

Translator Bahasa Indonesia ke Bahasa Daerah (atau sebaliknya): Suatu Tinjauan Literatur

Dian Dwi Ramadhan
Program Studi Informatika – Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia
Yogyakarta, Indonesia
17523228@students.uii.ac.id

Chanifah Indah Ratnasari, S.Kom., M.Kom.
Program Studi Informatika – Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia
Yogyakarta, Indonesia
chanifah.indah@uui.ac.id

Abstract — Bahasa Indonesia digunakan oleh masyarakat untuk berkomunikasi dalam kehidupan. Selain penggunaan bahasa Indonesia sebagai bahasa nasional, masyarakat juga sering kali menggunakan bahasa daerah sebagai alat berinteraksi dalam kehidupan sehari-hari. Dengan adanya perbedaan bahasa antar daerah di Indonesia yang sangat beragam ini, dibutuhkan mesin translator agar masyarakat di Indonesia dapat memahami makna dari bahasa daerah itu sendiri. Peneliti akan membahas tentang penerjemahan bahasa Indonesia ke bahasa daerah atau sebaliknya. Mesin translator merupakan media yang membantu mengalihbahasakan dari satu bahasa ke bahasa yang lainnya secara otomatis. Mesin translator berfungsi sebagai penjematan perbedaan antara suatu bahasa dengan bahasa lainnya. Pada penelitian ini dilakukan tinjauan literatur mengenai translator bahasa Indonesia ke bahasa Daerah atau sebaliknya. Tujuan penelitian yang peneliti lakukan untuk mengetahui metode apa saja yang digunakan pada translator dalam konteks bahasa Indonesia ke bahasa daerah atau sebaliknya, sehingga menjadi dasar penentuan metode pada pengembangan translator ke depannya.

Keywords — Bahasa daerah ke Bahasa Indonesia, Mesin translator, Bahasa Indonesia ke bahasa daerah, Translator

I. PENDAHULUAN

Manusia biasa berinteraksi dan bersosialisasi satu dengan lainnya menggunakan bahasa. Bahasa digunakan untuk menyampaikan suatu ide-ide maupun pemikiran dan keinginan ke orang lain dalam berinteraksi. Penggunaan bahasa dapat mempermudah interaksi antar manusia. Tanpa bahasa akan menyulitkan seseorang untuk menyampaikan apa yang ingin disampaikan. Jadi bahasa merupakan suatu yang sangat penting, sebagai media dalam berkomunikasi.

Bahasa Indonesia merupakan bahasa sehari-hari masyarakat Indonesia untuk berkomunikasi dan bersosialisasi. Jika dilihat penggunaan pada bahasa Indonesia dalam kehidupan sehari-hari ini sangat bervariasi dalam pemakaiannya. Variasi pemakaian bahasa ini dapat disebut dengan ragam bahasa. Ragam bahasa muncul karena masyarakat yang beraneka ragam dan lingkungan yang berbeda-beda dan adanya kebutuhan dalam penggunaan bahasa untuk berkomunikasi, bersosialisasi, dan bekerjasama sesuai dengan kebutuhan sosialnya [1]. Indonesia terdapat beragam suku dan budaya yang pada dasarnya memiliki berbagai macam bahasa yang terdapat dalam satu bangsa yang sama.

Selain pemakaian bahasa Indonesia untuk berbahasa nasional, masyarakat juga memakai bahasa daerah untuk berkomunikasi sehari-hari. Bahasa daerah bagian dari kekayaan bangsa Indonesia secara turun temurun yang harus

dijaga kelestariannya. Bahasa daerah di Indonesia sangatlah beraneka ragam dan biasa digunakan dalam lingkup daerahnya saja. Bahasa daerah di Indonesia selalu terus bertambah di setiap tahunnya. Pada tahun 2018 Indonesia memiliki 668 bahasa daerah dan pada tahun 2019 terdapat lonjakan yang sangat tinggi, yaitu hingga 801 bahasa [12]. Keanekaragaman bahasa ini dapat mengakibatkan ketidakpahaman antara satu orang dengan orang yang lainnya ketika berkomunikasi menggunakan bahasa daerah yang berbeda yang tidak saling dimengerti oleh satu sama lain. Dengan perbedaan bahasa antar daerah di Indonesia yang sangat beragam ini dibutuhkan alat penerjemah atau translator agar masyarakat di Indonesia dapat memahami bahasa daerah itu sendiri. Penelitian ini akan membahas tentang kajian literatur mengenai mesin translator dari bahasa Indonesia ke bahasa daerah atau sebaliknya.

Mesin translator merupakan sebuah alat yang membantu dalam proses pengalihbahasaan satu bahasa ke dalam bahasa yang lainnya secara otomatis. Pada hal ini translator dibutuhkan untuk menjembatani perbedaan antara suatu bahasa dengan bahasa lainnya agar masyarakat dapat saling berinteraksi. Translator selain menggunakan mesin, dapat juga dilakukan oleh manusia. Hasil translator manual dengan manusia dapat berlangsung dengan akurat yaitu menyesuaikan dengan struktur kalimat dan juga pemaknaannya. Akan tetapi sumber daya manusia terbatas, mahal, dan membutuhkan waktu yang lama. Melihat keterbatasan penerjemah manusia mulailah dibuat translator dengan mesin, walaupun hasilnya tidak seakurat penerjemah manusia, tetapi translator dengan menggunakan mesin translator dapat mempercepat hasil terjemahan dan meningkatkan efisiensi dalam menerjemahkan bahasa.

Tujuan studi ini adalah untuk mengetahui metode yang digunakan pada mesin translator sehingga dapat memberikan gambaran, khususnya kepada peneliti lain, dan umumnya kepada para pembaca, sehingga dapat memiliki wawasan mengenai hal tersebut dan memudahkan dalam menentukan metode yang tepat dalam pengembangan mesin translasi ke depannya.

II. METODE

Penelitian ini dilaksanakan oleh peneliti untuk menjadikan suatu tinjauan pengembangan mesin translator bahasa Indonesia ke bahasa daerah dengan menggunakan mesin translasi atau translator. *Keywords* yang digunakan untuk melaksanakan penelitian ini, yaitu: (1) "Mesin translator", (2) "Translator", (3) "Bahasa Indonesia ke bahasa daerah", (4) "Bahasa daerah ke Bahasa Indonesia". Literatur yang dipilih berdasarkan pencarian menggunakan **Google Scholar** dan **Mendeley**. Pemilihan literatur dalam

melaksanakan penelitian ini wajib memenuhi salah satu yang berasal kriteria yang di berikan, seperti:

- Pembahasan literatur terkait dengan bahasa Indonesia dan bahasa daerah atau sebaliknya.
- Pembahasan literatur mencakup mesin translator.
- Pembahasan literatur mencakup metode penerjemahan bahasa.

Diperoleh sebanyak 8 literatur terpilih yang memenuhi kriteria di atas. Pemilihan literatur berdasarkan abstrak, hasil, diskusi, dan kesimpulan dari *literature review* yang berkaitan pada kriteria yang telah diberikan. Literatur akan diurutkan dalam penggunaan metode mesin translator yang akan diberikan pada Tabel 1.

Tabel 1 Literature Review

No	Penulis	Tahun	Topik	SMT	NMT
1	Hadi, Ibnu.	2014	Uji Akurasi Mesin translator Statistik (MPS) Bahasa Indonesia Ke Bahasa Melayu Sambas dan Mesin translator Statistik (MPS) Bahasa Melayu Sambas Ke Bahasa Indonesia.	✓	
2	R. Nugroho Aditya, T. Adji Bharata, and B. Hantono S.	2015	Penerjemahan Bahasa Indonesia dan Bahasa Jawa Menggunakan Metode Statistik Berbasis Frasa.	✓	
3	M. Hasbiansyah, H. Sujaini, and N. Safriadi.	2016	Tuning for Quality Untuk Uji Akurasi Mesin translator Statistik (Mps) Bahasa Indonesia - Bahasa Dayak Kanayatn.	✓	
4	S. Mandira, H. Sujaini, and A. B. Putra.	2016	Perbaikan Probabilitas Lexical Model Untuk Meningkatkan Akurasi Mesin translator Statistik.	✓	
5	B. W. Yohanes, T. Robert, and S. Nugroho.	2017	Sistem Penerjemah Bahasa Jawa-Aksara Jawa Berbasis Finite State Automata.	✓	
6	M. G. Asparilla, H. Sujaini, and R. D. Nyoto.	2018	Perbaikan Kualitas Pasangan kalimat untuk Meningkatkan Kualitas Mesin translator Statistik (Studi Kasus : Bahasa Indonesia – Jawa Krama).	✓	
7	H. Sujaini.	2018	Peningkatan Akurasi Penerjemah Bahasa Daerah dengan Optimasi Pasangan kalimat Paralel.	✓	
8	M. Wahyuni, H. Sujaini, and H. Muhandi.	2019	Pengaruh Kuantitas Pasangan kalimat Monolingual Terhadap Akurasi Mesin translator Statistik.	✓	
9	R. Sennrich, B. Haddow, and A. Birch	2016	Improving neural machine translation models with monolingual data.		✓
10	W. He, Z. He, H. Wu, and H. Wang	2016	Improved neural machine translation with SMT features.		✓
11	F. Meng, Z. Lu, H. Li, and Q. Liu	2016	Interactive attention for neural machine translation.		✓
12	X. Wang, Z. Tu, D. Xiong, and M. Zhang	2017	Translating phrases in neural machine translation.		✓
13	Z. Abidin	2018	Translation of Sentence Lampung-Indonesian Languages with Neural Machine Translation Attention Based Approach.		✓

III. HASIL

Pada Tabel 1, literatur yang diperoleh dikelompokkan berdasar metode yang digunakan, yaitu *Statistical Machine Translation* (SMT) dan *Neural Machine Translation* (NMT). Terdapat delapan literatur yang menggunakan SMT dan lima menggunakan NMT. Pada literatur 1, penerapan mesin translator untuk memperoleh hasil tingkat akurasi dan kualitas terjemahannya dengan menggunakan mesin translator statistik. Cara mengembangkan dan mengimplementasikan penerjemahan bahasa Indonesia ke dalam bahasa Melayu atau sebaliknya dengan metode *Statistical Machine Translation* dengan melakukan pengambilan parameter dari hasil analisis paralel pasangan kalimat. Pelaksanaan pengujian dilakukan dua kali, pengujian pertama dilakukan terhadap bahasa Indonesia ke bahasa Melayu Sambas dengan mendapatkan

skor BLEU sebesar 0,5555 atau setara dengan 55% sedangkan dalam pengujian kedua yaitu dari bahasa Melayu Sambas ke bahasa Indonesia menghasilkan skor sejumlah 0,4950 atau setara dengan 49% [2]. Literatur 2, mengenai Penerjemahan bahasa Indonesia yang disebutkan dengan menggunakan metode statistik berbasis frasa untuk mengevaluasi hasil *translate* dari bahasa Indonesia ke bahasa Jawa. Data yang digunakan bersumber dari Alkitab dengan banyak data 5.080 kalimat bahasa Indonesia dan bahasa Jawa, 104.568 kata bahasa Jawa, dan 93.369 kata bahasa Indonesia. Hasil penelitian yang didapatkan terdapat dua hasil, yaitu hasil pertama 44.02% untuk Indonesia-Jawa dan hasil kedua 48.77% untuk Jawa-Indonesia. Hasil evaluasi dievaluasi menggunakan *Bilingual Evaluation Understudy* (BLEU). Perbedaan hasil tersebut karena corpus parallel mempunyai peran penting dalam menjaga kualitas mesin translator. Penataan

corpus parallel diperlukan untuk memperoleh hasil terjemahan yang baik [3]. Literatur 3, penggunaan mesin translator statistik sebagai uji ketepatan mesin translator statistik dalam bahasa Indonesia ke bahasa Dayak Kanayatn atau disebut *tuning for quality*. Proses ini adalah salah satu cara untuk memperbaiki skor hasil uji akurasi mesin translator statistik. Data yang diambil dari buku biografi berjudul Chairul Tanjung Si Anak Singkong dengan total 3667 pasangan pasangan kalimat paralel. BLEU adalah sebagai parameter untuk pembandingan dari hasil proses *tuning* nantinya. Pengujian dilakukan sebelum dan setelah proses *tuning*. Sebelum dilaksanakan proses *tuning* mendapatkan BLEU dengan skor sebesar 89.47 dan setelah dilaksanakan proses *tuning* mendapatkan skor BLEU sebesar 92.19. Dinyatakan bahwa proses *tuning* menyebabkan peningkatan nilai tingkat akurasi sebesar 3.04% [4]. Literatur 4, mengenai topik yang disampaikan dengan menggunakan mesin translator statistik untuk memperbaiki peluang dalam model *lexical* sehingga dapat meningkatkan hasil akurasi mesin translatoran. Data yang digunakan sebanyak 5.108 kalimat. Pengujian dilakukan dengan dua metode, yaitu metode BLEU dan pengujian oleh linguist Jawa. Hasil yang didapatkan eskalasi hasil dari BLEU sebesar 0.30% (95.32% menjadi 95.61%) pada pemeriksaan otomatis dan 10.69% (88.03% menjadi 97.44%) pada pengujian ahli bahasa Jawa [5]. Literatur 5, penerjemahan bahasa Jawa – aksara Jawa dalam studi ini menggunakan metode *Finite State Automata* dengan Graphite *Javanese Scripts* sebagai penyokong penyusunan huruf Jawa. Sistem penerjemah ini mendapatkan hasil pengujian yang maksimal dari bahasa Jawa latin ke aksara Jawa dan sebaliknya dengan memberikan solusi masukan kata yang mengikuti pelafalannya dalam menggunakan bahasa Jawa. Tingkat keberhasilan untuk penerjemahan dari bahasa Jawa ke aksara Jawa mendapatkan skor sebesar 92% dan penerjemahan aksara Jawa ke bahasa Jawa latin mendapatkan tingkat keberhasilan dengan skor 93.8% [6]. Literatur 6, menggunakan perbaikan dalam bobot pasangan kalimat untuk meningkatkan studinya. Tujuan dalam penelitian tersebut untuk melihat seberapa berpengaruh dalam memperbaiki bobot pasangan kalimat dengan menyaring pasangan kalimat yang berkualitas. Data yang diperoleh dari studi ini berasal dari novel kesusastran yang berasal dari daerah Jawa dengan jumlah 5000 pasangan kalimat bahasa Indonesia dan bahasa Jawa krama. Pengujian dilaksanakan dengan *Automatic Evaluation* dan manual. Hasil pengujian dari peningkatan optimasi pasangan kalimat dalam penerjemahan bahasa Indonesia ke bahasa Jawa krama pada presentasi mesin translator statistik dengan menguji dari pasangan kalimat sebanyak 15 kalimat sebesar 10.53% dan 100 kalimat uji dari pasangan kalimat dengan hasil optimasi sebesar 11.63%. Pengujian manual menggunakan seorang ahli bahasa dengan menghasilkan tingkat akurasi optimasi pasangan kalimat sebesar 0.03% [7]. Literatur 7, mengenai topik yang dibahas untuk memperbaiki bobot dalam kalimat yang berpasangan bahasa Indonesia ke bahasa Melayu dan bahasa Indonesia ke bahasa Jawa tanpa mengecek kalimat yang ada pada pasangan kalimat. Digunakan data pasangan kalimat bahasa Indonesia ke bahasa Melayu sebanyak 12.000 kalimat dan pasangan kalimat monolingual Melayu dan Jawa sendiri-sendiri sebanyak 50.000 kalimat. Data lainnya adalah 5.100 pasangan kalimat untuk bahasa Indonesia ke bahasa Jawa

dan 10.000 pasangan kalimat monolingual. Hasil yang didapatkan menggunakan pengujian BLEU terdapat peningkatan tingkat akurasi sebesar 6.97% untuk Indonesia-Melayu dan 5.55% untuk Indonesia-Jawa [8]. Literatur 8, penerapan mesin translator statistik sebagai menentukan tingkat akurasi dari pengaruh kuantitas pasangan kalimat monolingual dari bahasa Inggris ke dalam bahasa Indonesia umumnya membutuhkan data pasangan kalimat yang sangat banyak, tetapi pada literatur tersebut dikatakan bahwa akurasi pada mesin translator statistik menggunakan pasangan kalimat monolingual yang hanya berisi teks dalam satu bahasa agar mempermudah dalam pembuatan pasangan kalimat. Penggunaan data dalam penelitian tersebut berasal dari alamat web berita dua bahasa (<http://berita2bahasa.com/>) dengan mengumpulkan teks-teks yang kemudian diolah menjadi teks pasangan kalimat paralel dan pasangan kalimat monolingual yang terdiri dari 6000 pasang kumpulan pasangan kalimat bahasa Inggris ke bahasa Indonesia dan 54000 kumpulan kalimat monolingual bahasa Indonesia. Pengujian dilakukan dengan *Automatic Evaluation* dan pengujian manual. Hasil pengujian *Automatic Evaluation* dengan menggunakan *Bilingual Evaluation Understudy* (BLEU) dengan skor tingkat kecermatan sebesar 10.13% dilakukan dengan berangsur-angsur terhadap 2000 kalimat uji. Pengujian kedua dilakukan oleh seorang linguist Inggris dengan 100 kumpulan kalimat uji mendapatkan skor sebesar 10.07% [9]. Literatur 9, dalam studi ini penggunaan *Neural Machine Translation* (NMT) telah memperoleh kinerja mutakhir untuk beberapa pasangan bahasa, sementara hanya menggunakan data paralel untuk pelatihannya. Data monolingual sisi target memainkan peran penting dalam meningkatkan kelancaran untuk terjemahan mesin statistik berbasis frase dan menyelidiki penggunaan data monolingual untuk NMT. Studi ini mengeksplorasi strategi untuk melatih dengan data monolingual tanpa mengubah arsitektur jaringan saraf. Dengan memasang data pelatihan monolingual dengan terjemahan balik otomatis studi ini dapat memperlakukannya sebagai data pelatihan paralel tambahan, dan memperoleh peningkatan substansial pada tugas WMT 15 Bahasa Inggris-Jerman (+2,8–3,7 BLEU), dan untuk IWSLT Turki-Inggris (+2.1–3.4 BLEU), mendapatkan hasil baru yang canggih dan menunjukkan bahwa penyetulan halus pada data monolingual dan paralel dalam domain yang memberikan peningkatan substansial untuk Bahasa Inggris-Jerman [13]. Literatur 10, studi ini, peningkatan NMT dengan mengintegrasikan komponen SMT tambahan (misalnya tabel terjemahan, model bahasa) di bawah kerangka kerja log-linear, yang membuat pendekatan NMT mudah diperluas. Bagian ini secara singkat mengulas *encoder-decoder* RNN yang menggunakan pendekatan NMT yang baru-baru ini diusulkan berdasarkan jaringan saraf berulang dan model log-linear dengan kerangka dominan untuk SMT dalam dekade terakhir. Terjemahan mesin saraf (NMT) melakukan terjemahan ujung ke ujung dengan encoder bahasa sumber dan dekoder bahasa target yang membuat kinerja terjemahan yang menjanjikan. Namun, sebagai pendekatan yang baru muncul, metode ini memiliki beberapa keterbatasan. Sistem NMT biasanya harus menerapkan kosa kata dengan ukuran tertentu untuk menghindari pelatihan dan *decoding* yang memakan waktu, sehingga menyebabkan masalah *out-of-vocabulary* yang serius. Selain itu, dekoder tidak memiliki mekanisme

untuk menjamin semua kata sumber diterjemahkan dan biasanya lebih menyukai terjemahan pendek, menghasilkan terjemahan yang lancar tetapi tidak memadai. Untuk mengatasi masalah di atas, dengan menggabungkan fitur terjemahan mesin statistik (SMT), seperti model terjemahan dan model bahasa n-gram, dengan model NMT di bawah kerangka log-linear. Eksperimen yang dilakukan ini menunjukkan bahwa metode yang diusulkan secara signifikan meningkatkan kualitas terjemahan sistem NMT mutakhir pada tugas terjemahan bahasa Mandarin ke Inggris. Metode yang dilakukan ini menghasilkan perolehan hingga 2,33 skor BLEU pada set tes terbuka NIST. Eksperimen pada tugas terjemahan Bahasa Mandarin-ke-Inggris menunjukkan bahwa sistem ini dapat mencapai peningkatan yang signifikan di atas baseline pada sejumlah besar korpus pelatihan yang dirayapi dari web. Sebagai pendekatan baru, NMT masih memiliki lebih banyak ruang untuk perbaikan. *Encoder-decoder* RNN saat ini sebenarnya adalah sistem terjemahan berbasis kata [14]. Literatur 11, penggunaan *Neural Machine Translation* (NMT) berbasis perhatian konvensional melakukan penyesuaian dinamis dalam menghasilkan kalimat pada target. Mekanisme perhatian dapat meningkatkan NMT yang canggih. Dalam studi ini peneliti menggunakan mekanisme perhatian baru yang disebut *INTERACTIVE ATTENTION* yang memodelkan interaksi antara dekoder dan representasi kalimat sumber selama penerjemahan dengan operasi membaca dan menulis. *INTERACTIVE ATTENTION* dapat melacak riwayat interaksi dan karenanya meningkatkan kinerja terjemahan. Eksperimen pada tugas terjemahan Bahasa Mandarin-Inggris NIST menunjukkan bahwa *INTERACTIVE ATTENTION* dapat mencapai peningkatan yang signifikan baik dari baseline NMT berbasis perhatian sebelumnya dan beberapa varian canggih NMT berbasis perhatian. Pendekatan *INTERACTIVE ATTENTION* yang sederhana namun efektif yang memodelkan interaksi antara decoder dan representasi kalimat sumber selama penerjemahan dengan menggunakan operasi membaca dan menulis. Studi empiris pada terjemahan Cina-Inggris menunjukkan bahwa *INTERACTIVE ATTENTION* dapat secara signifikan meningkatkan kinerja NMT. Penerjemah mesin saraf dengan *INTERACTIVE ATTENTION* ini dapat mengungguli sistem NMT berbasis perhatian *open-source* Groundhog sebesar 4,22 poin BLEU dan sistem berbasis frase sumber terbuka Moses dengan rata-rata 3,94 poin BLEU pada beberapa set pengujian [15]. Literatur 12, Dalam literatur ini akan mempresentasikan model baru untuk menerjemahkan frasa sumber dan menghasilkan terjemahan frasa target di NMT dengan mengintegrasikan memori frasa ke dalam arsitektur *encoder-decoder*. Frase memainkan peran penting dalam pemahaman bahasa alami dan terjemahan mesin. Namun, sulit untuk mengintegrasikannya ke dalam terjemahan mesin saraf (NMT) saat ini yang membaca dan menghasilkan kalimat kata demi kata. Pada *decoding*, model SMT secara dinamis menghasilkan frase target yang relevan dengan informasi kontekstual yang diberikan oleh model NMT dan menulisnya ke memori frase. Kemudian model yang diusulkan membaca memori frase dan menggunakan penyeimbang untuk membuat perkiraan probabilitas untuk frase dalam memori frase. Akhirnya, dekoder NMT memilih sebuah frase dari memori frase atau sebuah kata dari kosa kata dengan kemungkinan tertinggi untuk dihasilkan. Hasil percobaan pada terjemahan bahasa

Mandarin-Inggris telah menunjukkan bahwa model yang diusulkan dapat secara signifikan meningkatkan kinerja terjemahan. Hasil eksperimen pada terjemahan bahasa Mandarin-Inggris menunjukkan bahwa model yang diusulkan mencapai peningkatan yang signifikan di atas baseline pada berbagai perangkat pengujian [16]. Literatur 13, studi yang dilakukan ini merupakan translator bahasa Lampung ke bahasa Indonesia menggunakan pendekatan *neural machine translation* (NMT) berbasis *attention*. NMT, metode pendekatan baru dalam teknologi terjemahan mesin telah bekerja dengan menggabungkan *encoder* dan *decoder*. *Encoder* di NMT adalah komponen jaringan saraf berulang yang mengenkripsi bahasa sumber ke beberapa vektor stabil panjang dan *decoder* adalah komponen jaringan saraf berulang yang menghasilkan hasil terjemahan secara komprehensif. Penelitian NMT telah dimulai dengan membuat 3000 kalimat paralel bahasa Lampung (dialek API) dan bahasa Indonesia. Kemudian dilanjutkan dengan menentukan model parameter NMT untuk proses pelatihan data. Langkah selanjutnya adalah membangun model NMT dan mengevaluasinya. Pengujian pendekatan ini menggunakan 25 monolingual korpus tanpa *out-of-vocabulary* (OOV), 25 kalimat tunggal dengan OOV, 25 paralel korpus tanpa OOV, dan 25 kalimat jamak dengan OOV. Hasil pengujian terjemahan menggunakan perhatian NMT menunjukkan nilai bilingual evaluation understudy (BLEU) adalah 51,96%.

IV. DISKUSI

Setelah mengkaji beberapa *literature review* dengan metode *Statistical Machine Translation* (SMT) dan *Neural Machine Translation* (NMT) terdapat 10 hasil yang diperoleh akan ditampilkan seperti Tabel 2 di bawah.

Tabel 2 Metode dan Tingkat Akurasi

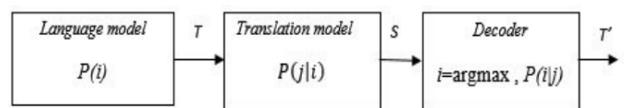
No	Literatur	SMT	NMT	Hasil
1	1	✓		55% untuk Bahasa Indonesia ke bahasa Melayu Sambas dan 49% untuk sebaliknya.
2	2	✓		Bahasa Indonesia ke bahasa Jawa 44.02%. Bahasa Jawa ke bahasa Indonesia 48.77%.
3	4	✓		Peningkatan akurasi 0.30% (95.32% menjadi 95.61%) pada pengujian otomatis dan 10.69% (88.03% menjadi 97.44%)
4	6	✓		Kenaikan 10.53% dengan 15 di luar pengujian kalimat Kenaikan 11.63% dengan 100 kalimat uji dari pasangan kalimat dengan hasil optimasi sebesar.
5	7	✓		Peningkatan akurasi bahasa Indonesia ke

				bahasa Melayu 6.97%. Peningkatan akurasi bahasa Indonesia ke bahasa Jawa 5.55%.
6	8	✓		Akurasi peningkatan kuantitas pasangan kalimat monolingual dari bahasa Inggris ke bahasa Indonesia sebesar 10.13%.
7	9		✓	Peningkatan substansial pada tugas WMT 15 Bahasa Inggris-Jerman (+2,8–3,7 BLEU), dan untuk IWSLT Turki-Inggris (+2.1–3.4 BLEU).
8	10		✓	menghasilkan perolehan hingga 2,33 skor BLEU pada set tes terbuka NIST
9	11		✓	<i>INTERACTIVE ATTENTION</i> dapat mengungguli sistem NMT berbasis <i>attention open-source</i> Groundhog sebesar 4,22 poin BLEU dan sistem berbasis frase sumber terbuka Moses dengan rata-rata 3,94 poin BLEU.
10	12		✓	Hasil pengujian terjemahan menggunakan perhatian NMT dengan poin BLEU sebesar 51,96%.

Dari Tabel 2 di atas didapatkan metode yang tepat sebagai alat penerjemah dengan mendapatkan hasil yang diinginkan sebagai mesin translator. Mesin yang dipakai untuk menerjemahkan bahasa Indonesia ke bahasa daerah ini menggunakan metode mesin translator statistik dan mesin saraf terjemah. Terjemahan mesin saraf (NMT) adalah pendekatan terjemahan mesin yang menggunakan jaringan saraf tiruan untuk memprediksi kemungkinan urutan kata, biasanya memodelkan seluruh kalimat dalam satu model terintegrasi. Mesin translator ini hanya memerlukan sebagian kecil dari memori yang dibutuhkan oleh model terjemahan mesin statistik (SMT) tradisional. Selanjutnya, tidak seperti sistem terjemahan konvensional, semua bagian dari model terjemahan saraf dilatih bersama (*end-to-end*) untuk memaksimalkan kinerja terjemahan. Mesin saraf terjemah menggunakan jaringan saraf untuk menerjemahkan teks sumber ke teks target, dan jaringan saraf dapat bekerja dengan set data yang sangat besar dan memerlukan sedikit pengawasan. Sistem terjemahan mesin neural memiliki dua bagian utama yang merupakan jaringan saraf yaitu: jaringan encoder dan jaringan decoder. Mesin translator statistik tersebut merupakan jenis mesin dengan pendekatan statistik yang kegunaannya sebagai penerjemah dari bahasa sumber ke bahasa target

[7]. Komponen yang akan dilibatkan dalam melakukan sebuah penerjemahan, yaitu *Language Model*, *Translation Model*, dan *Decoder* adalah suatu pengetahuan tentang mesin translator statistik berfungsi sebagai pemasangan teks masukan dalam bentuk bahasa sumber dan menghasilkan keluaran berupa bahasa yang telah diterjemahkan ke bahasa daerah sesuai dengan objek yang melakukan percakapan dalam bahasa daerah. Hal ini sangat berpotensi sebagai metode dalam pengimplementasian dalam menerjemahkan bahasa Indonesia ke bahasa daerah atau sebaliknya dengan memanfaatkan mesin translator statistik sehingga pengguna dapat memahami ketika sedang berinteraksi.

Dapat dilihat dari pembahasan pada literatur 6 yang memberitahu bahwa mesin translator statistik memiliki tiga elemen yang melibatkan proses *translate* kalimat dari bahasa sumber ke bahasa yang dituju yaitu, *translation model*, *decoder*, dan *language model* seperti terlihat di Gambar 1 [9].



Gambar 1 Elemen mesin translator statistik [9]

Diperlukan 3 komponen agar mesin translator statistik dapat bekerja, sebagaimana yang telah di ditampilkan pada Gambar.1

- *Language model*
Model bahasa merupakan komponen yang dilaksanakan pada sistem pengoperasian *Natural Language Processing* (NLP) seperti *syntactic parsing* dan *speech recognition*, *part-of-speech tagging*. Model bahasa statistik mematokkan peluang $P(W_{1,n})$ ke serangkaian n kata dengan rata-rata sebuah diseminasi peluang.
- *Model Translator*
Pada Gambar.1 *Translation model* adalah komponen yang sangat penting pada mesin translator statistik yang dapat memilah kalimat bahasa sumber menjadi barisan beberapa kata, mengalihbahasakan setiap frasa menuju tujuan, dan penyusunan ulang. Dalam mesin translator statistik ada dua model translator, yaitu model translasi berbasis kata atau biasa disebut dengan *word-based translation model* dan model translasi dengan berbasis frase juga biasa disebut *phrase-based translation model*[11].
- *Decoder*
Decoder merupakan komponen terakhir yang berperan dalam pencarian naskah bahasa yang dihasilkan kemudian memiliki peluang paling besar dalam memperhitungkan hasil model terjemahan dan model bahasa.

Dalam mesin translator statistik data yang diambil berupa data kalimat. Data kalimat tersebut akan dikumpulkan sebagai objek evaluasi pengujian tingkat akurasi dan hasil dari terjemahan satu bahasa ke bahasa lain. Pengujian tersebut menggunakan *Automatic Evaluation* dengan menggunakan *Bilingual Evaluation Understudy* (BLEU). *Automatic Evaluation* yang digunakan dalam studi ini merupakan suatu algoritma yang bekerja sebagai alat evaluasi untuk bobot dari sebuah hasil bahasa yang telah selesai diterjemahkan dari bahasa sumber ke bahasa daerah [7]. Dalam BLEU terdapat nilai

dari hasil perkalian antara *brevity penalty* dengan *mean geometri* dari *modified precision score*. Nilai dari hasil evaluasi terdapat pada antara 0 hingga 1. Semakin banyak kalimat yang menjadi acuan per kalimatnya, nilai BLEU yang akan didapat semakin akurat dan semakin tinggi nilainya. Untuk mendapatkan hal tersebut adalah semakin banyak dari suatu kalimat hasil, maka terjemahannya juga harus hampir sama dengan kalimat acuannya [9]. Rumus BLEU seperti Gambar 2 berikut:

$$BP_{BLEU} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

$$P_n = \frac{\sum_{C \in \text{corpus } n\text{-gram}} C \sum \text{count}_{clip(n\text{-gram})}}{\sum_{C \in \text{corpus } n\text{-gram}} C \sum \text{count}(n\text{-gram})}$$

$$BLEU = BP_{BLEU} \cdot e^{\sum_{n=1}^N w_n \log p_n}$$

Gambar 2 Rumus *Bilingual Evaluation Understudy* (BLEU) [5]

Keterangan:

BF	=	<i>brevity penalty</i>
c	=	total kata hasil <i>translate</i> otomatis
r	=	total kata rujukan
P_n	=	<i>modified precision score</i>
W_n	=	$1/N$ (standar nilai N untuk BLEU adalah 4)
p_n	=	total n-gram hasil terjemahan yang sesuai rujukan dibagi total n-gram total terjemahan

V. KESIMPULAN

Berdasarkan hasil dari pengamatan tinjauan literatur di atas, didapat berbagai sebagai berikut:

- *Statistical Machine Translator* (SMT) adalah mesin translator dengan melakukan pendekatan statistik guna sebagai translator satu bahasa sumber ke bahasa yang akan dituju. Mesin translator tersebut sangat membantu dalam menerjemahkan bahasa sumber ke bahasa yang ingin diterjemahkan.
- *Neural Machine Translator* (NMT) adalah pendekatan terjemahan mesin yang menggunakan jaringan saraf tiruan untuk memprediksi kemungkinan urutan kata dalam menerjemahkan bahasa sumber ke bahasa terjemahan.
- *Automatic Evaluation* dengan menggunakan *Bilingual Evaluation Understudy* (BLEU) berfungsi sebagai alat evaluasi dalam pencarian bobot dari sebuah *output* translator yang telah selesai diterjemahkan dari bahasa sumber ke bahasa lain. Semakin banyak kumpulan pasangan kalimat, maka tingkat akurasi yang didapat akan semakin tinggi.
- Didapatkan metode yang tepat sebagai alat penerjemah dengan mendapatkan nilai akurasi yang diinginkan untuk menerjemahkan dari Indonesia ke bahasa daerah atau sebaliknya. Mesin translator statistik atau *Statistical Machine Translation* (SMT) dan *Neural Machine Translator* (NMT) adalah jenis mesin

translator yang berguna sebagai penerjemah dari satu bahasa yang ingin diterjemahkan dan menghasilkan bahasa yang akan dituju.

Dengan adanya *paper* ini dapat memberikan gambaran mengenai metode pada mesin translasi, terutama pada bahasa Indonesia ke bahasa daerah, atau sebaliknya. Saran untuk keberlangsungan dalam pengaktualisasian *literatur review* berikutnya agar mengali referensi yang lebih terperinci mengenai mesin translator untuk penggunaan translator dari suatu bahasa yang ingin di terjemahkan ke bahasa tujuan.

REFERENSI

- [1] Rina Devianty, "Bahasa Sebagai Cermin Kebudayaan," *J. Tarb.*, vol. 24, no. 2, pp. 226–245, 2017.
- [2] I. Hadi, "Uji Akurasi Mesin translator Statistik (MPS) Bahasa Indonesia Ke Bahasa Melayu Sambas Dan Mesin translator Statistik (MPS) Bahasa Melayu Sambas Ke Bahasa Indonesia," *J. Sist. dan Teknol. Inf.*, vol. 2, pp. 1–6, 2014.
- [3] R. Nugroho Aditya, T. Adji Bharata, and B. Hantono S, "Penerjemahan Bahasa Indonesia dan Bahasa Jawa Menggunakan Metode Statistik Berbasis Frasa," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2015, no. Sentika, 2015.
- [4] M. Hasbiansyah, H. Sujaini, and N. Safriadi, "Tuning for Quality Untuk Uji Akurasi Mesin translator Statistik (Mps) Bahasa Indonesia - Bahasa Dayak Kanayaln," *J. Sist. dan Teknol. Inf.*, vol. 1, no. 1, pp. 3–7, 2016.
- [5] S. Mandira, H. Sujaini, and A. B. Putra, "Perbaikan Probabilitas Lexical Model Untuk Meningkatkan Akurasi Mesin translator Statistik," *J. Edukasi dan Penelit. Inform.*, vol. 2, no. 1, pp. 3–7, 2016, doi: 10.26418/jp.v2i1.13393.
- [6] B. W. Yohanes, T. Robert, and S. Nugroho, "Sistem Penerjemah Bahasa Jawa-Aksara Jawa Berbasis Finite State Automata," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 2, 2017, doi: 10.22146/jnteti.v6i2.306.
- [7] M. G. Asparilla, H. Sujaini, and R. D. Nyoto, "Perbaikan Kualitas Pasangan kalimat untuk Meningkatkan Kualitas Mesin translator Statistik (Studi Kasus : Bahasa Indonesia – Jawa Krama)," vol. 1, no. 2, pp. 66–74, 2018.
- [8] H. Sujaini, "Peningkatan Akurasi Penerjemah Bahasa Daerah dengan Optimasi Pasangan kalimat Paralel," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 7, no. 1, 2018, doi: 10.22146/jnteti.v7i1.394.
- [9] M. Wahyuni, H. Sujaini, and H. Muhandi, "Pengaruh Kuantitas Pasangan kalimat Monolingual Terhadap Akurasi Mesin translator Statistik," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 1, p. 20, 2019, doi: 10.26418/justin.v7i1.27241.
- [10] H. Maruli Manurung and H. Tanuwijaya, "Penerjemahan Dokumen Inggris-Indonesia Menggunakan Mesin translator Statistik Dengan Word Reordering Dan Phrase Reordering," *J. Ilmu Komput. dan Inf.*, vol. 2, no. 1, pp. 17–24, 2009, doi: 10.21609/jiki.v2i1.122.
- [11] Budiwiyanto, Adi. Kontribusi Kosakata Bahasa Daerah dalam Bahasa Indonesia. <https://badanbahasa.kemdikbud.go.id/lamanbahasa/content/kontribusi-kosakata-bahasa-daerah-dalam-bahasa-indonesia>. (Diakses pada 17 Oktober 2021)
- [12] Lidwina, Andrea. (2021). Papua Punya Bahasa Daerah Terbanyak di Indonesia. <https://databoks.katadata.co.id/datapublish/2021/05/14/papua-punya-bahasa-daerah-terbanyak-di-indonesia>. (Diakses pada 17 Oktober 2021)
- [13] R. Sennrich, B. Haddow, and A. Birch, "Improving neural machine translation models with monolingual data," *54th Annu. Meet. Assoc. Comput. Linguist. ACL 2016 - Long Pap.*, vol. 1, pp. 86–96, 2016, doi: 10.18653/v1/p16-1009.
- [14] W. He, Z. He, H. Wu, and H. Wang, "Improved neural machine translation with SMT features," *30th AAAI Conf. Artif. Intell. AAAI 2016*, no. 10, pp. 151–157, 2016.
- [15] F. Meng, Z. Lu, H. Li, and Q. Liu, "Interactive attention for neural machine translation," *COLING 2016 - 26th Int. Conf. Comput. Linguist. Proc. COLING 2016 Tech. Pap.*, pp. 2174–2185, 2016.

[16] X. Wang, Z. Tu, D. Xiong, and M. Zhang, "Translating phrases in neural machine translation," *EMNLP 2017 - Conf. Empir. Methods*

Nat. Lang. Process. Proc., pp. 1421–1431, 2017, doi: 10.18653/v1/d17-1149