

Accuracy of DeepL Translation in *Inside Out 2* Movie Compared to Official Subtitles

 <https://doi.org/10.31004/jele.v10i3.877>

*Sri Wahyuni Hasibuan, Berlin Sibarani, Dolli Rotua Sinaga, Alfina Gustiany Siregar^{abcd} 

¹STAIN Mandailing Natal, ²Universitas Negeri Medan, ³Universitas Prima Indonesia, ⁴Universitas Medan Area, Indonesia.

Corresponding Author: sriwahyunihsb@stain-madina.ac.id

ABSTRACT

This study aims to describe the level of translation accuracy produced by DeepL in the film *Inside Out 2* and compared it with the official subtitles. Using a qualitative descriptive method, the researcher selected 30 segments from the film's dialogues that contain various expressions, idioms, and contextual meanings. The accuracy of DeepL translations was assessed using the translation quality assessment model developed by Nababan, Nuraeni, and Sumardiono (2012), which classifies translations into three levels: accurate, moderately accurate, and inaccurate. The results showed that most of DeepL's translations fell into the "moderately accurate" category with total 67%, 17% as accurate and 17% also as inaccurate level. This was due to literal translation, misinterpretation of idiomatic expressions, and lack of contextual adaptation. In contrast, the official subtitles demonstrated higher accuracy by delivering contextually appropriate and culturally sensitive translations. These findings suggested that while DeepL offers grammatically fluent output, it still lacks the semantic and pragmatic sensitivity required for high-quality subtitle translation in films.

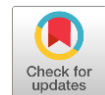
Keywords: *Machine Translation, DeepL, Subtitle, Translation Accuracy, Audiovisual Translation*

Article History:

Received 05th May 2025

Accepted 26th May 2025

Published 30th May 2025



INTRIDUCTION

The development of machine translation (MT) technology has had a significant impact on the way humans translate text, including in audiovisual contexts such as movie subtitles. One MT system that has grown rapidly in recent years is DeepL Translator, which relies on neural machine translation (NMT) technology. According to Kosem et al. (2022), DeepL often produces more natural and semantically accurate translations than other MT systems like Google Translate, especially for European languages. DeepL's architecture is based on a proprietary neural network that enables it to capture nuanced linguistic structures, which is critical in subtitle translation where conciseness and clarity are essential. DeepL is claimed to be able to produce translations that are more natural and closer to the quality of human translations than other systems such as Google Translate.

According to Groves and Mundt (2021) DeepL Translator provides more natural and fluent translations than Google Translate, especially in academic and idiomatic contexts. This refers to DeepL's ability to produce sentences that read more like those written by a human, using appropriate vocabulary, syntax, and idiomatic expressions, particularly in formal or complex texts such as academic writing. In academic contexts, clarity, formality, and accuracy are crucial. It is better at maintaining cohesion and coherence, avoiding literal or awkward phrasing often found in other MT outputs. Similarly, idiomatic expressions, phrases whose meanings are not deducible from the literal meanings of the words – are often mistranslated by systems like Google Translate.

However, although DeepL translations tend to be more grammatically fluent, the issue of meaning accuracy is still a big challenge, especially in texts that are rich in emotional and cultural nuances such as movie dialogs. Koehn (2020) asserts that neural machine translation has revolutionized the field by producing more fluent and context-aware translations, but accuracy and adequacy remain major concerns in domain-specific or culturally-loaded texts. This suggests that MT output, while appearing to flow linguistically, may not necessarily convey the right meaning according to context especially in the translation of highly contextualized movie subtitles.

Subtitle translation is a specific translation mode that represents one of the three main types of audiovisual translation, two others being dubbing and voice-over (Gottlieb, 2001). More practically, subtitle translation refers to the process of translating the spoken language in audiovisual works (like films or TV shows) into written text in another language, which appears as synchronized subtitles at the bottom of the screen to help the audience understand the content. In the world of audiovisual translation, especially in movie subtitles, accuracy is an important aspect that determines the integrity of the message from the source language to the target language. Subtitles not only function as a translation, but also as a medium that helps the audience understand the dialogue, cultural context, emotions, and humor conveyed by the characters in the film.

Cintas and Remael (2014) state that subtitling is a highly constrained form of translation that requires a balance between accuracy, readability, and synchronization with the image. This means that subtitling must pay attention to the limitations of display duration, character length, and visual relevance, while still maintaining accuracy of meaning.

In this context, accuracy in translation is defined as the extent to which the meaning of the source text can be transferred precisely and completely into the target text without omission, addition, or deviation of meaning. Nababan, Nuraeni, and Sumardiono (2012) define accuracy as the extent to which the meaning of the message in the source language can be transferred intact into the target language (p. 45). Thus, an accurate translation does not only transfer words, but also retains the message, nuance, and communication intent of the source text.

Based on the initial data from the Inside Out 2 movie subtitles, it can be seen that the translation of DeepL tends to be literal and does not adjust to the audiovisual context. For example, in the sentence "BRING IT IN!", DeepL translates it as "Bawa Masuk!" which sounds awkward and does not fit the context of the invitation to gather. Meanwhile, the official subtitle translates it more accurately as "Berkumpul", which fits the situation in the scene. Another example can be seen in the sentence "It should be nothing but smooth sailing from..." which DeepL translates to "Seharusnya tidak ada yang lain selain berlayar dengan lancar dari", while the official subtitle simplifies it to "semua akan berjalan lancar", which is more natural and easy to understand in Indonesian. This shows that the translation accuracy of DeepL is still problematic in capturing idiomatic and contextual meanings in movie subtitles.

The animated movie Inside Out 2 is an example of a movie full of idiomatic expressions, emotional conversations, and complex psychological contexts. Animated films, especially those aimed at families or children, tend to use language rich in idioms, metaphors and indirect expressions to convey messages in a creative and engaging manner. The speed of dialog and the nuances of language in this film make subtitling a challenge for MT systems such as DeepL. Therefore, it is important to evaluate whether DeepL is able to capture and convey meaning accurately, or if it produces translation errors such as mistranslation, omission, or addition that can interfere with audience understanding.

The research titled Accuracy Analysis of DeepL Translation in Inside Out 2 Compared to Official Subtitles focuses on this aspect of accuracy to compare the results of machine translation (DeepL) with official subtitles translated by humans. The research is to find out the translation accuracy level of Inside Out 2 movie subtitles generated by DeepL compared to the official subtitles and to describe the factors cause the difference in accuracy between the translation of DeepL and the official subtitles in the context of movie subtitles. Through this analysis, the research wants to know the extent to which DeepL is able to transfer the meaning of the dialog in the film completely and correctly, compared to the results of professional

translation. Given that the movie *Inside Out 2* contains many emotional expressions, idiomatics, and cultural contexts, the evaluation of accuracy is crucial

to assess whether DeepL can handle these nuances appropriately.

Accuracy in translation means reproducing in the receptor language the closest natural equivalent of the source-language message, first in terms of meaning and second in terms of style (Nida, . In translation studies, accuracy is the main aspect that shows the extent to which the message or meaning of the source language can be preserved intact in the target language. One of the most commonly used theories in assessing accuracy is the model developed by Nababan, Nuraeni, and Sumardiono (2012). They divided the level of accuracy into three categories, namely: (1) accurate, if the entire meaning of the message in the source language is correctly transferred into the target language; (2) moderately accurate, if most of the meaning is transferred well, but there are slight glitches or inaccuracies; and (3) inaccurate, if the meaning of the translation differs from the source language or there are serious errors in message transfer. This model is widely used in translation research because it is practical and applicable, including in the context of subtitle translation.

Table 1. Translation Accuracy Assessment Instrument (Based on Nababan et al., 2012)

Score	Category	Description
3	Accurate	The meaning of the source text is completely and correctly transferred into the target text. There is no distortion of meaning.
2	Moderately Accurate	The meaning of the source text is mostly transferred into the target text, but there are slight inaccuracies, distortions, or ambiguities that do not significantly hinder comprehension.
		The meaning of the source text is not accurately transferred. There are serious distortions, omissions, or additions that result in misinterpretation or loss of meaning.

METHOD

This research employs a descriptive qualitative method, aiming to analyze the level of accuracy in the Indonesian translation of *Inside Out 2* movie subtitles produced by DeepL Translator, by comparing it with the official human-translated subtitles. The focus of this study is on the transfer of meaning accuracy between the source language (English) and the target language (Indonesian), particularly in the context of idiomatic and contextual expressions frequently found in audiovisual media.

The primary method of data analysis is textual analysis, supported by a translation accuracy assessment framework. Through this method, the researcher systematically examines the quality of translation output, focusing on the extent to which the meaning from the original English subtitle is retained in both the machine-generated and human-produced translations.

Data and Data Sources

The data in this study consist of the original English subtitles (source text) from the movie *Inside Out 2*, the Indonesian translation generated by DeepL Translator, and the official Indonesian subtitles featured in the digital version of the film, as produced by a professional translation team. These data were collected by directly transcribing the dialogues from the original and translated versions, ensuring consistency in segment selection. The DeepL translations were generated manually by inputting the source text into the DeepL website to simulate typical user behavior. The official subtitles were extracted from the Indonesian digital release of the movie.

Data Collection Procedures

The research was conducted through a series of systematic steps to compare human and machine translations of the *Inside Out 2* movie subtitles. First, relevant dialogue segments were identified and transcribed from the original English subtitles. These segments were then translated into Indonesian using DeepL Translator. Simultaneously, the official Indonesian subtitles were retrieved from the film's authorized release. A comparative data table was compiled, consisting of three components: the source text (English), the official human translation, and the DeepL-generated machine translation. The corresponding segments from all three versions were aligned for direct comparison.

To assess translation quality, the study employed the accuracy scoring rubric developed by Nababan, Nuraeni, and Sumardiono (2012), which categorizes translations as accurate (score 3), moderately accurate (score 2), or inaccurate (score 1) based on the completeness and precision of meaning transfer. Each subtitle segment was analyzed both qualitatively and quantitatively to identify translation patterns and challenges. The analysis aimed to reveal the strengths and limitations of DeepL's machine translation in comparison to the professional human translation, leading to informed conclusions about translation accuracy and reliability.

Data Analysis Technique

Each subtitle segment was scored using the translation accuracy instrument mentioned above. The scores were then tabulated and interpreted to identify patterns of accuracy, common types of mistranslation, and recurring syntactic or semantic issues. In addition to score quantification, qualitative analysis was employed to provide detailed commentary on critical cases—particularly those involving idioms, cultural references, or emotion-laden language. The final analysis aims to describe the overall accuracy trend in DeepL's subtitle translation output and evaluate its reliability as a subtitling tool compared to human translation in terms of semantic precision, contextual sensitivity, and idiomatic appropriateness.

FINDINGS ND DISCUSSION

This study aims to evaluate the translation accuracy of subtitles in the movie *Inside Out 2*, focusing on a comparison between translations generated by DeepL (machine translation) and those produced by professional human translators (official subtitles). The evaluation framework is based on the model proposed by Nababan, Nuraeni, and Sumardiono (2012), which categorizes translation accuracy into three levels: Score 3 (Accurate), when the message is conveyed completely and precisely; Score 2 (Moderately Accurate), when most of the meaning is preserved but includes minor inaccuracies or ambiguities; and Score 1 (Inaccurate), when the translation is significantly flawed, missing essential meaning, or deviating from the original intent.

The analysis was conducted on 30 carefully selected subtitle segments that reflect a range of linguistic and contextual complexities, including idiomatic expressions, conversational language styles, complex grammatical structures, and culturally or emotionally charged content. Each segment was examined by comparing the English source text with its DeepL-generated translation and the corresponding human-translated official subtitle. This comparative approach allowed the researchers to assess how accurately each translation method conveyed meaning and to identify specific strengths and weaknesses in machine versus human translation performance. The findings show the following distribution of accuracy scores:

Table 2. Accuracy scores

Accuracy Score	Numbers of Segment	Percentage (%)
3 (accurate)	5	17%
2 (moderate accurate)	20	66%
1 (inaccurate)	5	17%
Total	30	100%

Most of the translations of DeepL fall into the moderately accurate category (66%), where the main meaning can still be understood despite minor distractions such as inappropriate diction or unnatural structures. However, 17% of the segments are inaccurate, mainly due to literal translation of idioms, cultural terms or figurative expressions. Only 17% of the segments were accurately translated by DeepL.

Based on the table above, here are some examples of representative segments from each accuracy category:

Score 3 (Accurate): The meaning of the message is transferred completely and precisely.

Data (1)

Source Text: *Joy is so old school.*

DeepL: *Joy sangat jadul.*

Official Subtitle: *Joy itu sangat kuno.*

In data (1) DeepL successfully uses the term “jadul” as an idiomatic equivalent of old school, which is culturally appropriate and common in Indonesian. Although the register is informal, the context is appropriate as the film is relaxed and the dialog is casual. This shows DeepL's ability to handle simple phrases that are already common.

Data (2)

Source Text: *This is Joy, coming to you live in Riley's mind, and we're expecting a GREAT CHAMPIONSHIP today with the Foghorns!*

DeepL: *Ini adalah Joy, yang hadir untuk Anda secara langsung di Riley, dan kami mengharapkan KEJUARAAN HEBAT hari ini dengan Foghorns!*

Official Subtitle: *Ini Joy, siaran langsung dari benak Riley dan semoga hari ini menjadi kejuaraan yang seru bagi Foghorns*

From the data (2) In this segment, DeepL is able to translate the meaning well. The sentence structure is more formal than the official subtitles, but there is no loss of meaning. The translation of the phrase “great championship” into “kejuaran hebat” is still appropriate and natural. Although DeepL's style is longer and more literal, the meaning of the message and communicative structure are still conveyed in full, making it accurate.

Score 2 (Moderately Accurate): The meaning is mostly transferred well, but there are minor inaccuracies.

Data (3)

Source Text: *Riley fans, get up on your feet, and make some NOISE!*

DeepL: *Para penggemar Riley, berdiri tegak, dan buatlah kegaduhan!*

Official Subtitle: *Penggemar Riley, berdiri dan beri sambutan meriah.*

Based on data (3) DeepL translates the sentence well enough, but the diction “kegaduhan” has a negative connotation that doesn't fit the positive context of the call to cheer. Meanwhile, the official subtitles choose “beri sambutan meriah” which is more idiomatic and natural. This reflects the limitations of DeepL in understanding emotional nuances and social registers.

Data (4)

Source Text (English): *Way down at the root level these memories were also creating beliefs.*

DeepL Translation: *Jauh di tingkat akar, ingatan ini juga menciptakan keyakinan.*

Official Subtitle: *Jauh di bawah pada tingkat dasar, memori juga menciptakan keyakinan.*

In this segment, the sentence "Way down at the root level these memories were also creating beliefs" contains a metaphorical expression – "root level" – which refers to a foundational or subconscious level of psychological development. The concept is abstract and contextual, requiring interpretive translation rather than literal rendering.

DeepL translates this as "Jauh di tingkat akar," which is a literal translation of "root level." While the phrase may be understandable, "tingkat akar" is not a common or natural expression in Indonesian and does not clearly reflect the metaphorical meaning intended in the source text. The rest of the sentence ("ingatan ini juga menciptakan keyakinan") is accurate and conveys the intended message correctly.

On the other hand, the official subtitle renders this phrase as "Jauh di bawah pada tingkat dasar," which is more idiomatic and pragmatically accurate in Indonesian. It successfully conveys the idea of depth and foundational development in a way that fits natural usage and

is contextually appropriate. The word "*tingkat dasar*" effectively replaces "*root level*" with a target-language equivalent that carries the same conceptual weight.

Score 1 (Inaccurate): The meaning is not appropriate or there are major errors in message transfer.

Data (5)

Source Text: *Bring it in!*

DeepL: BAWA MASUK!

Official Subtitle: *Berkumpul.*

Based on data (1) This is an example of literal translation that fails to convey the context of communication. "Bring it in" in the context of the movie is an invitation to the team or group to gather, not to literally "membawa sesuatu kedalam". The official subtitle captures this context and uses a more appropriate equivalent. DeepL fails to understand the pragmatic function of the sentence, hence the score of 1.

The findings of this study reveal a consistent pattern in the accuracy of subtitle translations produced by DeepL: while many of the translations retain the general meaning of the source text, they often lack contextual and cultural precision. This phenomenon is especially noticeable in segments involving idiomatic expressions, emotional tone, metaphorical language, and informal speech—elements that are prevalent in animated films like *Inside Out 2*. The DeepL translations, although grammatically correct and structurally coherent, tend to be literal and sometimes awkward, failing to capture the communicative intent of the original dialogue.

This issue arises from the fundamental limitations of machine translation systems, particularly those based on neural machine translation (NMT) like DeepL. As Koehn (2020) notes, NMT systems can generate fluent output but often struggle with semantic adequacy and contextual adaptation. DeepL does not "understand" the content it translates; rather, it predicts word sequences based on patterns in large data sets. As a result, it may accurately process general statements but falters when dealing with nuanced, domain-specific, or culture-bound content.

One example is the translation of the phrase "*Bring it in!*", which DeepL rendered as "*Bawa masuk!*". This literal interpretation ignores the pragmatic function of the utterance as a motivational group call (i.e., "gather around"). The official subtitle "*Berkumpul*" better represents the speaker's intent. Similarly, DeepL's translation of "*Way down at the root level*" as "*Jauh di tingkat akar*" is understandable but awkward and not idiomatically correct. In contrast, the human translator's version "*Jauh di bawah pada tingkat dasar*" shows a stronger awareness of natural target-language phrasing and metaphorical transfer.

These examples confirm House's (2015) theory that high-quality translation is not only about linguistic equivalence but also about functional and pragmatic adequacy. Human translators make interpretive decisions based on context, speaker tone, visual cues, and cultural relevance—all of which are essential in audiovisual translation but are beyond the reach of current machine translation capabilities.

Moreover, Nida's (1964) distinction between formal and dynamic equivalence explains the core difference observed: DeepL leans toward formal equivalence—translating words and structures—while human translators tend to prioritize dynamic equivalence, aiming for naturalness, effect, and audience reception. The greater accuracy found in the official subtitles underscores the importance of human judgment and cultural literacy in producing subtitles that resonate with viewers both linguistically and emotionally.

CONCLUSION

The analysis shows that DeepL's subtitle translations in *Inside Out 2* generally fall into the moderately accurate category (66%), with only 17% considered fully accurate, and another 17% rated as inaccurate. While DeepL performs adequately in translating literal and straightforward segments, it faces significant challenges when dealing with idiomatic expressions, cultural references, and metaphorical language, often leading to distortions in meaning. In contrast, the official human-translated subtitles consistently deliver higher accuracy, successfully capturing not only the literal meaning but also the pragmatic intent and

cultural nuances of the original dialogue. These findings reinforce the conclusion that machine translation tools like DeepL, while useful for basic translation tasks, are not yet fully capable of handling the complexity required in audiovisual translation.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Prof. Berlin Sibarani, lecturer of the *Trends in Translation* course, for his guidance, insight, and encouragement throughout the completion of this research. I also extend my thanks to my fellow classmates for their support, discussions, and collaboration during this academic journey.

REFERENCES

- Díaz Cintas, J., & Remael, A. (2014). *Audiovisual Translation: Subtitling* (2nd ed.). London: Routledge.
- Gottlieb, H. (2001). *Anglicism and TV subtitling in an anglicized world*. Amsterdam: John Benjamins
- Groves, M., & Mundt, K. (2021). Exploring the usability of DeepL Translator for academic writing. *Journal of English for Academic Purposes*, 52, 100976. <https://doi.org/10.1016/j.jeap.2021.100976>
- House, J. (2015). *Translation Quality Assessment: Past and Present* (2nd ed.). London: Routledge
- Kamaluddin, M. I., et al. (2024). *Accuracy Analysis of DeepL: Breakthroughs in Machine Translation Technology*. *Journal of English Education and Foreign Languages*, 4(2), 123–135.
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge: Cambridge University Press.
- Kosem, I., Lew, R., & Müller-Spitzer, C. (2022). Comparing machine translation tools: A case study of DeepL and Google Translate. *Lexikos*, 32(1), 1–21.
- Moleong, L. J. (2017). *Metodologi Penelitian Kualitatif (Edisi Revisi)*. Bandung: Remaja Rosdakarya.
- Saldanha, G., & O'Brien, S. (2013). *Research Methodologies in Translation Studies*. Manchester: St. Jerome Publishing.
- Mutia Asmara, D. S., & Repelitawati Br Kembaren, F. (2024). *Student's Perception Towards the Use of DeepL Translator in Writing Thesis or Journal for English Education Students*. *International Journal of Language Education and Cultural Review*, 10(1), 45–58.
- Nababan, M. R., Nuraeni, A., & Sumardiono. (2012). Pengembangan model penilaian kualitas terjemahan. *Kajian Linguistik dan Sastra*, 24(1), 39–57.
- Nida, E. A. (1964). *Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: E. J. Brill.
- Sun, R. (2024). *Evaluating the Translation Accuracy of ChatGPT and DeepL Through the Lens of Implied Subjects*. *Arab World English Journal for Translation & Literary Studies*, 8(4), 41–53.