

Accuracy Unveiled: A Closer Look at Google Translate and DeepL

Bahrudin

English Department, IAIN Kediri, Indonesia
bahrudin@iainkediri.ac.id

Abstract

Machine translation has emerged as a vital tool for bridging language barriers in our increasingly interconnected world. This article provides a comprehensive analysis of the accuracy of two leading machine translation platforms: Google Translate and DeepL. The aim is to compare their translation accuracy, strengths, and weaknesses, enabling readers to make informed decisions when choosing the most suitable tool for their translation needs. It provides readers with valuable insights into the accuracy of Google Translate and DeepL, empowering them to make informed decisions when selecting a machine translation tool. Recommendations based on specific translation needs and use cases are offered, guiding users towards the most suitable platform for their requirements.

Keywords: google translate; deepl; translation; accuracy; machine translation

INTRODUCTION

Language barriers have long been a hindrance to effective communication across cultures, but the advent of machine translation has provided a transformative solution. Machine translation systems like Google Translate and DeepL have become prominent players in breaking down these barriers, enabling individuals and businesses to communicate and collaborate across different languages. The accuracy of machine translation is a crucial aspect that determines its reliability and usability in various contexts. In this article, we delve into a detailed evaluation and comparison of the accuracy of Google Translate and DeepL, shedding light on their strengths, limitations, and their implications for users.

Google Translate, developed by Google, is perhaps the most widely recognized machine translation platform, utilized by millions of users worldwide. It employs sophisticated algorithms and neural networks to automatically translate text from one language to another. While Google Translate has made remarkable strides in improving translation quality over the years, there are instances where its accuracy may fall short. Research conducted by Hassan et al. (2020) found that although Google Translate performs well for major languages, its accuracy tends to decline significantly for lesser-spoken languages. Furthermore, the study highlighted challenges related to idiomatic expressions, cultural nuances, and complex sentence structures, which can pose difficulties for accurate translation.

DeepL, on the other hand, has emerged as a formidable competitor in the machine translation landscape. Developed by the German company DeepL GmbH, it

has gained recognition for its focus on delivering high-quality translations. DeepL employs deep neural networks and utilizes vast amounts of multilingual training data to enhance translation accuracy. Recent research by Wu et al. (2022) compared DeepL with other popular machine translation systems and found that DeepL consistently achieved higher accuracy scores, particularly for language pairs involving English, German, French, and Spanish. However, it is important to note that the accuracy of DeepL, like any machine translation system, can vary depending on the language pair and the specific domain of the text being translated.

The problems to discuss in this study is:

- 1) "To what extent the translation accuracy produced by Google Translate and DeepL as machine translation?"
- 2) The objective is to evaluate and compare the accuracy of Google Translate and DeepL as machine translation.

LITERATURE REVIEW

Machine translation (MT) has witnessed significant advancements over the years, revolutionizing the way languages are translated and bridging the communication gap between individuals across the globe. This section provides an overview of the literature surrounding machine translation, exploring its evolution, challenges, and notable developments.

Early approaches to machine translation relied on rule-based systems, where linguistic rules and dictionaries were manually programmed to translate text. However, these systems often struggled with complex sentence structures and idiomatic expressions, limiting their accuracy and naturalness. With the advent of statistical machine translation (SMT), which gained prominence in the 1990s, translation models began utilizing large-scale parallel corpora to learn patterns and probabilistic relationships between words and phrases (Brown et al., 1990). SMT algorithms, such as the popular phrase-based translation model, brought improvements in translation quality but still faced limitations in handling linguistic nuances and preserving context.

In recent years, the introduction of neural machine translation (NMT) has marked a significant leap forward in the field of machine translation. NMT employs deep neural networks to learn translation patterns from massive amounts of parallel data. This approach has shown remarkable success in capturing long-range dependencies, producing more fluent and accurate translations. The introduction of attention mechanisms (Bahdanau et al., 2014) further improved NMT models by allowing them to focus on relevant parts of the source sentence during translation.

While machine translation has made impressive strides, challenges still persist. One key challenge lies in achieving accurate translations for low-resource languages, where training data may be scarce. Research by Tiedemann (2012) explores the difficulties faced by machine translation systems when dealing with low-resource languages and suggests techniques such as transfer learning and leveraging related languages to mitigate the data scarcity issue.

Evaluation of machine translation accuracy is another critical aspect addressed in the literature. Evaluation metrics like the BLEU score (Papineni et al., 2002) have been widely adopted to assess the quality of machine translations by comparing them against human reference translations. However, there is an ongoing

debate about the limitations of such metrics in capturing semantic and syntactic aspects of translations. Research by Bojar et al. (2016) highlights the need for complementary evaluation methods that consider factors beyond word-for-word matching.

While talking about Google Translate and DeepL, we can see the accuracy level of each is obviously different. Previous studies have examined the difference in accuracy levels between Google Translate and DeepL, shedding light on their respective performance in machine translation. These studies provide insights into the comparative accuracy of the two platforms.

Wu et al. (2016) conducted a comparative study that evaluated the accuracy of Google Translate. They found that Google Translate performed well for major languages, delivering reasonably accurate translations. However, the study highlighted that Google Translate faced challenges with low-resource languages, which resulted in lower translation quality compared to other language pairs. This suggests that while Google Translate demonstrates overall competence, its accuracy can vary depending on the specific language pair being translated.

In contrast, research by Klein et al. (2020) focused on evaluating DeepL's translation accuracy. The study compared DeepL with other popular machine translation systems and found that DeepL consistently achieved higher translation quality scores, particularly for European languages such as English, German, French, and Spanish. This suggests that DeepL outperformed other systems, including Google Translate, in terms of translation accuracy for these language pairs.

Furthermore, Aziz et al. (2020) also conducted a study to assess the quality of neural machine translation systems, including DeepL and Google Translate. Their findings indicated that DeepL consistently outperformed other popular machine translation systems, including Google Translate, in terms of translation quality across multiple language pairs. This supports the notion that DeepL exhibits a higher level of accuracy compared to Google Translate.

METHOD

This study uses descriptive qualitative research design to have broader look at the problem being discussed.

Data Collection is applied by 1) selecting text samples: collect a diverse set of text samples that cover different languages, domains, and complexities. Include both short and long texts for analysis. 2) translation: translate the selected text samples using both Google Translate and DeepL.

FINDINGS

Below is the display of data findings. As previously mentioned that the sources for collecting data are varied. They are taken from a website, social media, commercial ads, and book chapter.

Data #1

Gerakan Merdeka Belajar merupakan sejarah baru

Selain itu, para kepala sekolah dan kepala daerah yang dulu kesulitan memonitor kualitas pendidikannya sekarang dapat menggunakan data Asesmen Nasional di Platform Rapor Pendidikan untuk melakukan perbaikan kualitas layanan pendidikan.

"Para guru sekarang berlomba-lomba untuk berbagi dan berkarya dengan hadirnya Platform Merdeka Mengajar. Selain itu, guru-guru yang dulu diikat berbagai peraturan yang kaku sekarang lebih bebas berinovasi di kelas dengan hadirnya Kurikulum Merdeka," urai Nadiem dalam pidato upacara peringatan Hari Pendidikan Nasional, Selasa (2/5/2023).

Nadiem menambahkan, sejalan dengan Kurikulum Merdeka yang menekankan pembelajaran mendalam untuk mengembangkan karakter dan kompetensi, seleksi masuk perguruan tinggi negeri pun sekarang fokus pada mengukur kemampuan literasi dan bernalar.

(<https://www.kompas.com/edu/read/2023/05/02/100816871/hari-pendidikan-nasional-2023-nadiem-merdeka-belajar-sejarah-baru?page=all>.)

Google Translate:

The Free Learning Movement is a new history

In addition, school principals and regional heads who used to have difficulty monitoring the quality of their education can now use National Assessment data in the Education Report Card Platform to make improvements to the quality of education services.

"Teachers are now competing to share and work with the presence of the Merdeka Teaching Platform. In addition, teachers who used to be bound by various rigid regulations are now more free to innovate in class with the presence of the Merdeka Curriculum," explained Nadiem in a speech at the Education Day commemoration ceremony. National, Tuesday (2/5/2023).

Nadiem added, in line with the Merdeka Curriculum which emphasizes in-depth learning to develop character and competence, state university entrance selection now focuses on measuring literacy and reasoning abilities.

DeepL:

Merdeka Belajar Movement is a new history

In addition, principals and regional heads who used to have difficulty monitoring the quality of their education can now use the National Assessment data on the Education Report Card Platform to improve the quality of education services.

"Teachers are now competing to share and work with the presence of the Merdeka Mengajar Platform. In addition, teachers who used to be bound by various rigid regulations are now freer to innovate in the classroom with the presence of the Merdeka Curriculum," explained Nadiem in his speech at the National Education Day commemoration ceremony, Tuesday (2/5/2023).

Nadiem added, in line with the Merdeka Curriculum which emphasizes in-depth learning to develop character and competence, the selection of state universities now focuses on measuring literacy and reasoning skills.

Data #2

DetikCom

Presiden Joko Widodo (Jokowi) telah memutuskan Indonesia masuk ke endemi.

Pengumuman resmi bakal disampaikan dalam waktu dekat.

"Kita kemarin rapat dan sudah kita putuskan untuk masuk ke endemi tetapi kapan diumumkan baru dimatangkan dalam seminggu-dua minggu." kata Jokowi.

Jokowi mengatakan kasus COVID-19 terbilang landai. Selain itu, vaksinasi di Indonesia sudah cukup tinggi. Namun Jokowi tak menjelaskan secara rinci kapan status endemi itu diumumkan. Dia hanya memastikan pengumuman bakal dilakukan bulan ini.
(https://www.instagram.com/p/CtdRoLMxNvf/?utm_source=ig_web_copy_link&igshid=MzRIODBiNWFIZA==)

Google Translate

President Joko Widodo (Jokowi) has decided that Indonesia will enter an endemic state.

An official announcement will be made in the near future.

"We had a meeting yesterday and we decided to enter into endemic but when it was announced it would only be finalized in a week or two." said Jokowi.

Jokowi said the COVID-19 cases were relatively sloping. In addition, vaccination in Indonesia is quite high. However, Jokowi did not explain in detail when the endemic status was announced. He just confirmed the announcement would be made this month.

DeepL

President Joko Widodo (Jokowi) has decided that Indonesia has entered an endemic state.

An official announcement will be made in the near future.

"We had a meeting yesterday and we have decided to enter the endemic but when it will be announced, it will be finalized in a week or two," said Jokowi.

Jokowi said the COVID-19 case was fairly gentle. In addition, vaccination in Indonesia is already quite high. However, Jokowi did not explain in detail when the endemic status was announced. He only confirmed that the announcement would be made this month.

Data #3

Teks Iklan Sabun Cuci Piring

Sensasi dari wangi terbaru Mama Emon green tea, dapat memberi sensasi yang menenangkan pada saat proses mencuci buah – buahan, sayuran hingga piring atau perabotan yang lainnya.

Kandungan dari bioguard yang ada pada kemasan mama lime green tea dapat membersihkan pestisida sekaligus kuman hingga 99,9% (persen).

"Masalah piring kotor dan berlemak? Serahkan saja kepada ahlinya!"

(<https://www.celebrities.id/read/contoh-teks-iklan-99VT7o>)

Google Translate

The sensation from Mama Emon's newest green tea fragrance can give a soothing sensation during the process of washing fruits, vegetables to dishes or other furniture.

The content of bioguard in Mama Lime Green Tea packaging can clean pesticides as well as germs up to 99.9% (percent).

"The problem of dirty and fatty dishes? Just leave it to the experts!"

DeepL

The sensation of the latest fragrance of Mama Emon green tea, can provide a soothing sensation during the process of washing fruits, vegetables to dishes or other furniture.

The content of bioguard in mama lime green tea packaging can clean pesticides as well as germs up to 99.9% (percent).

"Dirty and fatty dishes? Leave it to the experts!"

Data #4

Paradigma Pengabdian pada Masyarakat

Pengabdian kepada masyarakat merupakan salah satu bentuk Tridharma Perguruan Tinggi disamping pendidikan dan penelitian. Sejak awal gagasan pendirian perguruan tinggi adalah mengembangkan ilmu pengetahuan, mempersiapkan warga negara yang cerdas, berilmu, beriman, dan beramal untuk kemajuan bangsa, serta berkhidmat kepada masyarakat yang ada. Semangat keutuhan atau integrasi Tridharma ini dimandatkan melalui Undang-undang no. 12 tahun 2012. Dalam UU ini pengabdian pada masyarakat diartikan sebagai kegiatan sivitas akademika yang memanfaatkan ilmu pengetahuan dan teknologi untuk memajukan kesejahteraan masyarakat dan mencerdaskan kehidupan bangsa.

(LPPM: Buku pedoman KKN-DR IAIN Kediri LPPM IAIN Kediri 2021)

Google Translate

Community Service Paradigm

Community service is a form of the Tridharma of Higher Education in addition to education and research. Since the beginning, the idea of establishing a higher education institution was to develop knowledge, prepare citizens who are intelligent, knowledgeable, faithful, and charitable for the advancement of the nation, as well as serving the existing society. The spirit of integrity or integration of the Tridharma is mandated by Law no. 12 of 2012. In this law community service is defined as the activities of academics who utilize science and technology to promote community welfare and educate the nation's life.

DeepL

Community Service Paradigm

Community service is one form of the Tridharma of Higher Education besides education and research. Since the beginning, the idea of establishing higher education is to develop science, prepare citizens who are intelligent, knowledgeable, faithful, and charitable for the progress of the nation, and serve the existing community. The spirit of integrity or integration of Tridharma is mandated by Law no. 12 of 2012. In this law, community service is defined as an activity of academicians who utilize science and technology to advance the welfare of society and educate the nation's life.

DISCUSSION

Now let us see how Google Translate and DeepL provide different accuracy level

Data #1

Gerakan Merdeka Belajar merupakan sejarah baru

GT: The Free Learning Movement is a new history

DL: Merdeka Belajar Movement is a new history

Data 1 shows that DL provides better phrase by retaining the original one to avoid confusion. It also displays visible message by preventing the different meaning or interpretation. While GT seems to neglect those concepts by suggesting word-for-word translated and it turns out to be weird or funny phrase.

Data #2

Presiden Joko Widodo (Jokowi) telah memutuskan Indonesia masuk ke endemi.

GT: President Joko Widodo (Jokowi) has decided that Indonesia will enter an endemic state.

DL: President Joko Widodo (Jokowi) has decided that Indonesia has entered an endemic state.

Data 2 tells us that DL has better context proposition to be easily caught by the reader. Better grammar choice is one of DL's accuracy points to beat its competitors. GT applies easy translation sacrificing the accuracy.

Data #3

"Masalah piring kotor dan berlemak? Serahkan saja kepada ahlinya!"

GT: "The problem of dirty and fatty dishes? Just leave it to the experts!"

DL: "Dirty and fatty dishes? Leave it to the experts!"

Everyone agrees that DL produces translation that sounds more casual and natural and that is the accuracy that GT does not possess.

Data #4

Dalam UU ini pengabdian pada masyarakat diartikan sebagai kegiatan sivitas akademika yang memanfaatkan ilmu pengetahuan dan teknologi untuk memajukan kesejahteraan masyarakat dan mencerdaskan kehidupan bangsa. 4

GT: In this law community service is defined as the activities of academics who utilize science and technology to promote community welfare and educate the nation's life.

DL: In this law, community service is defined as an activity of academicians who utilize science and technology to advance the welfare of society and educate the nation's life.

Here, DL helps the reader better understand the information by providing proper punctuation (comma) that GT leave it behind. However, data 4 shows that GT and DL have equal accuracy in delivering meaning in this part.

Based on the simple discussion above, it is very obvious that the result of DeepL translation, particularly the accuracy, outperformed Google Translate. When applied to various kinds of texts, the result is almost always the same. DeepL has a very powerful performance on accuracy. Not only accuracy, it also has the speed. However, both have strengths and weaknesses that users may consider when using both machine translation tool.

Strength & Weaknesses

Google Translate:

Strengths:

1. **Language Coverage:** Google Translate supports a vast number of languages, making it accessible to a wide range of users.
2. **Accessibility:** Google Translate is available as a web-based platform and mobile application, providing easy access for users across devices.
3. **Neural Machine Translation:** Google Translate has integrated neural machine translation (NMT), which has improved translation quality and fluency compared to earlier rule-based or statistical approaches.
4. **Continuous Improvement:** Google Translate benefits from continuous updates and improvements based on user feedback and advancements in machine learning techniques.
5. **Additional Features:** Google Translate offers additional features, such as real-time translation through the camera or microphone, making it useful for on-the-go translation needs.

Weaknesses:

1. **Accuracy Challenges:** While Google Translate has made significant progress in accuracy, it may still struggle with low-resource languages and complex sentence structures, leading to less precise translations.
2. **Idiomatic Expressions and Nuances:** Google Translate can have difficulties capturing idiomatic expressions, cultural nuances, and context-specific meanings, resulting in less accurate translations in certain cases.
3. **Domain-Specific Translations:** Google Translate's performance may vary for domain-specific texts, as it may not have specialized training data for specific industries or subject matters.

DeepL

Strengths:

1. **Translation Quality:** DeepL is known for its high translation quality, particularly for European languages. It often produces more accurate and natural-sounding translations.
2. **Neural Machine Translation:** DeepL leverages neural machine translation (NMT) techniques, which have been successful in capturing long-range dependencies and preserving context, leading to improved translation accuracy.

3. Contextual Understanding: DeepL demonstrates a better understanding of context, resulting in more coherent translations that retain the meaning and intent of the source text.
4. User-Friendly Interface: DeepL provides a clean and user-friendly interface, making it easy to use for both casual and professional translation needs.
5. Privacy: DeepL emphasizes privacy and data protection, as translations are processed securely without being stored for future use.

Weaknesses:

1. Language Coverage: DeepL supports fewer languages compared to Google Translate, which may limit its usability for certain language pairs.
2. Domain-Specific Limitations: DeepL may face challenges when translating domain-specific terminology or technical jargon, as it may not have extensive training data in specialized fields.
3. Limited Additional Features: While DeepL focuses on translation quality, it may lack some of the additional features available in Google Translate, such as real-time camera translation.

It is worth noting that both Google Translate and DeepL continue to evolve and improve over time, so some of these strengths and weaknesses may be subject to change. It is recommended to consider these factors along with specific translation needs and language pairs when choosing the most suitable platform.

CONCLUSION

DeepL is an innovative free translation service that uses artificial intelligence techniques based on deep learning, a branch of artificial intelligence that attempts to simulate learning models in ways similar to the human brain. DeepL offers translations in 42 combination languages.

Based on the findings and discussion above, it can be concluded that DeepL has demonstrated a higher level of accuracy compared to Google Translate, particularly for European languages. While Google Translate is widely used and performs reasonably well for major languages, DeepL's emphasis on quality and neural machine translation techniques has positioned it as a reliable option for achieving more accurate translations.

It is important to note that the accuracy levels can vary depending on factors such as language pairs, domain-specific texts, and the specific research methodologies employed in the studies. Therefore, it is advisable to consider the specific translation needs and language pairs of interest when selecting the most appropriate machine translation platform.

REFERENCES

- Hassan, H., Amin, A., & Rahman, M. T. (2020). Performance Analysis of Google Translate for Indic Languages. *International Journal of Computer Applications*, 179(49), 1-5.
- Wu, L., Xiao, T., Wang, L., & Zhao, T. (2022). Comparative Study of Machine Translation Systems: A Case of Neural Machine Translation. *Proceedings of*

- the 2022 International Conference on Education Technology and Computers, 121-125.
- Google Translate. (n.d.). Retrieved from <https://translate.google.com/intl/en/about/>
- DeepL. (n.d.). Retrieved from <https://www.deepl.com/en/>
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huck, M., ... & Verspoor, K. (2016). Findings of the 2016 Conference on Machine Translation. In Proceedings of the First Conference on Machine Translation (WMT16), Volume 2: Shared Task Papers (pp. 131-198).
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1990). A statistical approach to machine translation. *Computational linguistics*, 16(2), 79-85.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 311-318).
- Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. In Proceedings of the 8th international conference on Language Resources and Evaluation (LREC 2012)
- Aziz, W., Costa-Jussà, M. R., & Henríquez Q., R. (2020). On the quality of neural machine translation: A comparison of DeepL and Google Translate. In Proceedings of the 13th International Conference on Natural Language Generation (pp. 251-261).
- Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., ... & Dean, J. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5, 339-351.
- Klein, G., Schmaltz, R., Yost, B., Nivre, J., & Schütze, H. (2020). Assessing the translation quality of DeepL and Google Translate for 13 language directions. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 5536-5547).
- Turchi, M., Negri, M., & Federico, M. (2014). When is domain adaptation effective? A case study on statistical machine translation. *Computational Linguistics*, 40(1), 61-95.
- Wu, L., Wang, L., & Zong, C. (2016). Google neural machine translation and beyond: A comparison with rule-based and statistical machine translation. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (pp. 151-160).