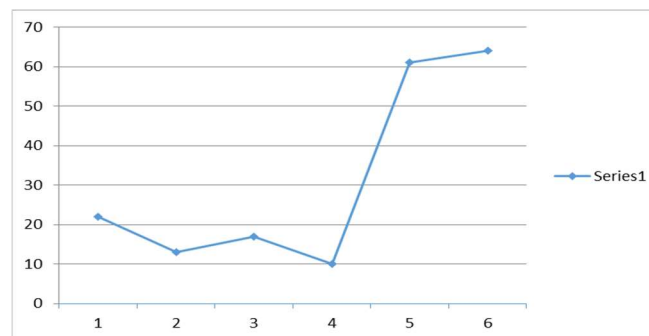


Figure 7. Trend of Top-Level Error Dimensions in 1380s (2005–2010)

According to Figure 7, the frequency of errors in the English abstracts from 1384 to 1389 (2005–2010) does not follow a consistent upward or downward pattern. From 1384 to 1387, the number of errors shows a declining trend, suggesting an improvement in translation quality during this period. However, from 1387 to 1389, the trend reverses, with the frequency of errors increasing again. Overall, the findings indicate fluctuations in error frequency across the period rather than a sustained improvement or deterioration. These variations may reflect differences in editorial practices, authors’

awareness of translation conventions, or the quality of the submitted abstracts, although such factors cannot be established directly from the present data.

Table 3 presents the total frequency of the top-level MQM error dimensions identified in the 1380s (2001–2010).

Table 3. Frequency of Each Top-Level Error Dimensions in 80s (in HT)

Top-Level Error Dimensions	Decades
Terminology	12
Accuracy	62
Linguistic conventions	97
Style	15
Locale conventions	0
Audience appropriateness	1
Design and markup	0
TOTAL	187

The percentages are displayed in Figure 8.

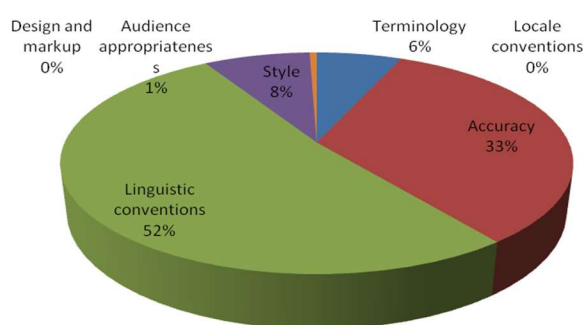


Figure 8. Top-Level Error Dimensions in 1380s (2005 - 2010)

As shown in Figure 8, “linguistic conventions” accounts for the largest proportion of the identified errors (52%), followed by accuracy (33%). Terminology accounts for 6% of the errors, while style represents 8%. Audience appropriateness accounts for only 1% of the errors. No errors were identified in the remaining two categories—design and markup and locale conventions.

4.2. Classification of Errors in AI Translation

Table 4 presents the frequencies of the top-level MQM error dimensions identified in the abstracts translated by AI.

Table 4. Frequency of Top-Level Error Dimensions in Each Year

Top-Level Error Dimensions	Years	1384 (2005)	1385 (2006)	1386 (2007)	1387 (2008)	1388 (2009)	1389 (2010)
Terminology		-	-	-	-	-	-
Accuracy		-	-	-	-	-	4
Linguistic Conventions		-	-	-	-	-	-
Style		-	-	-	-	-	-
Locale Conventions		-	-	-	-	-	-
Audience Appropriateness		-	-	-	-	-	-
Design and Markup		-	-	-	-	-	-

As shown in Table 4, only four Accuracy errors were identified in the AI-generated translations.

This section presents a qualitative analysis of selected instances from the corpus to provide a closer examination of the errors identified in the human-generated and AI-generated translations. It offers a deeper insight into individual cases, highlighting differences between the decisions made by human translators and the output generated by AI.

Example 1:

از این رو هدف از پژوهش حاضر مقایسه اثرات چهار الگوی باردهی تمرین با وزنه بر میزان قدرت و هایپرتروفی عضلانی افراد مبتدی می‌باشد.

Human Translation: Weight trainings widely are used to increase the power, strength and muscular hypertrophy of athletes. The aim of present study was to compare the effects of the four weight trainings loading patterns on the degree of muscle strength and hypertrophy in novice.

AI-generated Translation: Weight training is widely used to increase athletes' strength and muscle hypertrophy. Therefore, the aim of the present study was to compare the effects of four different weight training loading patterns on the strength and muscle hypertrophy of beginners.

An incorrect adverb placement can be observed at the beginning of the sentence. In the human translation, the phrase "widely are used" is used instead of the grammatically correct form "are widely used". The final word, "novice", also constitutes a translation error, as the human translator appears to have overlooked the plural form of the source phrase «افراد مبتدی». Moreover, even as a singular noun, "novice" is grammatically incomplete because it requires an article. It should therefore be "a novice" or, more appropriately in this scientific context, "novices", which preserves the plurality of the source expression.

Example 2:

در پایان دوره تمرینی، قدرت بیشینه و هایپرتروفی عضلانی آزمودنی‌ها به ترتیب با استفاده از آزمون یک تکرار بیشینه در حرکات پرس سینه، خم کردن آرنج، باز کردن زانو و محیط اندام سینه، بازو و ران اندازه‌گیری شد. داده‌ها با استفاده از آزمون آنالیز واریانس یک‌طرفه، آزمون تعقیبی توکی و t همبسته تحلیل شد. نتایج اختلاف معناداری بین چهار گروه تمرینی با گروه کنترل در کسب قدرت و هایپرتروفی عضلانی نشان داد.

Human Translation: At the end of training period, maximum strength and muscle volume of subjects were measured by 1-RM test in bench press, biceps curl, knee extensions and chest, arm and tight limb, respectively. Data analyzed by ANOVA, Tukey post hoc and paired t tests.

AI-generated Translation: At the end of the training period, the subjects' maximum strength and muscle hypertrophy were measured using the 1-RM test for the bench press, arm curl, and knee extension exercises, as well as by measuring the girth of the chest, arm, and thigh. Data were analyzed using one-way ANOVA, Tukey's post-hoc test, and paired t-tests.

Two consecutive sentences contain spelling-related errors. In the first, the Persian word «ران» is incorrectly rendered as «tight», which has a distinct meaning in English and may therefore mislead target readers who do not have access to the source text. Thus, this error cannot be regarded simply as a typographical error. The second error, "Tukey", is a genuine typographical error. In addition, the apostrophe is missing from the human translation. The human translator also omitted the phrase «یک‌طرفه» in «واریانس یک‌طرفه» (one-way ANOVA), thereby failing to convey the complete meaning of the source expression. In contrast, the AI-generated translation renders this phrase accurately.

Example 3:

بدین منظور ۷۴ دانشجوی پسر دانشگاه اصفهان با میانگین (سن 19.45 ± 1.36 سال، قد 173.71 ± 5.04 سانتی متر و وزن 69.16 ± 9.87 کیلوگرم) به صورت دردسترس انتخاب و بر اساس قدرت بیشینه حرکت پرس سینه یکسان سازی شدند و در دو گروه تمرینی (۶۰ نفر) و کنترل (۱۴ نفر) قرار گرفتند.

Human Translation: Seventy four male students from University of Isfahan participated in this study.

AI-generated Translation: For this purpose, 74 male students from the University of Isfahan (with a mean age of 19.45 ± 1.36 years, height of 173.71 ± 5.04 cm, and weight of 69.16 ± 9.87 kg) were selected via convenience sampling.

The whole phrase «با میانگین (سن 19.45 ± 1.36 سال، قد 173.71 ± 5.04 سانتی متر و وزن 69.16 ± 9.87 کیلوگرم)» is omitted in human translation, but correctly translated by AI. The same error has occurred in the next instance.

Example 4:

در حالی که ۶۰ نفر دیگر پیش از آغاز برنامه تمرینی «دوره قدرت بیشینه» به منظور سازگاری های ساختاری، فیزیولوژیکی و آمادگی برای قرارگیری در معرض بارهای بیشینه، به مدت ۸ هفته، هر هفته ۲ جلسه با شدت سبک تا متوسط (۷۰٪-۴۰٪ یک تکرار بیشینه) تحت تمرین قرار گرفتند.

Human Translation: While 60 others before the commence of training program of maximum strength period to structural, physiological adaptations and readiness for exposing to maximum loads exercised in a period 8 weeks, 2 times per week, with light to moderate intensity (40%-70% 1-RM).

AI-generated Translation: Meanwhile, the other 60 subjects underwent an 8-week "maximum strength preparation phase" prior to the main training program. This phase, consisting of two sessions per week at light to moderate intensity (40%-70% of 1-RM), was designed to induce structural and physiological adaptations and prepare them for maximal loads.

The purpose clause is omitted in the human translation. The phrase «به منظور سازگاری های ساختاری، فیزیولوژیکی و آمادگی برای قرارگیری در معرض بارهای بیشینه» in the source text is rendered as "to structural, physiological adaptations and readiness for exposing to maximum loads". The purpose connector "in order to" or "for the purpose of" is missing, resulting in an incomplete and grammatically incorrect construction. Moreover, the sentence as a whole is unintelligible in the human translation, and the intended meaning of the source text is substantially obscured.

Example 5:

با توجه به یافته های پژوهش حاضر می توان ادعا کرد، هنگامی که هدف از تمرین با وزنه افزایش قدرت و هایپرتروفی عضلانی می باشد، استفاده از الگوی باردهی هرمی مسطح بر سه الگوی دیگر مورد استفاده در پژوهش حاضر اولویت دارد.

Human Translation: According to our findings, when the purpose of weight training is strength and muscle volume enhancement, flat pyramid loading pattern prefer to three another patterns in current research.

AI-generated Translation: Based on the findings of the present study, it can be concluded that when the goal of weight training is to increase strength and muscle hypertrophy, using the Flat Pyramidal loading pattern is preferable to the other three patterns used in this study.

The phrase «سه الگوی دیگر مورد استفاده در پژوهش حاضر اولویت دارد» is translated by the human translator as "three another patterns in current research", which is grammatically incorrect and unnatural in English. A more appropriate rendering would be "the other three patterns used in this study", which accurately conveys the intended meaning and follows natural English usage.

Example 6:

در ابتدا و انتهای دوره، قدرت عضلانی و دامنه حرکتی مفاصل بیماران توسط دستگاه نیروسنج آیزوکنتیک (Biodex, Systems III) و گونیامتر ۱۸۰ درجه ای یونیورسال مورد ارزیابی قرار گرفتند.

Human Translation: The patients' muscle strength and joint range of motion were evaluated, through standard laboratory tools, using an isokinetic dynamometer (Biodex, Systems III) and a standard goniometer and the quality of life were evaluated by a standard questionnaire in the begining and end of period.

AI-generated Translation: At the beginning and end of the period, the patients' muscle strength and joint range of motion were assessed using an isokinetic dynamometer (Biodex, Systems III) and a universal 180-degree goniometer.

In the human translation, a spelling error ("begining") and the omission of "180-degree" can be observed. The latter error is more severe because it affects the accuracy of key information in the content.

Example 7:

جامعه آماری این تحقیق را کلیه بیماران هموفیلی مراجعه کننده به بیمارستان سیدالشهدا اصفهان تشکیل می دادند. تعداد ۲۰ مرد هموفیلی با شدت متوسط به صورت دردسترس و هدفمند انتخاب و سپس به صورت تصادفی در دو گروه آزمایش (۱۰ نفر) و کنترل (۱۰ نفر) قرار گرفتند.

Human Translation: The hemophilia patients who referred to Sayedo- Shohada hospital, were statistical society of this study. Twenty moderate hemophilia men selected accessible and targeted, then placed randomaly in experimental (10 persons) and control (10 persons) groups.

AI-generated Translation: The statistical population consisted of all hemophilia patients referred to Seyed-Al-Shohada Hospital in Isfahan. Twenty male hemophilia patients with moderate severity were selected via a purposive and convenience sampling method and then randomly assigned to either an experimental group (n=10) or a control group (n=10).

English abstracts are intended for an international audience. In this instance, the word "Isfahan", which refers to the location of the hospital where the research was conducted, was omitted, despite being important contextual information for the target readers. The human translation also contains an unjustified comma and inappropriate capitalization of the verb "were" in the middle of the sentence. In addition, the auxiliary verb required to render the Persian passive construction «برگزیده شد» has been omitted. The phrase "Were statistical society of this study", used in the human translation, is neither accurate nor natural. A more appropriate rendering would be "constituted the statistical population of this study". This error may have resulted from the literal translation of «جامعه» as "society", rather than the contextually appropriate term "population". Finally, the word "randomaly" is misspelled in the human translation.

Example 8:

قبل و بعد از برنامه بازتوانی عملکرد سیستولی (قطر پایان دیاستولی، قطر پایان سیستولی، حجم پایان دیاستولی، حجم پایان سیستولی، کسر تزریقی و حجم ضربه ای) و توده بطن چپ با استفاده از اکوکاردیوگراف دوبعدی و M-mode، عملکرد دیاستولی (E، A، E/A ratio، DT)، با استفاده از اکوکاردیوگراف پالس ویو داپلر ارزیابی شد و برای اندازه گیری ظرفیت عملکردی از آزمون ورزش بر اساس پروتکل نوتن استفاده شد.

Human Translation: Systolic function (LVEDD, LVESD, LVEDV, LVESV, EF, SV) and LV mass were measured by 2-Dimention guided M-mode method according to the guidelines of American society of Echocardiography

AI-generated Translation: Before and after the rehabilitation program, systolic function (End-Diastolic Diameter, End-Systolic Diameter, End-Diastolic Volume, End-Systolic Volume, Ejection Fraction, and Stroke Volume) and left ventricular mass were assessed using two-dimensional and M-mode echocardiography. Diastolic function (E, A, E/A ratio, DT) was evaluated using pulse wave Doppler

echocardiography, and functional capacity was measured using an exercise test based on the Naughton protocol.

The phrase «کسر تزریقی و حجم ضربه‌ای» is translated as “EF, SV”. Although these acronyms are commonly used in the relevant field, their full English forms, “Ejection Fraction” and “Stroke Volume”, are not introduced before the abbreviations. This omission may reduce clarity for some target readers. Moreover, substantial portions of the source text have been omitted in the human translation, resulting in a loss of important information.

Example 9:

هدف کلی تحقیق حاضر، بررسی تغییرات عملکرد سیستولی، عملکرد دیاستولی، توده بطن چپ و ظرفیت عملکردی در بیماران پس از سکته قلبی (post MI) در طول یک دوره ۸ هفته‌ای بازتوانی می‌باشد.

Human Translation: The main aim of this study is studying of systolic function, diastolic function's, mass of left ventricular and functional capacity variation in post myocardial infarction patients in a 8-weeks rehabilitation.

AI-generated Translation: The overall objective of the present study is to investigate the changes in systolic function, diastolic function, left ventricular mass, and functional capacity in patients after myocardial infarction (post-MI) over an 8-week rehabilitation period.

The phrase “studying of systolic function” is ungrammatical and should be rendered as either “the study of systolic function” or “studying systolic function”. Furthermore, the phrase «در طول یک دوره ۸ هفته‌ای بازتوانی» is translated in the human translation as “in a 8-weeks rehabilitation,” which should be replaced by “over an 8-week rehabilitation period”. This revision involves several grammatical and lexical changes, including replacing the preposition “in” with “over”, changing the article from “a” to “an”, using the singular form “week” in the hyphenated compound “8-week”, and adding “period” at the end of the phrase.

Example 10:

استفاده از تکرار وزنه تا سرحد خستگی برای برآورد یک تکرار بیشینه در حرکت پرس سینه

Human Translation: Using Repetitions to Fatigue to Predict One Repetition Maximum Bench Press

AI-generated Translation: Using Repetition with Submaximal Weights to Fatigue for Estimating One Repetition Maximum in the Bench Press Movement

The Persian phrase «تکرار وزنه تا سرحد خستگی» is incorrectly translated in the human-generated translation as “Using Repetitions to Fatigue”. The source expression refers to performing repetitions with a weight until reaching the point of fatigue, rather than to repetitions of fatigue itself. The human translation therefore fails to accurately convey the intended meaning of the source phrase. In addition, the preposition “in” and the definite article “the” are omitted from the phrase “... One-Repetition Maximum Bench Press”, resulting in a grammatically incomplete construction.

5. Conclusion

Based on the analysis of errors identified in the human-generated translations of abstracts from the 1380s, the findings indicate that the most prominent translation problems during this decade occurred in two primary dimensions: linguistic conventions and accuracy. “Linguistic conventions” accounts for the highest number of errors, with 97 instances (52% of all top-level errors), followed by accuracy, with 62 errors (33%). Style accounts for 15 errors (8%), while terminology accounts for 12 errors (6%). Locale conventions, audience appropriateness, and design and markup show minimal or no occurrence across the decade, with only one error identified in Audience Appropriateness.

Within the accuracy dimension, *omission* and *mistranslation* are the most frequent error subtypes. Omission errors total 33 across the corpus, with the highest frequencies recorded in 1388 (11 errors) and 1389 (9 errors). Mistranslation accounts for 16 errors, with a marked concentration in 1389, where 11 instances were identified. Addition accounts for 11 errors, occurring most frequently in 1388 (7 errors) and 1389 (3 errors).

Within the linguistic conventions dimension, the dominant error subtypes are *grammar*, *punctuation*, and *spelling*. Grammar is the most frequent subtype, with 41 errors overall, driven primarily by 1388, which accounts for 22 instances and represents a notable peak. Spelling accounts for 29 errors, distributed relatively evenly across the years, although 1388 and 1389 record the highest frequencies, with 9 errors each. Punctuation contributes 22 errors, with the highest frequency occurring in 1389 (11 errors), followed by 1388 (7 errors). The unintelligible subtype, although less frequent, is also noteworthy, with 5 errors in total, 3 of which occur in Document 89.

The style dimension consists primarily of *unidiomatic style* (12 errors) and *awkward style* (3 errors). The highest concentration of Unidiomatic Style errors occurs in 1389, with 6 instances. Within the terminology dimension, *Wrong term* is the predominant subtype, accounting for 11 of the 12 terminology errors. Notably, the year 1384 alone contains 10 instances of wrong term errors.

The findings indicate that the primary challenges in translating the abstracts involved avoiding omissions, accurately conveying the source meaning, and maintaining grammatical correctness. The years 1388 and 1389 stand out in particular, with 61 and 64 errors, respectively, accounting for a substantial proportion of the errors identified across the decade. Linguistic conventions and accuracy constitute the most prominent areas of difficulty, whereas lower-frequency dimensions such as locale conventions, audience appropriateness, and design and markup show little to no occurrence.

The error frequency across the decade also shows an unstable pattern rather than a consistent decline or increase. One possible explanation is that, during this period, many students may have relied on professional translators to prepare their English abstracts, particularly given the potentially limited English proficiency of some students. However, this interpretation cannot be established directly from the present data, as the study did not investigate how the abstracts were translated or who translated them.

Moreover, the analysis of errors identified in the AI-generated translations from the 1380s revealed that accuracy was the only MQM dimension in which errors were recorded, with four omission errors accounting for 100% of the identified top-level errors. No errors were identified in the remaining MQM dimensions, including linguistic conventions, style, terminology, locale conventions, audience appropriateness, and design and markup. These findings indicate that, within the present corpus, the AI-generated translations exhibited no identifiable problems in these dimensions according to the applied MQM criteria.

Despite its contributions, the present study has several limitations. First, the analysis was restricted to abstracts in the field of Physiology from the 1380s. Extending the analysis to other subfields of sports science, such as Sports Pathology, Motor Behavior, and Sports Management, could provide a broader basis for examining translation errors in sports-related academic texts. Second, the study examined only DeepSeek as the AI system. Future research could therefore investigate the performance of multiple AI systems, such as Grok, ChatGPT, Claude, and Copilot, when translating sports-related texts. Comparative studies could examine whether different systems exhibit distinct patterns of translation errors and how their outputs vary across MQM dimensions.

References

- Afrouz, M. (2022). Towards designing a framework for analyzing linguistic errors in the national-media news: Providing solutions for challenges. *Zabanpazhuhi (Journal of Language Research)*, 14(44), 145-180. <https://doi.org/10.22051/JLR.2022.37271.2076>
- Akbari, E. & Kheirabadi, R. (2015). A linguistic analysis of errors in news agencies and websites of Iran article in theory and practice in language studies. *Theory and Practice in Language Studies*, 5(11), 2340-2347. <http://dx.doi.org/10.17507/tpls.0511.19>
- Candra, N. (2005). *An analysis on some pronunciation mistakes in the speech of English forum announcers for Radio Republik Indonesia (RRI)*, Surabaya [Bachelor's thesis, Petra Christian University]. Petra Christian University Repository. <https://repository.petra.ac.id/id/eprint/6397/>
- Comelles, E., Atserias, J., Arranz, V., & Castellón, I. (2012). VERTa: Linguistic features in MT evaluation. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)* (pp. 3944–3950). European Language Resources Association (ELRA). <https://aclanthology.org/L12-1447/>
- Corder, S. P. (1967). *The significance of learners errors*. Oxford University Press.
- Costa, A., Ling, W., Luís, T., Correia, R., Coheur, L. (2015). A linguistically motivated taxonomy for machine translation error analysis. *Machine Translation*, 29(2), 127–161. <https://doi.org/10.1007/s10590-015-9169-0>
- Farrús, M., Costa-Jussà, M. R., Mariño, J. B., & Fonollosa, J. A. R. (2010). Linguistic-based evaluation criteria to identify statistical machine translation errors. In F. Yvon & V. Hansen (Eds.), *Proceedings of the 14th annual conference of the European Association for Machine Translation* (pp. 167–173). European Association for Machine Translation. <https://aclanthology.org/2010.eamt-1.12/>
- Federico, M., Negri, M., Bentivogli, L., & Turchi, M. (2014). Assessing the impact of translation errors on machine translation quality with mixed-effects models. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1643–1653). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1172>
- Ghasemi, H., & Hashemian, M. (2016). A comparative study of Google Translate translations: An error analysis of English-to-Persian and Persian-to-English translations. *English Language Teaching*, 9(3), 13–17. <https://doi.org/10.5539/elt.v9n3p13>
- Girardi, C., Bentivogli, L., Farajian, M. A., & Federico, M. (2014). MT-EQuAl: A toolkit for human assessment of machine translation output. In L. Tounsi & R. Rak (Eds.), *Proceedings of COLING 2014, the 25th international conference on computational linguistics: System demonstrations* (pp. 120–123). Dublin City University and Association for Computational Linguistics. <https://aclanthology.org/C14-2026/>
- Hasani, H. (2005). *Looking at the types of language errors in media*. In *Proceedings of the seminar on radio and television and in the media*. Tarhe Ayandeh.
- House, J. (2001). Translation quality assessment: Linguistic description versus social evaluation. *Meta*, 46 (2), 243-257. <https://doi.org/10.7202/003141ar>
- House, J. (2015). *Translation quality assessment: Past and present*. Routledge. <https://doi.org/10.4324/9781315752839>

- Font Llitjós, A., Carbonell, J. G., & Lavie, A. (2005). A framework for interactive and automatic refinement of transfer-based machine translation. In *Proceedings of the 10th EAMT conference: Practical applications of machine translation* (pp. 87–96). European Association for Machine Translation. <https://aclanthology.org/2005.eamt-1.13/>
- Lommel, A. R., Burchardt, A., & Uszkoreit, H. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica*, (12), 455–463. <https://doi.org/10.5565/rev/tradumatica.76>
- Mahdavi, E. (2017). *A TQA of Persian translations of M.A. theses abstracts based on Nord's model: A case study of M.A. theses of Shahid Bahonar University translation students* [Master's thesis, Shahid Bahonar University]. Ganj. <https://ganj.irandoc.ac.ir/>
- Mitselos, N. (2024). *Evaluation of a machine translation system for sports texts in the international press* [Postgraduate thesis, National and Kapodistrian University of Athens]. <https://pergamos.lib.uoa.gr/uoa/dl/object/uoadl:3419983>
- Mugambi, A., Mukuthuria, M., & Kobia, J. M. (2016). Error analysis in mass media in Kenya: A case study of radio news bulletins from Kenya Broadcasting Corporation Kiswahili Station and Radio Citizen. *International Journal of Business, Social Sciences & Education*, 2(3), 173–188. <http://hdl.handle.net/123456789/4472>
- Parvizi, G. R., Shafipour, M., & Mashayekh, J. (2016). An examination of the errors committed by Iranian MA students in their translation of advertisement slogans based on Keshavarz's taxonomy of errors. *Cross-Cultural Communication*, 12(12), 1-13. <http://dx.doi.org/10.3968/9114>
- Popović, M. (2018). Error classification and analysis for machine translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment: From principles to practice* (pp. 129–158). Springer. https://doi.org/10.1007/978-3-319-91241-7_7
- Vilar, D., Xu, J., D'Haro, L. F., & Ney, H. (2006). Error analysis of statistical machine translation output. In *Proceedings of the fifth international conference on language resources and evaluation (LREC'06)* (pp. 697–702). European Language Resources Association (ELRA). <https://aclanthology.org/L06-1244/>
- Zandi, M. A. (2020). *A translation quality assessment of thesis and article abstracts translated from Persian into English in the field of psychology* [Unpublished master's thesis]. University of Isfahan.