Direct Machine Translation Indonesian-Batak Toba

Zaenal Abidin
Doctoral Program of Mathematics and Natural Sciences
Universitas Lampung
Bandarlampung, Indonesia
2237061006@students.unila.ac.id

Akmal Junaidi
Department of Computer Science
Universitas Lampung
Bandarlampung, Indonesia
akmal.junaidi@fmipa.unila.ac.id

Wamiliana
Department of Mathematics
Universitas Lampung
Bandarlampung, Indonesia
wamiliana.1963@fmipa.unila.ac.id

Frengky Marulitua Togatorop
Faculty of Engineering and Computer Science
Universitas Teknokrat Indonesia
Bandarlampung, Indonesia
frengkytogatorop79@gmail.com

Imam Ahmad
Faculty of Engineering and Computer Science
Universitas Teknokrat Indonesia
Bandarlampung, Indonesia
imamahmad@teknokrat.ac.id

Ajeng Savitri Puspaningrum
Faculty of Engineering and Computer Science
Universitas Teknokrat Indonesia
Bandarlampung, Indonesia
ajeng.savitri@teknokrat.ac.id

2023 7th International Conference on New Media Studies (CONMEDIA) | 979-8-3503-0750-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONMEDIA60526.2023.10428332

***Abstract*— The translation application was developed using Python and Visual Studio Code as the primary programming environment. Technological progress has led to the younger generation forsaking the Batak language in their daily interactions, perceiving it as inconsequential to their social circle and contributing to the decline of regional languages. There is concern that regional languages, namely the Batak language, may lose significance or face extinction if this trend persists. This translation program is being implemented to aid in preserving regional languages, specifically the Batak language, enabling it to conform to contemporary circumstances. The development of this translation engine utilized the Batak language dictionary. The process of designing, coding, and testing is responsible for creating this direct machine translation. The BLEU score is employed to assess the caliber of machine translation. The study's findings indicate that the % BLEU value for ten sentences was 32.3%. This score falls within the category of sentences whose meaning can be understood, but the sentence structure still needs development. The direct machine translation developed in this Web platform is highly proficient and practical.**

Keywords—batak language; direct machine translation; BLEU; python; translator

I. INTRODUCTION

Residents utilize Indigenous languages in a certain nation region for their everyday interactions. Furthermore, the presence of a local language is a defining attribute of ethnic communities within a certain geographical area, setting them apart from ethnic groups in other regions. Indonesia has over 700 regional languages characterized by distinct vocabulary, speakers, dialects, areas, and cultural settings [1]. The Batak tribe constitutes the predominant portion of the people of North Sumatra. The Batak tribe has six categories: Toba Batak, Karo Batak, Pakpak Dairi Batak, Simalungun Batak, Mandailing Batak, and Angkola Batak [2]. Nur Hasanah Hrp et al. have researched text pre-processing, specifically focusing on stemming in the Angkola Batak language [2]. However, they still need to research machine translation. The findings of the stem analysis conducted by Nur Hasanah Hrp et al. will prove highly valuable for studying natural language processing in the Angkola Batak language. No research on stemming has been discovered in the Toba Batak language. The research on other Angkola Batak languages was undertaken using the Enhanced Confix Stripping technique. The research findings identify affixed words in papers written in the Angkola Batak language [3].

The Batak language is part of the Austronesian language family. Batak Toba is a linguistic dialect commonly used near Lake Toba and its environs. Samosir, Humbang Hasundutan, North Tapanuli, and Toba Samosir are constituent regions of the Batak Toba language area. The Batak language is currently employed for everyday communication in rural areas of North Sumatra, with the Batak Toba language being particularly prominent. The contemporary world also influences the survival of the Batak language. The younger generation today feels ashamed to speak the Batak language since it is seen as being outdated. One of the challenges facing the expansion of the Batak Toba language. Moreover, numerous factors contribute to the decline of the Batak language, such as the child's inadequate comprehension of the Batak Toba language, parents' failure to educate their children about the Batak Toba language, the impact of the surrounding environment, the shortcomings of specific groups or organizations in teaching local languages, and the influence of Indonesian and foreign languages. The Batak language is being marginalized in the face of technological advancements in the modern world. The validity of the Batak language diminishes when elements from another language are incorporated into it. An effective approach to address the abovementioned issue involves employing Natural Language Processing technology using the Toba Batak language dictionary.

The dictionary functions as a translation tool and written record, providing a reliable source for preserving the integrity and longevity of the Batak language. Nevertheless, dictionaries lose their attraction as technology advances. This is due to dictionaries' need to offer rapid and immediate translation. In order to locate a specific word, it is necessary to ascertain the

significance of a sentence initially, examine individual words sequentially and then organize them accordingly. This renders the dictionary less efficient in the current era of technological progress. The book "Machine Translation" by Indian researchers highlights the development of machine translation technology through three approaches: Direct Machine Translation (DMT) with a reference bilingual dictionary, Rule-based Machine Translation (RBMT) with a set of rules, and data-driven technology using a parallel bilingual corpus [4]. The primary element of the DMT experiment is the Indonesian-Batak Toba multilingual lexicon. The primary objective of this study is to gain proficiency in translating from Indonesian to Toba Batak.

The dictionary is employed as a repository in the development of DMT. One example of research that utilizes the Lampung language as the subject is the study of speech processing in the Lampung language [5]. Study on the translation of Kannada to Telugu using DMT [6]. Additional research has been conducted on DMT translation from Pali to Sinhala in Sri Lanka [7]. DMT The research conducted in Indonesia utilizes Android technology to translate from Indonesian to Balinese [8], the outcome of this research is an Android application. Subsequently, the DMT translation from Indonesian to Javanese was performed utilizing a mobile device [9]. Statistical Machine Translation (SMT) research encompasses various studies, such as the translation of Javanese into Indonesian using SMT with phrases, the translation of Sundanese into Indonesian using SMT with phrases and based on part of speech (PoS) Tag [10], the translation of English documents into Indonesian documents using SMT and word and phrase rearrangement [10], and the analytical exploration of language models using SMT for Indonesian-Dayak Kanayatn [11]. The study explores the relationship between two related research approaches: Neural Machine Translation (NMT) and Attention-based NMT [12]. Both approaches aim to translate from the Lampung language dialect of api to Indonesian language using a parallel corpus of 3000 sentences. Statistical Machine Translation (SMT) was employed to translate from the Lampung language dialect of api to the Indonesian language, utilizing a parallel corpus of 3000 sentences [13]–[16]. No study on Indonesian-Batak Toba has been completed using either SMT or NMT techniques.

The objective of this study is to (1) develop a machine translation system for the Indonesian-Batak Toba language using an Indonesian-Batak Toba dictionary, (2) perform a translation test on the Indonesian-Batak Toba language, and (3) evaluate the quality of the translations using the BLEU score approach. The Bilingual Evaluation Understudy (BLEU) score assesses the translation outcome [17], [18].

II. Machine Translation

A. Direct Machine Translation

The three fundamental models of machine interpretation are DMT, rule-based machine translation, and data-driven machine translation [4]. The process of directly interpreting sentences using a machine is achieved by systematically handling the conversion of sentence structures from the source language to the target language with the assistance of a bilingual dictionary. During the direct machine translation process, the machine translator does not analyze the sentence structure of the source language. Instead, it performs preprocessing and morphological analysis to turn the sentence into a list of words. The bilingual dictionary will sequentially compare the terms derived from the source language. The list of terms in the source language that correspond to the target language and are found in the bilingual dictionary will be collected once more for reorganization according to the target language's grammar. The ultimate phase involves morphologically creating the translation results to achieve sentence structure that aligns with the target language.

B. Automatic Evaluation Translation

The Bilingual Evaluation Understudy (BLEU) [17], [18] metric can be employed to assess the quality of machine-translated sentences from the Lampung language to Indonesian. BLEU (Bilingual Evaluation Understudy) evaluates machine translations by comparing them to one or more reference translations created by humans using the same source text. The BLEU approach evaluates the alignment between consecutive words in the automatic translation and consecutive phrases in the reference translation, quantifying the number of weighted matches. This matching is performed autonomously. A higher match rate signifies greater similarity to the reference translation and a higher score. Intelligence and language correctness are not seen as factors [17], [18]. The BLEU metric evaluates the appropriateness of a translation by considering the precision of individual words and the fluency by calculating the precision of n-grams for n = 1, 2, 3, 4. BLEU is computed based on three elements: (1) the precision of n-gram matching between the machine translation output and the reference translation, (2) a brevity penalty (BP) to avoid favoring shorter sentences, and (3) clipping to adjust for the frequency of word occurrences.

Precision is determined by the ratio of the number of matching n-grams to the total number of n-grams. The precision, denoted as p_n, is calculated using the following formula:

$$p_n = \frac{\sum_{p \in Hypothesis} ngram \sum_p Count_{clip}(ngram)}{\sum_{p \in Hypothesis} ngram \sum_p Count(ngram)} \quad (1)$$

BLEU calculates the highest occurrence of n-gram matches. In order to prevent the repetition of the same word (n-gram) from being counted numerous times, BLEU restricts the number of matching occurrences of the n-gram to the highest count observed in any individual reference. The formula for a clipped count is mathematically represented as

$$Count_{clip} = \min(ngram\ \ count, \underbrace{max}_{r \in R}(ngram\ count\ in\ r)) \quad (2)$$

The clipped count is the least between the maximum and original counts, always less than or equal to the original count. BLEU replaces the original count with the clipped count while calculating the adjusted precision. The purpose of using BP is to mitigate the impact of sentence length on the BLEU score. This condition is met when the length of the reference sentence |r| is shorter than that of the hypothesis sentence |h|. The equations for BP and BLEU are as follows

$$BP = 1\ \ if\ |h| > |r| \quad (3)$$

$$BP = e^{(1-|r|/|h|)} if\ |h| \leq |r| \quad (4)$$

$$BLEU = BP.\exp(\sum_{n=1}^{N} w_n log p_n) \quad (5)$$

In Formula (5), p_n represents the precision of the *n*-th gram, whereas w_n represents its weight. The BLEU score is computed by adding the weight and precision values and multiplying the result by BP. According to our most accurate understanding, the BLEU Score evaluation is displayed in Table 1 below.

TABLE I. BLEU SCORE

BLEU Score	Interpretation
< 10	Almost useless
10 - 19	Hard to get the gist
20 - 29	The gist is clear, but has significant grammatical errors
30 - 39	Understandable to good translations
40 - 49	High quality translations
50 - 59	Very high quality, adequate, and fluent translations
> 60	Quality often better than human

C. Characteristics of Batak Toba Language

In terms of form, the Batak Toba language sentence structure in writing is as follows: Subject (S) + Predicate (P) + Object (O) + Adverb (K) = SPOK [19] [20]. This section presents various sentences in the Batak Toba language along with their translations in English.

1) Single sentence: for example, in Batak Toba language, "*Ibana Nunga Ro*", it means "*He is here*".
2) Compound sentences: For example, in the Batak Toba language "*Nangpe au marsahit, tongtong do au lao tu sikkola*", it means "*Even though I'm sick, I still go to school*".
3) Imperative sentence: for example, in Batak Toba language, "*Bahen jo tes manis!*", it means "*Make Sweet Tea!*".
4) Interrogative sentence: For example, in the Batak Toba language, "*Nunga mangan ho?*", it means "*Have you eaten?*"
5) News sentence: for example, in the Batak Toba language "*Ulang taon si Dina do sadarion*", it means "*Today is Dina's birthday*".
6) Perfect sentence: for example, in the Batak Toba language "*Nyak ngebattu ulun tuhani di ghani sunday*", it means "*Andi is playing guitar at home*".

III. RESEARCH METHODOLOGY

The research stage encompasses the systematic and deliberate efforts undertaken by researchers to achieve specified objectives in a planned and consistent manner. Figure 1 illustrates the stages of research or research methodology employed in this study.

A. Data Collection

Data collection is crucial in the research process, as obtaining accurate data significantly facilitates the overall procedure. Figure 2 illustrates the Indonesian-Batak Toba dictionary that was used in the study. The primary constituents employed in this study are the Indonesian-Batak Toba lexicon.

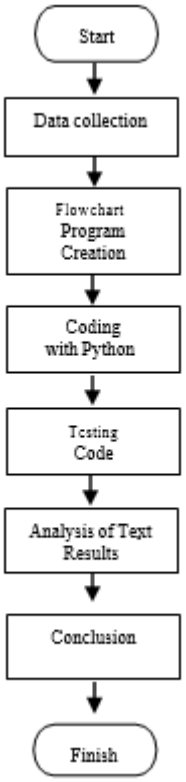

Fig. 1. Stages of Research

Fig. 2. Indonesian-Batak Toba Dictionary

B. Flowchart of The Coding Program

Figure 3 depicts the processes taken in the dictionary-based investigation of Indonesian-Batak Toba language sentence translators. The execution of the application follows a predetermined plan. The user enters an Indonesian sentence as text into the textbox. The system then proceeds to translate the Indonesian message into the Toba Batak language. During the translation process, the system performs text preprocessing, which involves case folding and tokenizing. Case folding refers to standardizing the usage of uppercase letters by converting them to lowercase letters while removing non-alphanumeric characters. Once case folding is completed, the tokenization step is carried out. Tokenization is the process of dividing phrases into individual units known as tokens, typically separated by spaces. Tokenizing yields either a list or a vocabulary list.

The sentences will be converted into a vocabulary list following the text preprocessing phase. Each vocabulary will then be sequentially transferred into the database, with the corresponding Toba Batak language translation stored as the value of the database key. These translations will be stored in a list or temporary list. Suppose the database does not include an analogous vocabulary. In that case, the application will provide a word result based on the first word. Subsequently, the

Indonesian counterpart of the Toba Batak language will be systematically merged, with word boundaries being delineated, resulting in the formation of a sentence in the Toba Batak language.

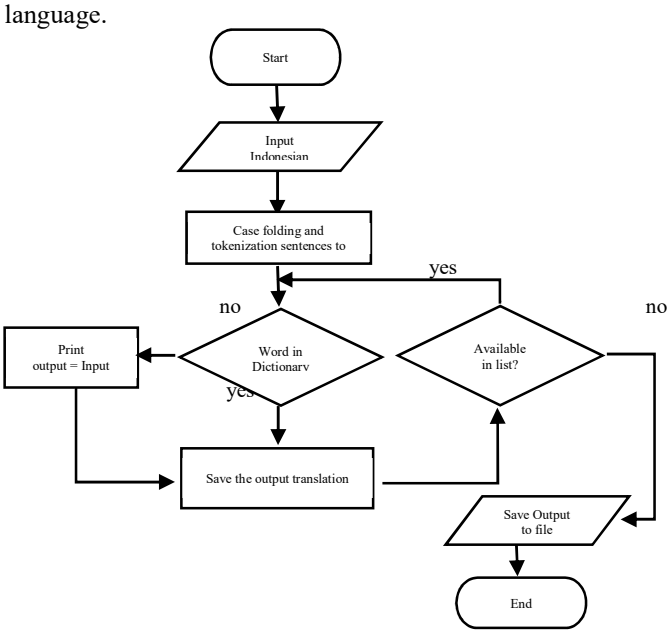

Fig. 3. Flowchart of DMT Indonesian – Batak Toba

C. Coding with Python and Testing Code

This stage involves the application of modeling that has been converted into a user interface using a programming language. The Python programming language is used in conjunction with the Django development framework. The word translation process follows, in which each word stored in the data array is translated into the Batak language based on the data available in the database, and the process of changing or moving the words contained in the database into an array variable that holds the previous word is carried out. Figure 4 shows the Python program code snippet that was created.

```
teks.addEventListener('keyup', e => {
    var xhr = new XMLHttpRequest();
    xhr.onreadystatechange = function(){
        if(this.readyState == 4 && this.status == 200){
            teksoutput.innerHTML = xhr.responseText;
            teks.style.height = "200px";
            let panjang = e.target.scrollHeight;
            teks.style.height = `${panjang}px`;
        }
    };
    xhr.open('GET', '/hai?teks='+teks.value, true);
    xhr.send();
});
```

Fig. 4. Python program code snippet

The process of assembling words into sentences follows, displaying the translation output in the output textfield. Ajax first processes the input before inserting it into the user's textfield. Ajax is required to execute the system asynchronously, where program execution occurs without a queue. Figure 5 shows a piece of the Ajax application code that was created below. In daily interactions, the Indonesian sentence translation system was tested using single, compound, command, question.

```
def hai(request):
    cur = connection.cursor()
    teks = request.GET['teks']

    tekscasefolding = teks.lower()
    tekstokenizing = tekscasefolding.split()

    teksakanditerjemahkan = []

    for kata in tekstokenizing:
        cur.execute("select batak from pythonlearn_bahasa where indonesia = '%s'" % (kata))
        data = cur.fetchall()
        if(data != ()):
            teksakanditerjemahkan.append(data[0][0])
        else:
            teksakanditerjemahkan.append(kata)
    data = " ".join(teksakanditerjemahkan)

    return HttpResponse(data)
```

Fig. 5. A piece of Ajax application code

IV. Result and Discussion

A web Django-based application was developed to help with the process of translating sentences. This application includes automatic machine translation from Indonesian to Batak Toba, as demonstrated in Figure 6 below. This is the page where people can translate Indonesian into Batak. This website has a form for entering Indonesian and displaying the translation results.

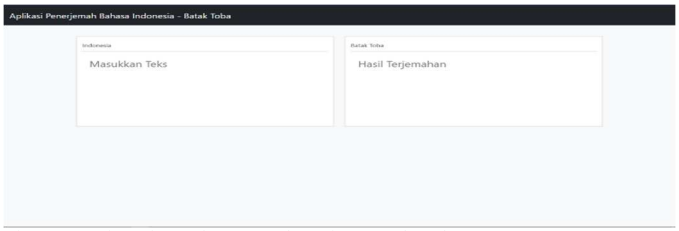

Fig. 6. An interface of DMT Indonesian-Batak Toba

After inputting Indonesian sentences that would be translated into the Toba Batak language, the testing phase of the Indonesian to Batak language translation machine application was carried out. Testing must be performed to look for any remaining faults or vulnerabilities.This translator program has been thoroughly tested. While testing this sentence translator application, the translator application is required to meet the goal of appropriately composing sentences. Some sentences could not be translated correctly and adequately by the translation program during testing.

A. Single sentence and Compound Sentences Testing

During testing, the user enters an example of a single sentence in Indonesian "*dia sudah datang*" into the given textfield. The translation results are shown in figure 7. When this single sentence example was tested, it translated well according to expectations, with the translation result in Batak being "*ibana nunga ro*". In this test, the user enters samples of compound sentences in Indonesian, such as "*aku tetap pergi ke sekolah walau aku tetap sakit*" into the given textfield. Figure 8 shows the outcomes of the translation. The Indonesian translation entered yields " *au tongtong laho tu singkola nangpe au lagi sahit* ". The Batak language sentence structure is still missing in the translation, which should be " *tongtong do au laho tu singkola nangpe au lagi marsahit* " However, if you know the meaning of the Batak language, you can understand this translation.

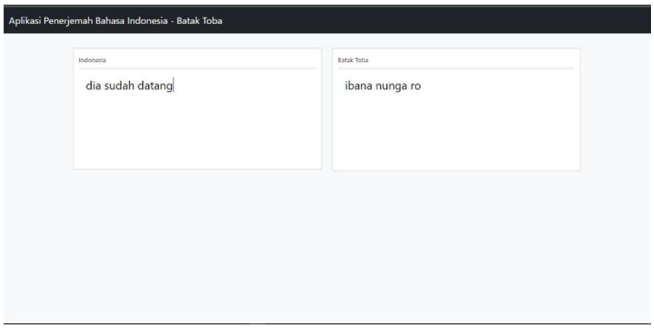

Fig. 7. A single sentece used in application testing

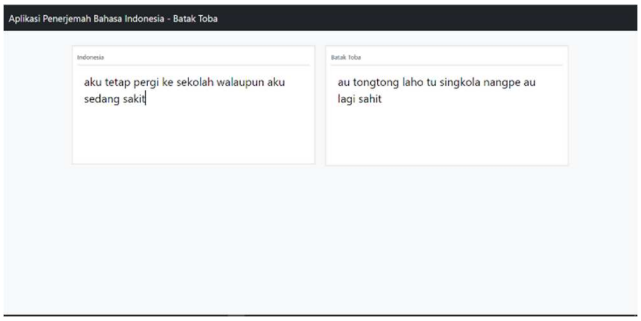

Fig. 8. A compound senteces used in application testing

B. Command sentence and Interrogative Sentence Testing

The user writes an example command sentence in Indonesian, "*jangan makan di kasur !*" into the textfield provided in this test. Figure x depicts the outcomes of the translation. The result of the Batak translation is "*unang mangan di tilam!*" Machine translation is capable of producing good translations that meet expectations.

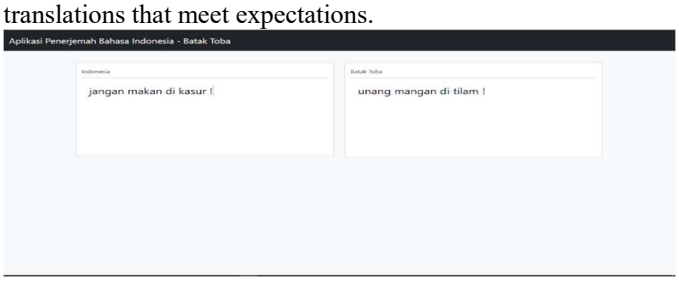

Fig. 9. A command sentece used in application testing

In this test, the user inserts an example of an Indonesian query sentence "kapan kamu pergi ke sekolah?" into the textfield of the given translation program. Figure 10 depicts the translation. The results of the Batak language translation "andigan ho laho tu singkola?".

The next step is to attempt to calculate the BLEU score by comparing machine and human translation. In this study, someone fluent in the Toba Batak language will translate it into human translation. A BLEU score will be determined using the translated sentence as a reference. This test assesses the quality of translation results a translation machine produces from a source language sentence to the target language.

Fig. 10. A interrogative sentece used in application testing

C. Ten Indonesian Sentences Translated to Batak Toba

In this section, we will test ten Indonesian sentences and translate them into Toba Batak using the DMT application. Machine translation was used to translate the Indonesian sentences, then compared to human translations collected from four Batak respondents who speak the Batak language. Tables 2 and 3 below exhibit the BLEU scores from 10 sentences with four reference sentence sources. A value of 32.3% was included in the excellent group, understandable although the sentence structure was jumbled, based on the results of assessing the ten sentences. These ten sentences can illustrate this machine translator, demonstrating that it can translate and be understood by humans despite poor sentence form.

TABLE II. TEN SENTENCES DMT INDONESIA-BATAK TOBA

No	Indonesian Test Sentences	*The DMT application produces translation results.*
1	Aktivitas kamu apa sekarang	Aktipitas ho aha nuaeng
2	Obat segala penyakit adalah hati yang gembira	Ubat sasude sahit ima roha na sonang
3	Kunci sukses adalah konsistensi	Hunsi hasea ima konsistensi
4	Aku suka menulis di kertas	Au lomo manurat di harotas
5	Lemak perut sangat menyebalkan	Tabo-tabo butuha mansai mansogohon
6	Dia adalah cinta pertama ku	Ibana ima holong parjolo hu
7	Kenapa banyak orang bercerai	Boasa godang jolma marsirang
8	Begitu lah akhir hidup orang jahat	Songoni ma ujung ngolu jolma jahat
9	Jangan melakukan hal yang tidak baik	Unang ulaon taringot nan dang denggan
10	Kamu telihat cantic memakai pakaian itu	Ho tarbereng uli mamangke abit i

TABLE III. DMT COMPARISON RESULT WITH FOUR RESPONSES

The DMT application produces translation results.	Native speaker 1	Native speaker 2	Native speaker 3	Native speaker 4	*BLEU Score*
Aktipitas ho aha nuaeng	Aha ni ulaon mi saonari	Marhua ho nuaeng	Aha ulaon mu nuaeng	Marhua ho nuaeng	21%
Ubat sasude sahit ima roha na sonang	Obat ni sude sahit ima hati na sonang	Ubat ni sude sahit roha na sonang	Ubat ni sian sasude sahit ima las ni roha	Ubat ni sude sahit roha na sonang	42.7%
Hunsi hasea ima konsistensi	Kunci ni sukses ima konsisten	Hunsi ni hasea ima arga hata	Hunsi ni hasea ima arga hata	Hunsi ni hasea ima arga hata	27.5%

Au lomo manurat di harotas	lomo rohakku manurat i kertas	Lomo rohakku manurat di harotas	Naung lomo rohakku manurat di harotas	Lomo rohakku manurat di harotas	42.7%
Tabo-tabo butuha mansai mansogohon	Ndang lomo rohakku mamereng tabo-tabo I butuha	Sogo rohakku mangida tabo-tabo ni butuhaon	Sogo do rohakku mamereng tabo-tabo ni dibutuhaon	Tabo-tabo ni butuha mansai songgot roha	21.4%
Ibana ima holong parjolo hu	Ibana ma cinta parjolo ku	Ibana haholongan ni rohakku na parjolo	Ibana haholongan ni rohakku na parjolo	Imana holong ni rohakku na parjolo	14%
Boasa godang jolma marsirang	Boha godang halak na marsirang	Boasa tung godang jolma marsirang	Boasa tung godang jolma marsirang	Boasa godang jolma sirang	84%
Songoni ma ujung ngolu jolma jahat	Songoni ma ujung ni ngolu halak na bolison	Songoni ma ujung ni ngolu ni jolma na jahat	Songoni ma ujung ni ngolu ni jolma na jahat	Songoni ma ujung ni ngolu jolma na jahat	28.4%
Unang ulaon taringot nan dang denggan	Unang ulaon si karejoan na ndang burju	Unang sai ulahon akka naso suman	Unang sai ulahon akka naso suman	Unang bahen naso denggan	24.2%
Ho tarbereng uli mamangke abit i	Songon na bagak ho mamangke baju i	Mansai bagak do ho mamangke abit i	Mansai bagak do ho mamakke i	Bagak ho molo mamangke pakaian i	17.9%
Average Result					32.3%

V. Conclusion and Future Works

This study develops a Digital Machine Translation (DMT) system for translating Indonesian to the Toba Batak language. The DMT system is implemented as a Django web application. The Translation Engine can convert Indonesian to Batak language word by word. However, a limitation arises when the input word has numerous interpretations. After conducting a test on ten sentences, it was found that the result was 32.3%. This percentage indicates that the sentence construction is understandable but of poor quality. The DMT can translate several types of sentences, including single, compound, imperative, and interrogative. After analyzing the test results of 10 sentences, it was found that the BLEU score achieved an accuracy of 32.3%. This indicates that the translations can be understandable to good. Developers are advised to incorporate stemming or word separation techniques to improve the attached terms' translation.

Acknowledgment

Thank you to the Universitas Teknokrat Indonesia for funding a Hiputs research project under contract number 006/UTI/LPPMI/E.1.1/VIII/2023.

References

[1] A. Fikri Aji *et al.*, "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia," Long Papers.

[2] N. H. Hrp, M. Fikry, and Y. Yusra, "Algoritma Stemming Teks Bahasa Batak Angkola Berbasis Aturan Tata Bahasa," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 3, pp. 642–648, May 2023, doi: 10.47065/josyc.v4i3.3458.

[3] M. A. Muchtar *et al.*, "Separation of Basic Words in Angkola Batak Text Documents using Enhanced Confix Stripping Stemmer Case: Mandailing Ethnic," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Oct. 2019. doi: 10.1088/1757-899X/648/1/012024.

[4] P. Bhattacharyya, *Machine Translation*, Boca Raton: Taylor & Francis, Group. 2015

[5] S. Ningsih, "EKSPERIMEN PENGENALAN UCAPAN AKSARA LAMPUNG DENGAN CMU SPHINX 4," 2018.

[6] D. V Sindhu and B. M. Sagar, "Dictionary Based Machine Translation from Kannada to Telugu," *IOP Conf Ser Mater Sci Eng*, vol. 225, p. 012182, Aug. 2017, doi: 10.1088/1757-899x/225/1/012182.

[7] B. Hettige, S. & Hettige, R. M. M. Shalini, and B. Hettige, "Dictionary Based Machine Translation System for Pali to Sinhala EnSiMaS View project Octopus View project Dictionary Based Machine Translation System for Pali to Sinhala," 2017. [Online]. Available: https://www.researchgate.net/publication/323425834

[8] K. Trya, C. Resmawan, I. Ketut, R. Arhana, I. Made, and G. Sunarya, "PENGEMBANGAN APLIKASI KAMUS DAN PENERJEMAH BAHASA INDONESIA-BAHASA BALI MENGGUNAKAN METODE RULE BASED BERBASIS ANDROID," *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, vol. 4, no. 2, 2015.

[9] N. Afifah, T. B. Santoso, M. Yuliana, L. Sinyal, P. Elektronika, and N. Surabaya, "Pembuatan Kamus Elektronik Kalimat Bahasa Indonesia dan Bahasa Jawa untuk Aplikasi Mobile Menggunakan Interpolation Search," 2010.

[10] A. A. Suryani, D. H. Widyantoro, A. Purwarianti, and Y. Sudaryat, "Experiment on a phrase-based statistical machine translation using PoS Tag information for Sundanese into Indonesian," *2015 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2015 - Proc.*, pp. 1–6, 2016, doi: 10.1109/ICITSI.2015.7437678.

[11] H. Sujaini, "khazanah informatika Jurnal Ilmu Komputer dan Informatika 51 Meningkatkan Peran Model Bahasa dalam Mesin Penerjemah Statistik (Studi Kasus Bahasa Indonesia-Dayak Kanayatn)."

[12] A. B. zaenal Abidin, Adi Sucipto, "Penerjemahan Kalimat Bahasa Lampung-Indonesia Dengan Pendekatan Neural Machine Translation Berbasis Attention," *Journalbalitbangdalampung.Org*, vol. 4, no. 2, pp. 82–86, 2018, [Online].

[13] Z. Abidin, Permata, I. Ahmad, and Rusliyawati, "Effect of mono corpus quantity on statistical machine translation Indonesian-Lampung dialect of nyo," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1742-6596/1751/1/012036.

[14] Z. Abidin and P. Permata, "PENGARUH PENAMBAHAN KORPUS PARALEL PADA MESIN PENERJEMAH STATISTIK BAHASA INDONESIA KE BAHASA LAMPUNG DIALEK NYO," *Jurnal Teknoinfo*, vol. 15, no. 1, p. 13, Jan. 2021, doi: 10.33365/jti.v15i1.889.

[15] Z. Abidin, P. Permata, and F. Ariyani, "Translation of the Lampung Language Text Dialect of Nyo into the Indonesian Language with DMT and SMT Approach," *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 5, no. 1, pp. 58–71, Feb. 2021, doi: 10.29407/intensif.v5i1.14670.

[16] Z. Abidin, Permata, R. I. Borman, U. Ardiyatno, F. Rossi, and Y. Jusman, "Computer-aided Translation Based on Lampung Language as Low Resource Language," in *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering, ICOMITEE 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 7–11. doi: 10.1109/ICOMITEE53461.2021.9650324.

[17] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation."

[18] S. Lee *et al.*, "A Survey on Evaluation Metrics for Machine Translation," *Mathematics*, vol. 11, no. 4, Feb. 2023, doi: 10.3390/math11041006.

[19] V. No and E. Februari, "Bentuk Afiksasi Bahasa Batak Toba Pada Masyarakat," vol. 2, no. 1, pp. 1–4, 2022.

[20] J. Bahasa dan Sastra Indonesia BASASASINDO Hal *et al.*, "BENTUK AFIKSASI BAHASA BATAK TOBA PADA MASYARAKAT KELURAHAN PADANG MASIANG KECAMATAN BARUS."