



Translation Techniques and Quality of Bilingual Public Signs in Medan

Muhamad Safi'i^{1*}, Hanna Maria Panggabean², Ronika Apriyanti Manik³

¹⁻³English Study Program, Politeknik Negeri Medan, Indonesia

Email: muhamadsafi'i@polmed.ac.id¹, hannamaria@polmed.ac.id², ronikaapriyanti@polmed.ac.id³

*Corresponding author: muhamadsafi'i@polmed.ac.id

Abstract. Public signs have a functional and pragmatic function in multilingual cities; however, there is still a dearth of research on both the translation techniques and translation quality of Indonesian-English public signs outside Java and Bali, especially in Medan. In this study, 250 Indonesian-English translation units were analyzed from public signs located in transportation facilities, heritage/tourism, commerce/retailing, and governmental/public services in Medan from August to September 2026. The coding of translation techniques was done following Molina and Albir's (2002) classification, whereas the translation quality assessment utilized Nababan et al.'s (2012) framework by two independent raters, where interrater agreement was calculated based on Cohen's Kappa. The most common primary techniques found in the data set were Transposition (25.6%), Literal Translation (23.6%), and Established Equivalent (18.4%). In terms of translation quality, as judged by two independent raters against the Nababan et al. (2012) rubric rather than through an empirical comprehension test with target-language readers, the translations can be considered of generally high quality ($M=2.76/3.00$; 72.4% were high). The interrater reliability of Accuracy and Readability was found to be fair, and that of Acceptability as slight, a limitation that tempers the strength of the quality claims, particularly regarding translation naturalness. Techniques focused on literal and conventional equivalence showed higher quality ratings than those focused on reductive and highly interpretive ones.

Keywords: Inter-Rater Reliability; Linguistic Landscape; Public Signs; Translation Quality Assessment; Translation Techniques.

1. INTRODUCTION

English has developed into a global lingua franca that influences the linguistic landscape (LL) of multi-ethnic Indonesian cities in which public signs act as mediators of communication between locals, domestic tourists, and foreigners (Gorter & Cenoz, 2023; Landry & Bourhis, 1997). In Medan, studies of recent LL have indicated that Indonesian, English, and bilingual signs co-exist and that digital signs increasingly replace traditional ones (Nurmala et al., 2025). Public signs form a unique text category. As Newmark (1988) defines texts, they are mainly informative and vocative texts, intended for immediate and situational understanding and action on the part of the reader who might neither understand nor be familiar with the source language and source culture. Translation mistakes on public signs are therefore not only aesthetically but potentially communicatively ineffective, for example, leading to incorrect directions being taken, misunderstanding safety instructions, or reducing the credibility of institutions and authorities. Studies of English-language signs in Indonesia and Thailand describe various kinds of spelling and grammar mistakes, loss or distortion of information, and bad word choices (Laras, 2021; Sangseekaew, 2026).

Translation of public signs is more than just being able to translate as a language for specific purposes (LSP). The spatial limitations in signage necessitate the translators to choose a concise yet functionally equivalent rendering, and the target audience lacks the cultural

knowledge presumed by the source text, thereby necessitating both intercultural and linguistic competency. There are two models that would be effective in assessing this form of translation. One is the Taxonomy of 18 Translation Techniques, developed by Molina & Albir (2002), which provides a systematic description of the techniques used in translation at a micro level, and the other one is the Translation Quality Assessment (TQA) model by Nababan, Nuraeni, and Sumardiono (2012), which assesses the translation in terms of its accuracy, acceptability, and readability, giving an overall quality of the product. Web, movie, and academic texts have also been studied recently in Indonesia through the combination of technique analysis and quality assessment (Dama & Hisasmaria, 2026; Ghithrif & Nugroho, 2025; Simarmata et al., 2025).

In general, studies on public sign translation in Indonesia has been limited to Java and Bali. In terms of Bali, there are studies that investigate grammatical errors, translation techniques, meaning shifts, and adjustments in bilingual signs (Ariani & Artawa, 2022; Prayogi & Manipuspika, 2025; Susini et al., 2021, 2022, 2023), while in terms of Jakarta area there are studies which focus on technique use and verb shifts in transport signs (Estetika & Gusthini, 2024; Mardhiya et al., 2024; Zaman et al., 2025). The contribution from other parts of the country has been less significant. Some studies have focused on signs in Medan (Laras, 2021; Nurmala et al., 2025), Gorontalo (Napu, 2024), and Lombok (Syaia et al., 2026). However, these studies mainly discuss unnatural renderings, dominant techniques, language choice, or power relations in isolation. Specifically, Laras (2021) catalogued categories of unnatural English renderings on Medan's public signs but did not code translation techniques systematically or assess quality using a structured, numerically scored instrument, while Nurmala et al. (2025) examined language choice and the shift toward digital signage in central Medan without analyzing translation technique or quality at all. Neither study paired technique analysis with a structured, inter-rater-validated quality assessment of the kind Cohen's (1960) kappa coefficient enables. The present study aims to fill this gap by jointly analyzing translation technique and translation quality in one corpus of signs, and tests the reliability of the quality scores obtained, thus demonstrating how the technique chosen is linked to the resultant quality rather than treating them as separate lines of inquiry. This joint, reliability-tested design, rather than the Medan location alone, constitutes the primary contribution of the present study.

2. THEORETICAL REVIEW

This part presents the theoretical background for the study: the linguistic landscape and the functional nature of public signs; translation strategy and translation quality evaluation methods used; statistical grounds for measuring inter-rater reliability; empirical studies in Indonesian public sign translation.

Linguistic Landscape and the Function of Public Signs

Linguistic Landscape (LL) is the visibility and prominence of languages as reflected in signs of all kinds in a particular place or region, such as road signs, advertisements, stores, and public institutional signs (Landry & Bourhis, 1997). The authors have further identified the presence of an informational dimension in LL, which involves demarcating the extent of a language community and providing information and a symbolic dimension, which is the symbolic construction of public spaces wherein the relative power and status of language communities are constructed and perceived (Ben-Rafael et al., 2006). Moreover, Backhaus (2007), when studying multilingual signposts in Tokyo, differentiates official signs (produced by public institutions) from non-official signs (produced by individuals or private organizations)—the latter being referred to by Ben-Rafael et al. (2006) as top-down and bottom-up, respectively—stating that they differ in terms of their language use and translation. This distinction is important as both institutional and commercial signage will be considered in this study, including heritage sites and transport as well as coffee shops and shops. The following findings reveal the difference in function of those signs: the language on state warning signs in Bali emphasizes Indonesian, while on private signs it is English (Parthama et al., 2025); official signs in Lombok are more inclined to Indonesian and commercial signs highlight English (Syaia et al., 2026); finally, at Chinese border hospital English is used only as an emblem (Wang & Zhou, 2025). Gorter & Cenoz (2023) offer a recent overview of the field.

Within Translation Studies, for instance, Newmark (1988) categorizes texts in terms of their communicative function as being either expressive, informative, or vocative. Public signs are generally more informative and vocative in nature, and this is because such texts are meant to inform readers about something and evoke a response from them. Unlike literary translation, which places a great deal of importance on style, public sign translation is unique in its function because it aims at communicating something to the reader. Studies that have reviewed the literature on Chinese-English sign translation suggest that there is an abundance of functional theories in this area of research and no feedback from readers is obtained (Amenador & Wang, 2022).

Functionalist Translation and Language for Specific Purposes

The functionalist, skopos-oriented translation theory by Nord (2026) suggests that a translation's goal (skopos) in the target context, not formal equivalency to the source text, should be considered in making decisions. It is especially relevant when applied to public signs, as signs are an example of language for special purposes (LSP). Due to the spatial constraints of the sign and the fast reading, a technically perfect translation that cannot be acted upon by a target reader does not fulfill the communicative purpose of the sign. Studies provide evidence for this approach: Susini et al. (2022) have demonstrated that the modifications of the form of Indonesian public signs in Bali resulted in a change of their contextual meaning, including less polite and more polite forms, while Hazaea et al. (2024) found politeness used on public signs in Abha, Saudi Arabia, at a tourist destination.

Translation Techniques: Molina and Albir's (2002) Taxonomy

Analysis of the specific strategies used by translators on the micro-level can be traced to the comparative stylistics approach developed by Vinay & Darbelnet (1995) of comparing French and English languages through the identification of a list of translation “procedures” (for instance, borrowing, calque, literal translation, transposition, modulation, equivalence, and adaptation) to distinguish between direct and oblique translations. (Molina & Albir (2002) updated the list to produce a more dynamic functional taxonomy of 18 different translation techniques, and they made a clear distinction between “technique,” which is “the end-result focused choice of text,” “method,” which is the general translator’s approach to translation (free or literal), and “strategy.” According to the classification used by Molina and Albir (2002), the use of more than one technique in a translation unit is possible, and the choice of a technique is considered functional depending on the context, target audience, and purpose of translation instead of being based on a concept of equivalence. The classification was selected for this study due to its functionalist approach and possibility to code techniques as primary and secondary ones. It has already been applied to translation of Indonesian public signs, including signs of Jabodebek LRT (Estetika & Gusthini, 2024) and Jakarta integrated transport system (Mardhiya et al., 2024), as well as information boards of commuter lines (Zaman et al., 2025).

Translation Quality Assessment

Models of TQA make use of certain criteria for assessing the adequacy of the translation product. An example of such a model is the one made famous by House (2015), who makes a distinction between overt and covert translation and employs the analysis of register and genre of both source and target text as the means of assessing the quality of the translation, producing

thus a qualitative analysis of mismatches rather than quantitative evaluation in terms of number. Another example of a translation quality assessment model tailored specifically to the needs of Indonesian translation studies is the one developed by Nababan et al. (2012). Their model assesses the translation product in three dimensions: accuracy (*keakuratan*), acceptability (*keberterimaan*), and readability (*keterbacaan*). Each dimension is scored on a 1-3 scale by the raters and calculated as a weighted quality score using the formula $\text{Quality Score} = (\text{Accuracy} \times 3 + \text{Acceptability} \times 2 + \text{Readability} \times 1) \div 6$, resulting in a composite quality score that ranges from 1.00 to 3.00 and accounts for accuracy being the most significant dimension, followed by acceptability and readability. For this study, the Nababan et al. (2012) framework was preferred over the one offered by House (2015), since the former's numerically expressed criteria and weighting of the dimensions are more suitable for quantitative analysis along with translation technique and have been previously used in studies conducted in Indonesia concerning bilingual signboard translations and other non-literary texts (Dama & Hisasmaria, 2026; Ghithrif & Nugroho, 2025; Mardhiya et al., 2024; Nuraeni et al., 2016).

Inter-Rater Reliability in Translation Quality Research

Since the assessment of TQA scoring relies on human judgment, the quality score provided by only one rater runs the risk of being biased by the rater's individualized criteria as opposed to a sound basis of criterion. The kappa coefficient developed by Cohen (1960) solves this problem by determining the degree of agreement between two independent raters using a categorical scale above what would be achieved by mere chance. It is usually used to assess the 1-3 scores of the Nababan et al. (2012) rubric, and since these scores are ordered, the use of weighted kappa can prove valuable by rewarding near agreement. A kappa coefficient interpretation scale by Landis & Koch (1977): less than 0.00 denotes poor agreement, 0.00-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.00 almost perfect agreement. Kappa is also sensitive to rating distribution; if most of the ratings are in the same category, then there will be high agreement by chance, hence making the value of kappa low even when there is high observed agreement; this phenomenon is referred to as the kappa paradox (Zec et al., 2017). It, therefore, becomes important for the observed agreement to be presented together with the value of kappa. The inclusion of Cohen's Kappa within TQA research, like in this paper, provides a clear numerical way of checking reliability of quality scores.

Previous Studies on Indonesian Public Sign Translation

A number of studies have been conducted on translation techniques and/or quality of Indonesian public signs, but most of them were conducted in Java and Bali. In Bali, Ariani &

Artawa (2022) found errors such as incorrect construction of nouns, auxiliary verbs, spelling, and tenses, among others, in the English translation of five public signs in Ubud because of the improper use of the translation techniques and the interference of Indonesian language. Susini et al. (2021) found that signs in Indonesian language in Bali had been translated using substitution, amplification, reduction, condensation, and reordering; Susini et al. (2023) found that translators made alterations on signs by using sentences and non-sentences; and Prayogi & Manipuspika (2025) found that literal translation along with poor consideration of context resulted in ambiguity and errors in bilingual traffic signs. In Jakarta, Estetika & Gusthini (2024), in their analysis of 13 LRT signs of Jabodebek with the use of Molina and Albir's model, found the presence of six out of 18 methods used (adaptation, substitution, literal translation, established equivalent, modulation, and description), but without providing the frequency of use. Mardhiya et al. (2024) in analyzing 190 verb translations on the signs of Jakarta's integrated transportation system, found that the translation shifts were frequent and provided an average quality score of 2.12, described as inaccurate and unacceptable. Zaman et al. (2025) discovered that literal translation and established equivalents had high readability of commuter line information boards, while borrowing, modulation, and amplification had medium or low readability.

Less evidence is available on the issue in other regions. Concerning Medan, for instance, Laras (2021) identified spelling mistakes, missing or modified information, grammar mistakes, and poor vocabulary usage as the main problems with unnatural English translations of Indonesian public signs and suggested borrowing, adaptation, and transposition as possible solutions, while Nurmala et al. (2025) examined language selection and digital media in relation to the signs of central Medan. In Napu's research (2024), most of the 123 bilingual signs in Gorontalo were translated literally and not translated at all, while in Syaia et al.'s research (2026), regulatory signs in Lombok were based on literal translation, but commercial signs were adaptive. Similar conclusions can be drawn from research on other language pairs, such as Algryani (2021) on Arabic–English translation problems in signs and Sangseekaew (2026) on orthographic, grammatical, and semantic mistakes in Chiang Mai English signs and their causes.

3. RESEARCH METHODOLOGY

This study utilized descriptive qualitative design coupled with frequency-based quantitative design. The qualitative portion entailed identification of the translation techniques

used in translating each translation unit, whereas the quantitative side involved computation of frequencies, percentages, mean score ratings, and inter-rater reliability coefficients.

Data included 250 translation units of Indonesian and English found in 84 bilingual public sign photos taken on various purposively chosen sites in Medan. Photos were included if they (a) displayed an Indonesian source text paired with a visible, legible English target text on the same physical sign, (b) were located at a site accessible to the general public, and (c) were not exact duplicates of a sign already photographed at the same location; signs with only one language, or with a third language in addition to Indonesian and English, were excluded. Nine sites were sampled, namely Medan Railway Station (76 units), Museum Sumut (70), Sun Plaza Mall (52), Medan Electric Bus (35), Center Point (6), Starbucks (4), Tjong A Fie Mansion (3), Kesawan (2), and Jalan Jawa (2). Sites were purposively selected to represent some functional domains in which Indonesian–English bilingual public signs are commonly found: transportation ($n = 113$), heritage/tourism ($n = 73$), and commercial/retail ($n = 64$). Table 1 shows how each of the nine sites maps onto these domains; each site falls entirely within a single domain. In case one photo contains more than one independent phrase, all phrases were split and presented in different translation units. In other words, in this study, one unit of analysis consists of one source-target phrase pair, not the entire sign.

Table 1. Mapping of Sampling Site to Functional Domains.

Site	Domain	n	%
Medan Railway Station	Transportation	76	30.4%
Medan Electric Bus	Transportation	35	14.0%
Jalan Jawa	Transportation	2	0.8%
Museum Sumut	Heritage/Tourism	70	28.0%
Tjong A Fie Mansion	Heritage/Tourism	3	1.2%
Sun Plaza Mall	Commercial/Retail	52	20.8%
Center Point	Commercial/Retail	6	2.4%
Starbucks	Commercial/Retail	4	1.6%
Kesawan	Commercial/Retail	2	0.8%
Total		250	100.0%

There were two instruments used in this study. The first one was a list of translation techniques that consisted of 18 translation strategies created by Molina and Albir (2002) (such as literal translation, transposition, modulation, established equivalent, borrowing, amplification, reduction, discursive creation) and coded technique(s) used for translation of every unit; a unit could have one or more than one technique used. In this study, the primary technique refers to the technique that best accounts for the overall transformation of the unit (e.g., the strategy used for rendering the core proposition or the dominant structural change), whereas a secondary technique refers to an additional technique applied to a specific lexical

item or sub-segment within the same unit that does not by itself account for its overall transformation (e.g., an explanatory word added to clarify a term otherwise rendered through established equivalent or transposition). When more than one technique appeared to apply to the same transformation, the technique that most directly explained the change between source and target text was coded as primary. All 250 units were coded by the second author, with a written justification recorded for every decision; these were periodically discussed with the research team, but no independent second coder re-coded the data and no formal inter-coder agreement statistic was computed. This is acknowledged as a limitation in the Conclusion.

The second instrument was a Translation Quality Assessment (TQA) rubric created by Nababan et al. (2012). This rubric allowed us to evaluate accuracy, acceptability, and readability of the translation of every unit and assign each dimension one of the numbers from 1 to 3, resulting in the overall quality score from 1.00 to 3.00 using the formula: $\text{Quality Score} = (\text{Accuracy} \times 3 + \text{Acceptability} \times 2 + \text{Readability} \times 1) / 6$, yielding a composite score ranging from 1.00 to 3.00.

The translation quality was rated by two raters, both lecturers in the English Study Program at Politeknik Negeri Medan holding a Master's degree related to English/Applied Linguistics, with experience in English teaching and translation. Both received a shared written orientation to the Nababan et al. (2012) rubric, in which a score of 3 denotes a fully accurate, acceptable, or readable rendering; 2 denotes minor deviations that do not obscure the intended message; and 1 denotes a significant mismatch in meaning, an unnatural or ungrammatical form, or a rendering difficult for a target reader to understand. No joint calibration session or pilot round was conducted. The raters rated all 250 units independently and without knowledge of each other's scores, and final scores were the simple average of the two raters' weighted scores regardless of the degree of disagreement, with no post-hoc adjudication. This protected rater independence but likely also contributed to the low kappa for Acceptability ($\kappa = 0.124$); a calibration session is recommended for future replications.

The data gathering took place in three phases. Firstly, public signs at the selected sites were taken pictures of and these pictures were documented. Secondly, each sign was segmented into translation units, each one of which had a unique number (MED-001 to MED-250). Then the source and target texts of these units were transcribed and coded using the classification system provided by Molina & Albir (2002) for translation technique with the reason behind each coding documented. Lastly, the two raters independently rated all translation units on all three TQA dimensions.

Data on translation techniques were analyzed using frequency counts and percentages for primary and secondary techniques. With regard to translation quality, reliability of inter-rater agreement was first examined using Cohen's Kappa for each TQA dimension, with coefficients classified based on the criteria by Landis and Koch (1977): <0.00 - poor, 0.00–0.20 - slight, 0.21–0.40 - fair, 0.41–0.60 - moderate, 0.61–0.80 - substantial, 0.81–1.00 - almost perfect agreement. In addition, observed agreement and quadratic-weighted kappa were calculated since the scale was ordinal and the ratings could be skewed to one category (Zec et al., 2017). After that, the final quality score of each unit was determined as an average of both raters' weighted scores and classified into three groups with cut-off scores 1.60 and 2.60: high (2.60-3.00), moderate (1.60-2.59), and low (1.00-1.59). Finally, means of quality scores were cross-tabulated with primary translation technique. As an explanatory analysis, differences between techniques with at least ten units were examined using Kruskal-Wallis test, and grouped comparisons used the Mann–Whitney U test.

4. RESULTS AND DISCUSSION

Translation Techniques Used in Bilingual Public Signs

Table 2 presents the frequency distribution of the primary translation technique (Technique 1) applied across the 250 translation units.

Table 2. Distribution of Primary Translation Techniques.

No	Translation Technique	n	%
1	Transposition	64	25.6%
2	Literal translation	59	23.6%
3	Established equivalent	46	18.4%
4	Modulation	27	10.8%
5	Amplification	16	6.4%
6	Reduction	15	6.0%
7	Borrowing	11	4.4%
8	Generalization	4	1.6%
9	Discursive creation	3	1.2%
10	Adaptation	2	0.8%
11	Description	2	0.8%
12	Particularization	1	0.4%
	Total	250	100.0%

Transposition was the most commonly used technique (25.6%), with Literal Translation (23.6%) and Established Equivalent (18.4%) coming second and third, respectively. Altogether, these three techniques comprised 67.6% of all units. The prominence of transposition was due to the translation of institution and museum names, since 36 out of 64 transposed units (56.3%) were found on the museum signs. In most cases, transposition involved reordering the head-modifier sequence of Indonesian institutional and place names (typically Head Noun + Modifying Noun) into the English premodifying Noun+Noun

convention common for institution names (Modifier + Head Noun), a structural reordering rather than a change in word class (e.g., Istana Maimun → Maimun Palace; Ruang Etnografi → Ethnography Room). A smaller subset did involve a genuine shift in grammatical category, for instance the Indonesian noun Islam rendered as the English adjective Islamic in Ruang Islam → Islamic Room. Prominence of literal translation is consistent with the finding by Napu (2024), who noted that most of the bilingual signs in Gorontalo used literal translation or did not translate at all, and with that of Syaia et al. (2026), who found literal translation on regulatory signs in Lombok, whereas the abundance of established equivalents (for example, Lift → Elevator; *Alat pemadam kebakaran* → Fire extinguisher) mirrors high readability reported by Zaman et al. (2025) for established equivalents on commuter-line boards. Such interpretation techniques as Modulation (10.8%), Discursive Creation (1.2%), and Adaptation (0.8%) were considerably less frequent and were usually applied to culturally-specific notions or safety rules where meaning had to be modified in addition to being transferred structurally. Most of the examples of Modulation involved politeness shift, such as "*Dilarang melewati pembatas ini...*" → "Please do not cross the line..." and "*Maaf, jangan disentuh*" → "Please do not touch," which is in accordance with the findings for the softening of messages on Balinese signs (Susini et al., 2022) and politeness strategies on tourist signs in Abha (Hazaea et al., 2024).

In addition to the primary technique, 76 of 250 units (30.4%) exhibited a secondary technique applied concurrently, as shown in Table 3.

Table 3. Distribution of Secondary Translation Techniques.

No	Secondary Technique	n	%
1	Amplification	23	9.2%
2	Established equivalent	15	6.0%
3	Transposition	11	4.4%
4	Reduction	10	4.0%
5	Particularization	3	1.2%
6	Modulation	3	1.2%
7	Literal translation	3	1.2%
8	Borrowing	3	1.2%
9	Generalization	2	0.8%
10	Description	2	0.8%
11	Adaptation	1	0.4%
	Total units with secondary technique	76	30.4%

Among the secondary techniques, amplification was applied most frequently, occurring in 23 of the 250 translation units (9.2% of the full corpus, or 30.3% of the 76 units that carried a secondary technique), where additional clarifying words were used (for example, "building," "team," "safety") that were not originally part of the source text to help make sense of it to the target audience. The strategy of amplification is one of the five strategies used by Susini et al. (2021) for public signs in Bali, thus indicating that a typical way of making explicit short signs

for foreigners is the addition of a clarifying word. Other common secondary techniques were Established Equivalent and Transposition.

Translation Quality of Bilingual Public Signs

Before combining both raters' scores, inter-rater agreement was tested using Cohen's Kappa for each TQA dimension. Results are presented in Table 4.

Table 4. Inter-Rater Reliability Results (Cohen's Kappa).

Dimension	N	Po	Pe	Kappa	Interpretation
Accuracy	250	0.752	0.671	0.247	Fair
Acceptability	250	0.696	0.653	0.124	Slight
Readability	250	0.792	0.735	0.214	Fair

Based on the categorization of Landis & Koch (1977), agreement in Accuracy ($\kappa = 0.247$) and Readability ($\kappa = 0.214$) was categorized as fair, while Acceptability ($\kappa = 0.124$) was categorized as slight. This is in comparison with the observed agreement that was between 69.6% and 79.2%. This difference was due to the fact that the ratings were highly skewed towards the highest rating where 74.8% to 89.6% of each rater's ratings were 3 on every dimension. As a result, the expected chance agreement ($P_e = 0.653$ – 0.735) increased while the kappa decreased, a phenomenon that is known as the kappa paradox (Zec et al., 2017). Quadratic-weighted kappa, which credits near agreement on an ordinal scale, was higher (Accuracy 0.37, Acceptability 0.27, Readability 0.30) but still in the fair range. Acceptability is characterized by lower agreement because the evaluation of the translation's naturalness and adherence to the target language is a more subjective procedure than the assessment of the meaning accuracy and comprehensibility, and the raters' mean ratings differ the most on this criterion (Rater 1 = 2.81, Rater 2 = 2.64); in contrast, they differ very little on the criterion of Accuracy (2.76 and 2.76). Given that the raters performed the task independently without prior discussion or calibration (see Research Methodology), the fair-to-slight agreement levels obtained are consistent with this absence of a calibration step rather than indicating a moderate level of agreement; the resulting mean quality scores, especially for Acceptability, should therefore be interpreted as estimates that may mask disagreement between raters rather than as a fully reconciled single judgment.

Quality scores were calculated using Nababan et al.'s (2012) weighted formula: Quality Score = (Accuracy $\times$ 3 + Acceptability $\times$ 2 + Readability $\times$ 1) / 6. Table 5 presents the mean score for each quality dimension across all 250 units.

Table 5. Mean Translation Quality Scores by Dimension.

Quality Dimension	Mean Score
Accuracy	2.76
Acceptability	2.73
Readability	2.82
Overall Weighted Quality Score	2.76

The overall quality score average was 2.76 out of the possible total score of 3.00, suggesting that the bilingual translation of public signs in Medan is, in general, of high quality. The score is numerically higher than the 2.12 average reported for verb translations of transport signs in Jakarta (Mardhiya et al., 2024) and the 2.65 average reported for human-translated food product names on a commercial site (Ghithrif & Nugroho, 2025); however, these figures are not directly comparable, as the studies differ in corpus composition (verbs only vs. full phrase units vs. product names), unit of analysis, and possibly in how raters applied the TQA rubric, so this contrast should not be read as evidence that public sign translation in Medan is categorically superior. Readability received the highest score (2.82), indicating that the translations were easy to comprehend by target readers, whereas Acceptability received the lowest score (2.73), which is also the dimension with the greatest discrepancy between the two raters. Nuraeni et al. (2016) observed the same pattern in terms of universal school signboards in Surakarta.

To provide a more detailed picture, final quality scores were categorized into three bands using cut-off values of 1.60 and 2.60: High (2.60–3.00), Moderate (1.60–2.59), and Low (1.00–1.59). The resulting distribution is shown in Table 6.

Table 6. Distribution of Translation Quality Categories.

Quality Category	n	%
High (2.60–3.00)	181	72.4%
Moderate (1.60–2.59)	61	24.4%
Low (1.00–1.59)	8	3.2%
Total	250	100.0%

Most units were classified as high-quality (72.4%), which suggests that the majority of bilingual signs under study were well-translated and acceptable. A smaller percentage of units (24.4%) were classified as moderate, while only 3.2% (8 units) were classified as low-quality. Low-quality units were usually complete failures of sense or form, such as MED-022 (from “*Dilarang membawa barang menyengat*” to “No strong odor good”; 1.00), MED-221 (from “*Dilarang membuang air kecil di lantai toilet*” to “Do not urinate on the bathroom float”; 1.25), two word-by-word translations of union names (MED-001 and MED-002), as well as spelling mistakes like “Nurersy Room” (MED-189). Such issues as spelling errors, changed content, and inappropriate words coincide with Laras’s (2021) categories of unnatural translations for Indonesian–English bilingual signs in Medan and with Sangseekaew’s (2026) orthographic and semantic mistakes for English bilingual signs in Chiang Mai.

Weighted averaging may also mask critical issues. MED-242 ("*Dilarang Masuk*" → "Exit Only") involves changes to the directive and was rated 1 in accuracy by both raters, but got 1.83 due to having more points for its acceptability and readability; MED-087 ("*Ambil souvenir disini*" → "Prize out") received 2.00 solely due to a complete disagreement between the two raters (Rater 1 rated 3 in each aspect, Rater 2 rated 1). This issue of mismatch is most critical in directional and safety signs, as misinterpretations can alter their meaning (Prayogi & Manipuspika, 2025).

Because the weighted formula gives Accuracy the greatest weight but still blends it with Acceptability and Readability, a unit can receive a moderate overall score despite a severe meaning failure, as MED-242 illustrates. For practical use, we recommend reporting a separate “critical accuracy failure” flag alongside the weighted score, defined as any unit on which both raters independently assigned the lowest Accuracy score (1). In the present corpus, three units met this criterion (MED-002, MED-022, and MED-242; 1.2% of the corpus), and these would benefit from prioritized correction or a dedicated safety review independent of their composite score.

Relationship Between Translation Technique and Translation Quality

To explore how technique choice relates to translation quality, the final quality score of each unit was grouped by its primary translation technique. Results, ordered from highest to lowest mean quality score, are presented in Table 7.

Table 7. Mean Translation Quality Score by Translation Technique.

Translation Technique	n	Mean Quality Score
Borrowing	11	2.95
Established equivalent	46	2.87
Particularization	1	2.83
Amplification	16	2.80
Transposition	64	2.79
Literal translation	59	2.78
Modulation	27	2.65
Adaptation	2	2.62
Description	2	2.50
Reduction	15	2.45
Generalization	4	2.44
Discursive creation	3	2.42

Borrowing (M = 2.95) and Established Equivalent (M = 2.87) showed the highest mean quality scores. Although they are conceptually distinct techniques under Molina and Albir’s (2002) taxonomy (borrowing retains a source-language form largely unchanged, whereas established equivalent substitutes a conventionally recognized target-language form), both rely on a form that is already fixed or widely recognized rather than newly constructed by the translator, which plausibly explains their comparatively low risk of meaning distortion. Borrowed units comprised mostly proper names and loan words (e.g., “Raz Museum &

Gallery,” “Toilet,” “Taxi”); hence, it may be possible that borrowing produced the highest score while in a study by Zaman et al. (2025), borrowing, modulation, and amplification produced medium and low readability in commuter-line boards since the material being borrowed could be a factor. On the other hand, Discursive Creation ($M = 2.42$), Generalization ($M = 2.44$), and Reduction ($M = 2.45$) produced the lowest mean scores, although Discursive Creation ($n = 3$) and Generalization ($n = 4$) are based on very few units and should be read as descriptive only, not as evidence of technique effectiveness; Reduction ($n = 15$) is the only one of the three large enough for the exploratory tests below. This is consistent with the nature of these methods, since discursive creation results in images that are inherently difficult to predict from the source text and thus susceptible to meaning change; reduction involves losing some of the information contained in the source text; and generalization involves lowering the specificity of translated expressions. Likewise, Algryani (2021) cited generalization and omission as some of the methods employed in Arabic–English signs and attributed inaccurate translations to inappropriate use of methods.

Due to the extreme imbalance of the number of units within each method, the means for techniques having less than ten units (generalization, discursive creation, adaptation, description, and particularization) are only suggestive. For the seven techniques with at least ten units, an exploratory Kruskal-Wallis test revealed differences in quality, $H(6) = 22.28$, $p = .001$, $\eta^2H = .07$ (small-to-medium effect). Two targeted contrasts were then examined rather than all pairwise comparisons: Reduction ($n = 15$, $M = 2.45$) scored significantly lower than the four traditional techniques (Borrowing, Established Equivalent, Literal Translation, and Transposition; $n = 180$, $M = 2.82$), $U = 692$, $p < .001$, $r = .49$ (large effect); and techniques oriented toward literal or conventional equivalence ($n = 180$, $M = 2.82$) scored higher than reductive or interpretive ones (Reduction, Generalization, and Discursive Creation; $n = 22$, $M = 2.44$), $U = 2,996.5$, $p < .001$, $r = .51$ (large effect). No formal correction for multiple comparisons was applied, as only these two contrasts were tested. The “literal/conventional” versus “reductive/interpretive” grouping was derived post hoc from the technique-level means and should be read as an exploratory, hypothesis-generating pattern, not a confirmatory test. Moreover, the 250 units were drawn from 84 photographed signs (median one unit per photo, but up to 40 from a single sign), so units are not fully independent. The Kruskal-Wallis and Mann-Whitney tests do not account for this clustering; the p -values and effect sizes therefore apply to the pooled units, and future work should use a clustering-aware approach (e.g., a mixed-effects model with photo as a random effect). Techniques oriented toward literal or conventional equivalence thus appear to produce higher and more consistent quality on public

signs than interpretive or reductive ones, although this association does not imply causation, because reductive or interpretive techniques may be used for difficult-to-translate source texts.

All these results suggest that public signs translation follows the principles of pragmatic function of public signs discussed before (Newmark, 1988; Nord, 2026), since such signs need to be understood instantly by those people who are not acquainted with the source culture, and, therefore, translators prefer to use strategies that retain clarity of the structure and semantics of the signs, and the strategies themselves are correlated with the higher translation quality. Unlike Ariani & Artawa (2022) and Prayogi & Manipuspika (2025) whose research shows that literal translation leads to errors and distorted meaning of Balinese signs, literal translation in the current corpus was near to the average score ($M = 2.78$), what means that it is rather safe to translate short conventional phrases but rather risky to translate long or culture-specific ones. The rather low level of inter-rater agreement on Acceptability, at last, shows that even TQA instrument that is quite well structured can leave some space for subjectivity of raters' judgment, especially for linguistic naturalness assessment.

5. CONCLUSION

This study investigated translation strategies and quality of 250 translation units between Indonesian and English public sign translations in Medan. Transposition (25.6%), Literal Translation (23.6%), and Established Equivalent (18.4%) contributed to 67.6% of the major translation strategies, implying that the translators opted for more transparent translation. The overall quality of the translation units was quite high ($M = 2.76$ out of 3.00; 72.4% of the units were of high quality), though inter-rater reliability was only fair for Accuracy and Readability and slight for Acceptability. Established Equivalent and Borrowing achieved the highest mean quality scores, while Reduction ($n = 15$) showed significantly lower quality than the four most literal/conventional techniques among the seven strategies with at least ten units, $H(6) = 22.28$, $p = .001$; Generalization ($n = 4$) and Discursive Creation ($n = 3$) also showed low descriptive means but, given their very small samples, this pattern is suggestive rather than confirmed. The present study was confined by several limitations, including the city-specific design, unequal sampling across sites and techniques, the absence of a rater calibration session, low inter-rater reliability on Acceptability, the lack of inter-coder agreement verification for the technique coding, and the non-independence of translation units nested within the same photographed sign, which was not accounted for in the inferential tests.

ACKNOWLEDGEMENT

The authors would like to thank Pusat Penelitian dan Pengabdian kepada Masyarakat (P3M) of Politeknik Negeri Medan for funding this research through the Fundamental Research (*Penelitian Fundamental*) scheme under the DIPA budget, as stipulated in contract number B/331/PL5/PT.01.05/2026. The authors also extend their appreciation to all parties who contributed to data collection and rating during this research.

REFERENCES

- Algryani, A. (2021). On the Translation of Linguistic Landscape: strategies and quality assessment. *Khazar Journal of Humanities and Social Sciences*, 24(2), 5–21. <https://doi.org/10.5782/2223-2621.2021.24.2.5>
- Amenador, K. B., & Wang, Z. (2022). The Translation of Culture-Specific Items (CSIs) in Chinese-English Food Menu Corpus: A Study of Strategies and Factors. *SAGE Open*, 12(2), 1–7. <https://doi.org/10.1177/21582440221096649>
- Ariani, N. M., & Artawa, K. (2022). Analysis of the grammatical errors of English public signs translations in Ubud, Bali, Indonesia. *Journal of Language and Linguistic Studies*, 18(2), 899–909. <https://www.jlls.org/index.php/jlls/article/view/2974>
- Backhaus, P. (2007). *Linguistic Landscape: A Comparative Study of Urban Multilingualism in Tokyo* (J. Edwards, Ed.). Multilingual Matters.
- Ben-Rafael, E., Shohamy, E., Hasan Amara, M., & Trumper-Hecht, N. (2006). Linguistic Landscape as Symbolic Construction of the Public Space: The Case of Israel. *International Journal of Multilingualism*, 3(1), 7–30. <https://doi.org/10.1080/14790710608668383>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104/ASSET/CD05F7AD-79E4-455F-B905-CA634CFDF48D/ASSETS/001316446002000104.FP.PNG>
- Dama, I., & Hisasmaria. (2026). View of An Analysis of Abstract Translation by Pohuwato University Students: Translation Techniques, Quality, and Local Academic Register. *Jurnal Politik Dan Sosial Kemasyarakatan*, 18(1), 145–160. <https://e-jurnal.unisda.ac.id/index.php/MADANI/article/view/13251/4748>
- Estetika, M., & Gusthini, M. (2024). Analysis of Translation Techniques on Public Signs in Jabodebek Light Rail Transit (LRT). *Jurnal Ilmu Sosial Dan Humaniora*, 2, 381–396. <https://journal.ikmedia.id/index.php/jishum>
- Ghithrif, A. N., & Nugroho, R. A. (2025). Measuring the Translation Quality of Bilingual Commercial Web in Indonesian: The Urgency of Avoiding Machine Translation. *Journal of English Language Teaching and Linguistics*, 10(2), 317–334. <https://doi.org/10.21462/JELTL.V10I2.1554>
- Gorter, D., & Cenoz, J. (2023). A panorama of linguistic landscape studies. In *A Panorama of Linguistic Landscape Studies*. Multilingual Matters. <https://doi.org/10.21832/GORTER7144>

- Hazaea, A. N., Al-Sofi, B. B. M. A., & Alfaifi, A. (2024). Tourism representation and strategies on public signs: linguistic landscape of a Saudi southern tourist destination. *International Journal of Tourism Cities*, 10(1), 302–323. <https://doi.org/10.1108/IJTC-07-2023-0130>
- House, J. (2015). *Translation Quality Assessment: Past and Present*. Routledge.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Landry, R., & Bourhis, R. Y. (1997). Linguistic landscape and ethnolinguistic vitality: An empirical study. *Journal of Language and Social Psychology*, 16(1), 23–49. <https://doi.org/10.1177/0261927X970161002>
- Laras, S. (2021). The Unnaturalness Of Indonesian Public Signs Translation Into English In Medan. *Indonesian Journal of Education and Mathematical Science*, 2(3), 62–67. <https://doi.org/10.30596/IJEMS.V2I3.8101>
- Mardhiya, S. N., Sukaesih, I., & Sinaga, M. O. (2024). Verb translations in public signs from Jakarta's integrated transportation. *LinguAmerta*, 1(1). <https://doi.org/10.32722/JLA.V1I1.6811>
- Molina, L., & Albir, A. H. (2002). Translation techniques revisited: A dynamic and functionalist approach. *Meta*, 47(4), 498–512. <https://doi.org/10.7202/008033ar>
- Nababan, M., Nuraeni, A., & Sumardiono. (2012). Pengembangan Model Penilaian Kualitas Terjemahan. *Kajian Linguistik Dan Sastra*, 24(1), 39–57. <https://doi.org/10.23917/kl.v24i1.101>
- Napu, N. (2024). Linguistic landscapes in multilingual urban settings: Insights from translation perspectives. *Studies in English Language and Education*, 11(1), 530–548. <https://doi.org/10.24815/SIELE.V11I1.29559>
- Newmark, P. (1988). *A Textbook of Translation*. Prentice Hall.
- Nord, Christiane. (2026). *Translating as a purposeful activity: functionalist approaches explained*. Routledge. <https://www.routledge.com/Translating-as-a-Purposeful-Activity-Functionalist-Approaches-Explained/Nord/p/book/9781041159025>
- Nuraeni, A., Kusumastuti, F., & Nababan, M. R. (2016). A Translation Study on School Signboards in Surakarta: Types, Functions, And Quality. *Humaniora*, 28(2), 198–207. <https://doi.org/10.22146/JH.16402>
- Nurmala, D., Sopha, D., Ningsih, A. M., Kristiana, V., Kiki Wardana, M., Muslim, U., & Al-Washliyah, N. (2025). A LINGUISTIC LANDSCAPE IN THE DIGITAL ERA: A CASE STUDY IN MEDAN. *Teaching English and Language Learning English Journal (TELLE)*, 05(03), 434. <https://doi.org/10.36085/telle.v5i3.9398>
- Parthama, I. G. N., Giri, N. L. K. Y., & Setianingsih, N. K. A. I. (2025). Negotiating Language Policy in Tourist Spaces: A Linguistic Landscape Study of Bilingual Warning Signs in Bali. *Humanus: Jurnal Ilmiah Ilmu-Ilmu Humaniora*, 24(2), 248–258. <https://doi.org/10.24036/humanus.v24i2.133436>
- Prayogi, I. K. I., & Manipuspika, Y. S. (2025). Penerjemahan Rambu Lalu Lintas Bilingual di Bali: Antara Makna dan Konteks. *SPHOTA: Jurnal Linguistik Dan Sastra*, 17(2), 116–129. <https://doi.org/10.36733/SPHOTA.V17I2.11770>

- Sangseekaew, K. (2026). Error Analysis of English Public Signs in Tourist Attractions of Chiang Mai Province. *Humanities and Social Sciences Journal of Pibulsongkram Rajabhat University*, 20(1), 260–272. <https://doi.org/10.14456/PSRUHSS.2026.15>
- Simarmata, N. B., Pratiwi, Z., & Aripin, A. A. Z. (2025). From Javanese Suroboyoan to English: A Study of Cultural Terms, Translation Techniques, and Quality Assessment. *Buletin Al-Turas*, 31(2), 215–228. <https://doi.org/10.15408/BAT.V31I2.45712>
- Susini, M., Ana, I. W., & Sujaya, N. (2022, June 6). Shifts of Contextual Meanings in the Translation of Indonesian Public Signs. *Proceedings of the 1st Warmadewa International Conference on Science, Technology and Humanity (WICSTH 2021)*. <https://doi.org/10.4108/EAI.7-9-2021.2317741>
- Susini, M., Ana, I. W., & Sujaya, N. (2023). Adjustments in Indonesian-English Translation of Public Signs in Bali. *Journal of Language Teaching and Research*, 14(2), 368–374. <https://doi.org/10.17507/JLTR.1402.12>
- Susini, M., Sujaya, N., & Ana, I. W. (2021). Translation Alternatives of Indonesian Public Signs. *Theory and Practice in Language Studies*, 11(9), 1034–1040. <https://doi.org/10.17507/TPLS.1109.08>
- Syaia, S., Baharuddin, & Putera, L. J. (2026). Power and Translation: A Study of Multilingual Signs in Lombok. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 11(02), 213–227. <https://doi.org/10.23969/JP.V11I02.49432>
- Vinay, J.-P., & Darbelnet, J. (1995). *Comparative Stylistics of French and English: A methodology for translation* (J. C. Sager & M.-J. Hamel, Trans.) (J. C. Sager & M.-J. Hamel, Eds.). John Benjamins Publishing Company.
- Wang, M., & Zhou, J. (2025). Multilingual Healthcare Landscapes at a China's Border Hospital. *Asian Social Science*, 21(3), 39. <https://doi.org/10.5539/ass.v21n3p39>
- Zaman, M. N., Supriatnoko, & Alhilmy, M. N. (2025). Landscape lexical diversity for readability and safety mitigation on the commuter line bilingual information board. *E3S Web of Conferences*, 604, 07001. <https://doi.org/10.1051/E3SCONF/202560407001>
- Zec, S., Soriani, N., Comoretto, R., & Baldi, I. (2017). High Agreement and High Prevalence: The Paradox of Cohen's Kappa. *The Open Nursing Journal*, 11(1), 211. <https://doi.org/10.2174/1874434601711010211>