

INTISARI

Mesin penerjemah kini telah berkembang semakin pesat dengan temuan teknologi komputer yang disebut kecerdasan buatan (*artificial intelligence*). Mesin penerjemah masa kini menghasilkan terjemahan teks yang mendekati hasil terjemahan manusia. Dengan kemajuan tersebut, manusia semakin dimudahkan dan semakin banyak manusia yang menggunakan mesin penerjemah untuk kebutuhan menerjemahkan teks.

Meski dapat menghasilkan terjemahan yang mendekati natural, mesin penerjemah masih terus dievaluasi dan hasil evaluasi membuktikan bahwa terdapat kesalahan dalam hasil terjemahan mesin. Dalam penelitian ini, peneliti menganalisis kesalahan hasil terjemahan mesin. Mesin penerjemah yang dievaluasi ialah DeepL dan ChatGPT. Kedua mesin tersebut merupakan mesin penerjemah yang relatif baru di dunia mesin penerjemah, namun juga merupakan mesin penerjemah yang digunakan oleh banyak orang yang diantaranya adalah kaum pemelajar.

Dalam menganalisis kesalahan hasil terjemahan mesin, digunakan teori analisis kesalahan hasil terjemahan mesin milik Koponen (2010). Analisis kesalahan dilakukan oleh 5 evaluator yang hasilnya disimpulkan dengan menggunakan metode *percent agreement interrater reliability*. Selain itu juga digunakan matriks evaluasi mesin penerjemah otomatis *Bilingual Language Evaluation Understudy*, yang digunakan sebagai pembandingan dalam menyimpulkan hasil evaluasi kedua mesin penerjemah. Melalui penelitian ini disimpulkan bahwa mesin penerjemah DeepL memiliki jumlah kesalahan yang lebih rendah dibandingkan dengan mesin penerjemah ChatGPT. Skor BLEU juga menunjukkan bahwa skor hasil terjemahan DeepL lebih tinggi dibandingkan dengan skor BLEU ChatGPT.

Kata kunci: analisis kesalahan, evaluasi mesin penerjemah, BLEU, DeepL, ChatGPT

ABSTRACT

Machine translation has now developed significantly with the invention of computer technology known as artificial intelligence. Today's machine translation produces text translations that are close to human translations. With these advances, humans are increasingly facilitated and more and more people are using machine translation for translation needs.

Despite being able to produce near-natural translations, machine translation is still being evaluated and the evaluation results prove that there are errors in machine translation results. In this study, we analyzed machine translation errors. The machine translations evaluated are DeepL and ChatGPT. Both machines are relatively new in the world of machine translation, but they are also machine translations that are used by many people, including learners.

In analyzing machine translation errors, Koponen's (2010) theory of machine translation error analysis was used. Error analysis was conducted by 5 evaluators whose results were summarized using the percent agreement interrater reliability method. In addition, the automatic machine translation evaluation matrix of Bilingual Language Evaluation Understudy was also used, which was used as a comparison in summarizing the evaluation results of the two translation machines. Through this study, it is concluded that the DeepL machine translation has a lower number of errors compared to the ChatGPT machine translation. The BLEU score also shows that the DeepL translation score is higher than the ChatGPT BLEU score.

Keywords: error analysis, machine translation evaluation, BLEU, DeepL, ChatGPT