

Efek Peningkatan Jumlah Paralel Korpus Pada Penerjemahan Kalimat Bahasa Indonesia ke Bahasa Lampung Dialek *Api*

¹Permata, ²Zaenal Abidin & ³Farida Ariyani

¹ Program Studi Informatika, ² Program Studi Sistem Informasi, Univeristas Teknokrat Indonesia, Jalan Z.A Pagar Alam No.9-1, Bandarlampung, Indonesia

³ Program Studi Magister Pendidikan Bahasa dan Kebudayaan Lampung, Univeristas Lampung, Jalan Prof. Soemantri Brojonegoro No. 1, Bandarlampung, Indonesia

e-mail : permata@teknokrat.ac.id, zabin@teknokrat.ac.id, farida.ariyani@fkip.unila.ac.id

Abstract — Experimental observations of the effect of the number of parallel corpus on Indonesian translation into the Lampung dialect *api* were carried out using the statistical machine translation (SMT) method. SMT utilizes a parallel Indonesian corpus and its translation in the Lampung dialect *api* as a material for training data. The research strategy was carried out in three ways, namely first strategy with a corpus parallel number of 1000 sentences, the second strategy with a corpus parallel number of 2000 and the third strategy with a corpus parallel number of 3000 sentences. The research starts from the preprocessing phase followed by the training phase, namely the parallel corpus processing phase to obtain a language model and translation model. Then the testing phase, and ends with the evaluation phase. SMT testing uses 25 single sentences without out-of-vocabulary (OOV), 25 single sentences with OOV, 25 compound sentences without OOV and 25 compound sentences with OOV. The test results of translating Indonesian sentences into Lampung dialectic *api* are shown through the accuracy value of Bilingual Evaluation Undestudy (BLEU) obtained in testing 25 single sentences without out-of-vocabulary (OOV) in the first strategy, the second and the third are 21.49%, 59.58% and 73.21%. In testing 25 single sentences with out-of-vocabulary (OOV) obtained in the first strategy, the second and the third are 23.22%, 44.33% and 68.72%. In testing 25 compound sentences without out-of-vocabulary (OOV) obtained in the first strategy, the second and the third are 18.22%, 39.4% and 69.18%. In testing 25 compound sentences with out-of-vocabulary (OOV) obtained in the first strategy, the second and the third are 25.94%, 28.22% and 71.94%.

Keywords: Paralel Corpus; Monocorpus; Statistical Machine Translation; Bilingual Evaluation Undestudy; out-of-vocabulary.

1. PENDAHULUAN

Beberapa upaya eksperimen otomatisasi penelitian penerjemahan bahasa Indonesia ke berbagai bahasa daerah mulai telah dilakukan dalam kurun waktu 10 tahun terakhir ini termasuk bahasa Lampung[1][2]. Secara umum di provinsi Lampung terdapat bahasa Lampung yang memiliki dua dialek utama yaitu dialek *api* dan dialek *nyo*. Pada penelitian ini difokuskan pada upaya eksperimen otomatisasi penerjemahan bahasa Indonesia ke bahasa Lampung dialek *api*, menggunakan *tool* Moses. *Tool* ini dapat diakses di laman (<http://www.statmt.org/moses/>). Penelitian ini bertujuan untuk investigasi awal dalam pengembangan mesin penerjemah pada bahasa Indonesia dan bahasa Lampung dialek *api* berdasarkan pendekatan *Statistical Machine Translation* (SMT). Penelitian penerjemahan bahasa Lampung dialek *api* ke bahasa Indonesia telah dilakukan[1][2][3]. Sedangkan penelitian bahasa Lampung dari aspek riset *speech* pertama kali dilakukan oleh Saniati,dkk[4]. Bahasa Indonesia sebagai bahasa pemersatu bangsa dan bahasa Lampung sebagai sebuah kekayaan budaya bagi masyarakat di provinsi Lampung.

Penelitian mesin penerjemah di Indonesia telah dilakukan oleh para peneliti diantaranya penerjemahan bahasa Jawa dan Bahasa Indonesia dengan SMT berbasis frasa[5], penerjemahan bahasa Sunda ke Bahasa Indonesia dengan SMT berbasis frasa dan memanfaatkan *part of speech* (PoS) *Tag*[6], bahasa Lampung juga memiliki aspek dari sisi morfologisnya[7], penerjemahan dokumen bahasa Inggris ke dokumen bahasa Indonesia dengan SMT dan *word reordering* dan *phrase reordering*[8], investigasi peran model bahasa pada penelitian SMT bahasa Indonesia-Dayak Kanayatn[9], pengaruh kuantitas korpus pada penelitian SMT bahasa Bugis

Wajo ke bahasa Indonesia[10], penelitian penerjemahan kalimat bahasa Lampung dialek *api* telah menjadi objek penelitian penerjemahan secara otomatis dengan menggunakan pendekatan *Neural Machine Translation* (NMT) tanpa *attention*[1] dan NMT dengan *attention*[2], berbagai model penerjemahan bahasa Indonesia ke bahasa Jepang juga telah dilakukan[11], pengukuran hasil penerjemahan yang dihasilkan oleh mesin penerjemah menggunakan perhitungan skor *Bilingual Evaluation Understudy* (BLEU)[12], pengamatan aspek morfologi bahasa Lampung telah dilakukan oleh para peneliti bahasa Lampung[13], sedangkan teori SMT diambil dari buku utama[14][15].

Penerjemahan bahasa Indonesia ke bahasa Lampung dialek *api* dapat dilakukan dengan menggunakan kamus bahasa Lampung yang digunakan di sekolah-sekolah. Penerjemahan, menggunakan kamus sebagai rujukan utama dalam menerjemahkan, bahasa Lampung ke bahasa Indonesia sedang diselidiki di tahun 2020 ini oleh peneliti dalam waktu yang bersamaan. Penelitian dengan metode SMT ini bertujuan untuk melakukan pengamatan terhadap hasil penerjemahan bahasa Indonesia ke bahasa Lampung dialek *api* dengan tiga strategi eksperimen pada bagian korpus paralel yang digunakan. Penelitian ini tidak menggunakan bahasa Lampung dialek *nyo* dan tidak menggunakan naskah bernuasa sastra Lampung dialek *api* atau *nyo*. Penelitian ini tidak menyertakan aspek morfologi pada bahasa yang digunakan sebagai salah satu batasan masalahnya.

2. METODOLOGI PENELITIAN

Eksperimen penerjemahan bahasa Indonesia ke bahasa Lampung dialek *api* dilakukan dengan menggunakan *tool* Moses Decoder. Informasi tentang SMT dengan Moses Decoder dapat diakses melalui tautan (<http://www.statmt.org/moses/>). Saat ini dalam menerjemahkan bahasa Indonesia ke bahasa Lampung dialek *api* dilakukan secara manual yaitu dengan menggunakan kamus bahasa Lampung dengan cara pencarian kata per kata yang berkesesuaian. Hasil wawancara dengan guru bahasa Lampung menunjukkan bahasa Lampung memiliki kemiripan struktur tata bahasa dengan bahasa Indonesia. Sebuah kamus bahasa Lampung – Indonesia yang diterbitkan oleh penerbit gunung pesagi Bandar Lampung berisi 4071 kata dialek *Api*, sedangkan kamus bahasa Indonesia edisi ke-5 yang diterbitkan oleh penerbit gramedia pustaka utama memiliki 118000 kata. Beberapa contoh kalimat dalam bahasa Indonesia dan terjemahannya dalam bahasa Lampung *api* dapat dilihat pada tabel 1 di bawah ini.

Tabel 21. Bahasa Indonesia dan terjemahannya dalam bahasa Lampung

Bahasa Indonesia	Bahasa Lampung
Ani tolong ambilkan air itu	<i>Ani tulung akukko wai sina</i>
Ibuku sedang mencuci baju di sungai	<i>Makku lagi mepoh kawai di wai</i>
Kita pergi ke pasar	<i>Gham lapah mit pasagh</i>
Mari kita sholat di masjid	<i>Lapah gham sholat di masjid</i>
Ketua kelas yang mengatur barisan di kelasnya	<i>Ketua kelas sai ngator barisan di kelasni</i>
Dia tetap di kelas satu karena tidak naik kelas tiga kali	<i>Ya tetap di kelas sai ulah mak cakak kelas telu kali</i>

Pada tabel 1 di atas terdapat hal yang menarik yaitu pada kalimat pertama terdapat air yang dalam bahasa Lampung dialek *api* yaitu *wai* sedangkan pada kalimat kedua kata sungai yang dalam bahasa Lampung dialek *api* yaitu *wai*. Dua kata yaitu air dan sungai memiliki padanan kata yang sama yaitu *wai*. Beberapa kata yang memiliki makna berbeda tergantung konteks kalimat atau posisi pada kalimat, seperti kata *lapah*, *sai*. Pengguna kamus bahasa Indonesia - Lampung harus menerjemahkan satu per satu dan memilih makna kata yang tepat secara manual agar kalimat tersebut dapat dipahami dengan baik. Penerjemahan kalimat bahasa Indonesia ke bahasa Lampung dialek *api* adalah sangat penting dengan memperhatikan konteks kalimat yang digunakan agar dapat menghasilkan terjemahan yang sesuai. Pada bahasa Lampung juga terdapat beberapa kata yang

memiliki satu makna yaitu *lapah, mit, tandak, lijing* yang semuanya bermakna *pergi*. Beberapa partikel seperti *kudo, do, pai, tah* semuanya memiliki satu makna yaitu *lah*, sebuah partikel dalam bahasa Indonesia.

Dalam eksperimen SMT bahasa Indonesia ke bahasa Lampung dialek *api* dibutuhkan korpus paralel antara bahasa Indonesia ke bahasa Lampung dialek *api* dan mono korpus bahasa Lampung dialek *api* sebagai bahan membuat *language model*. Pembuatan korpus paralel yang dilakukan secara manual. Korpus paralel akan dilakukan prapemrosesan kalimat secara manual baik pada bahasa Lampung dialek *api* maupun bahasa Indonesia yaitu dengan cara : (1) kalimat dibuat menggunakan semua huruf kecil, (2) antar kalimat dipisahkan oleh spasi termasuk tanda koma, tanda tanya, tanda seru dan titik.

Kalimat-kalimat, bahasa Indonesia dan terjemahannya dalam bahasa Lampung dialek *api*, yang digunakan dalam penelitian ini dibuat secara mandiri dengan cara mengetikkan ulang karena faktor ketiadaan bahan yang dapat diambil secara daring dari laman. Jumlah korpus paralel yang digunakan adalah 3000 kalimat adalah berasal dari buku sekolah dasar (SD) dan lembar kerja siswa (LKS) tingkat SD yang digunakan di provinsi Lampung, buku percakapan sehari-hari bahasa Lampung serta buku afiksasi verba bahasa Lampung. Untuk keperluan penelitian 3000 kalimat korpus paralel disediakan dalam tautan <https://drive.google.com/drive/folders/1eIQcgZCXSytAoVqSbp0w9JcFgWGBjM?usp=sharing>

2.1. Tahapan Ekperimen Penerjemahan Bahasa Indonesia – Lampung Dialek Api

Pada penelitian ini bahan yang digunakan satu buah korpus paralel yaitu 3000 kalimat bahasa Indonesia, 3000 kalimat bahasa Lampung dialek *api* dan satu buah mono korpus 4000 kalimat bahasa Lampung dialek *api*. Korpus paralel bahasa Indonesia – bahasa Lampung dialek *api* digunakan untuk membuat atau membangun *language model* sedangkan mono korpus bahasa Lampung dialek *api* digunakan untuk membuat atau membangun model translasi.

Tahapan eksperimen dilakukan dengan 3 strategi, yaitu :

1. strategi pertama, eksperimen menggunakan satu buah korpus paralel yaitu 1000 kalimat bahasa Indonesia, 1000 kalimat bahasa Lampung dialek *api* dan satu buah mono korpus 4000 kalimat bahasa Lampung dialek *api*,
2. strategi kedua, eksperimen menggunakan satu buah korpus paralel yaitu 2000 kalimat bahasa Indonesia, 2000 kalimat bahasa Lampung dialek *api* dan satu buah mono korpus 4000 kalimat bahasa Lampung dialek *api*,
3. strategi ketiga, eksperimen menggunakan satu buah korpus paralel yaitu 3000 kalimat bahasa Indonesia, 3000 kalimat bahasa Lampung dialek *api* dan satu buah mono korpus 4000 kalimat bahasa Lampung dialek *api*.

Strategi pertama digunakan sebagai *baseline* untuk pengamatan hasil penerjemahan bahasa Indonesia ke bahasa Lampung dialek *api*, kemudian akan dilihat hasilnya pada strategi kedua dan ketiga. Adapun pengujian penelitian ini dilakukan dengan menggunakan 100 kalimat uji terdiri dari 25 kalimat tunggal tanpa *out of vocabulary* (OOV) disingkat menjadi TTOOV, 25 kalimat tunggal dengan OOV disingkat menjadi TDOOV, 25 kalimat majemuk tanpa OOV disingkat menjadi MTOOV dan 25 kalimat majemuk dengan OOV disingkat menjadi MDOOV.

2.2. Statistical Machine Translation

Dalam melakukan uji coba penerjemahan secara statistik bahasa Indonesia ke bahasa Lampung dialek *api*, peneliti menggunakan arsitektur SMT yang digunakan dalam penelitian menggunakan penelitian[5]. Ujicoba ini di mulai dari fase prapemrosesan yang merupakan tahapan awal untuk mempersiapkan korpus paralel. Kemudian dilanjutkan dengan fase *training*, yakni fase pengolahan korpus paralel untuk memperoleh *language model* dan *translation model*. Selanjutnya fase *testing*, dan diakhiri dengan fase evaluasi. Untuk alat-alat (*tools*) yang digunakan dalam ujicoba SMT pada bahasa Indonesia-Lampung dialek *api* adalah :

1. Linux ubuntu 16.04 LTS, sistem operasi yang digunakan dalam menjalankan Moses.
2. Moses, perangkat lunak untuk proses *preprocessing* dan *decoding*.

3. KenLM, perangkat lunak telah terpasang di Moses, untuk *Language Model*.
4. Giza++, perangkat lunak untuk *Translation Model*.
5. BLEU, perangkat lunak untuk proses evaluasi.

2.3. Implementasi SMT Pada Moses Decoder

Fase prapemrosesan SMT pada Moses Decoder terdiri dari beberapa bagian, yakni penjajaran kalimat, tokenisasi, *cleaning*, *lowercase filtering*, dan *truecase*. Penjajaran kalimat adalah proses menjajarkan korpus paralel bahasa Indonesia ke bahasa Lampung dialek *api* sebagai terjemahannya. Tokenisasi diperlukan untuk memberi jarak antar kata termasuk juga memberi jarak antara kata dengan tanda baca yang ada, sedangkan *lowercase* merupakan proses untuk menyeragamkan besar-kecilnya huruf. Detail sintaks yang digunakan dalam penelitian ini terdapat di tautan

(<https://drive.google.com/drive/folders/1eIOQcgZCXSYtaoVoqSbp0w9JcFgWGBjM?usp=sharing>)

3. HASIL DAN PEMBAHASAN

Hasil eksperimen pada SMT bahasa Indonesia ke bahasa Lampung dialek *api* disajikan secara detail pada tautan (<https://drive.google.com/drive/folders/1eIOQcgZCXSYtaoVoqSbp0w9JcFgWGBjM?usp=sharing>). Eksperimen menggunakan tiga strategi pada bahan penelitian memberikan hasil yang menarik untuk diamati secara numerik dan analisis pada masing-masing bagian-bagian pada 100 kalimat uji terdiri dari 25 kalimat tunggal tanpa *out of vocabulary* (OOV) disingkat menjadi TTOOV, 25 kalimat tunggal dengan OOV disingkat menjadi TDOOV, 25 kalimat majemuk tanpa OOV disingkat menjadi MTOOV dan 25 kalimat majemuk dengan OOV disingkat menjadi MDOOV. Akurasi hasil penerjemahan dari bahasa Indonesia ke bahasa Lampung dialek *api* disajikan di tabel 2 dalam persen (%).

Tabel 2. Hasil skor BLEU eksperimen SMT bahasa Indonesia ke bahasa Lampung dialek *api*

Kalimat Uji	Strategi 1	Strategi 2	Strategi 3
TTOOV	21.49	59.58	73.21
TDOOV	23.22	44.33	68.72
MTOOV	18.22	39.4	69.18
MDOOV	25.94	28.22	71.94

Pada strategi pertama, hasil yang didapat adalah rendahnya akurasi penerjemahan bahasa Indonesia ke bahasa Lampung dialek *api* baik pada bagian TTOOV, TDOOV, MTOOV dan MDOOV. Nilai akurasi tertinggi hanya dicapai pada MDOOV, eksperimen penerjemahan kalimat uji berupa kalimat majemuk dengan ada OOV didalamnya, sebesar 25.94 %. Hal menarik yang muncul pada strategi pertama ini adalah hasil akurasi pada uji coba kalimat tunggal dan kalimat majemuk yaitu terjadi peningkatan akurasi ketika dengan sengaja diberikan kata yang tidak ada di dalam korpus paralel atau disebut OOV. Hal ini disebabkan minimnya informasi n-grams yang diperoleh dari hasil pembangkitan model bahasa menggunakan korpus paralel sejumlah 1000 kalimat Indonesia, 1000 kalimat bahasa Lampung dialek *api*.

Pada strategi kedua, eksperimen menggunakan satu buah korpus paralel yaitu 2000 kalimat bahasa Indonesia, 2000 kalimat bahasa Lampung dialek *api* dan satu buah mono korpus 4000 kalimat bahasa Lampung dialek *api*. Hasilnya menunjukkan bahwa penambahan korpus paralel 1000 kalimat secara signifikan meningkatkan akurasi pada TTOOV sebesar 38.09 % dari sebelumnya strategi pertama, pada TDOOV sebesar 21.11 % dari sebelumnya strategi pertama, pada MTOOV sebesar 21.18 % dari sebelumnya strategi pertama dan pada MDOOV sebesar 2.28 % dari sebelumnya strategi pertama. Peningkatan yang paling kecil terjadi di bagian MDOOV.

Tabel 3. Hasil peningkatan akurasi

Peningkatan Akurasi	Dari Strategi 1 menjadi Strategi 2	Dari Strategi 2 menjadi Strategi 3
TTOOV	38.09	13.63
TDOOV	21.11	24.39
MTOOV	21.18	29.78
MDOOV	2.28	43.72

Pada strategi ketiga, eksperimen menggunakan satu buah korpus paralel yaitu 3000 kalimat bahasa Indonesia, 3000 kalimat bahasa Lampung dialek *api* dan satu buah mono korpus 4000 kalimat bahasa Lampung dialek *api*. Hasilnya menunjukkan bahwa penambahan korpus paralel 2000 kalimat secara signifikan meningkatkan akurasi pada TTOOV sebesar 51.72 % dari sebelumnya strategi pertama, pada TDOOV sebesar 45.5 % dari sebelumnya strategi pertama, pada MTOOV sebesar 50.96 % dari sebelumnya strategi pertama dan pada MDOOV sebesar 46 % dari sebelumnya strategi pertama. Peningkatan jumlah korpus paralel secara signifikan meningkatkan akurasi hasil penerjemahan dari bahasa Indonesia ke bahasa Lampung dialek *api*.

3.1 Analisis Hasil TTOOV

Hasil eksperimen pada SMT bahasa Indonesia ke bahasa Lampung dialek *api* yang disajikan pada tabel 2 menunjukkan informasi peningkatan akurasi dari strategi 1, strategi 2 menuju strategi 3. Detail hasil ada di tautan(<https://drive.google.com/drive/folders/1eIOQcgZCXSytAoVoqSbp0w9JcFgWGBjM?usp=sharing>). Hasil eksperimen, menggunakan 25 kalimat tunggal tanpa *out of vocabulary* (OOV) disingkat menjadi TTOOV, menunjukkan peningkatan akurasi sebesar 38.09 % dari strategi 1 menuju strategi 2 sedangkan dari strategi 2 menuju strategi 3 hanya menambah 13.63 % akurasinya. Detail hasil SMT TTOOV ada di tautan di bagian wah ini.

Hasil pengamatan yang diberikan pada tautan tersebut menunjukkan pada strategi pertama SMT memberikan hasil penerjemahan berupa kalimat yang berisi campuran antara kata bahasa Indonesia dan kata dalam bahasa Lampung dialek *api* walaupun demikian SMT mampu memberikan hasil berupa kalimat. Sebuah contoh hasil SMT diberikan pada tabel 4 di bawah ini. Pada strategi 2, SMT secara umum memberikan hasil sepenuhnya kalimat berbahasa Lampung *api* walaupun susunan posisinya kalimatnya masih tertukar sedangkan pada strategi 3 hasil SMT lebih baik dari strategi 2.

Tabel 4. Contoh hasil TTOOV

Kalimat Uji (TTOOV)	saya tidak mencuci baju di sungai
Hasil terjemahan acuan oleh penutur bahasa Lampung Api	nyak mak mepoh kawai di wai
Output strategi 1	nyak mak mencuci baju dipa-dipa sungai
Output strategi 2	nyak mak mepoh kawai di wai
Output strategi 3	nyak mak mepoh kawai di wai

3.2 Analisis Hasil TDOOV

Hasil eksperimen pada SMT bahasa Indonesia ke bahasa Lampung dialek *api* yang disajikan pada tabel 2 menunjukkan informasi peningkatan akurasi dari strategi 1, strategi 2 menuju strategi 3. Detail hasil ada di tautan(<https://drive.google.com/drive/folders/1eIQcgZCXSYtaoVoqSbp0w9JcFgWGBjM?usp=sharing>).

Pada TDOOV pada prinsipnya menggunakan kalimat yang sama seperti pada TTOOV. Perbedaan mendasarnya adalah adanya penambahan kata yang berperan sebagai *out of vocabulary* (OOV). Hasil eksperimen, menggunakan 25 kalimat tunggal dengan *out of vocabulary* (OOV) disingkat menjadi TDOOV, menunjukkan adanya pengaruh pada hasil penerjemahan SMT sehingga peningkatan akurasi hanya sebesar 21.11 % dari strategi 1 menuju strategi 2 akan tetapi dari strategi 2 menuju strategi 3 hasilnya lebih tinggi yaitu menambah 24.39 % akurasinya.

Hasil pengamatan yang diberikan pada tautan tersebut menunjukkan pada strategi pertama SMT memberikan hasil penerjemahan berupa kalimat yang berisi campuran antara kata bahasa Indonesia dan kata dalam bahasa Lampung dialek *api* walaupun demikian SMT mampu memberikan hasil berupa kalimat. Sebuah contoh hasil SMT diberikan pada tabel 5 di bawah ini. Pada strategi 2, SMT secara umum memberikan hasil sepenuhnya kalimat berbahasa Lampung *api* walaupun demikian kata OOV belum terakomodasi dengan baik bahkan hilang sedangkan pada strategi 3 hasil SMT lebih baik dari strategi 2.

Tabel 5. Contoh hasil TDOOV

Kalimat Uji (TDOOV)	saya tidak mencuci baju di sungai brantas
Hasil terjemahan acuan oleh penutur bahasa Lampung <i>Api</i>	nyak mak mepoh kawai di wai brantas
<i>Output</i> strategi 1	nyak mak mencuci baju dipa-dipa sungai brantas
<i>Output</i> strategi 2	nyak mak brantas mepoh kawai di wai
<i>Output</i> strategi 3	nyak mak mepoh kawai di wai brantas

3.3 Analisis Hasil MTOOV

Hasil eksperimen pada SMT bahasa Indonesia ke bahasa Lampung dialek *api* yang disajikan pada tabel 2 menunjukkan informasi peningkatan akurasi dari strategi 1, strategi 2 menuju strategi 3. Detail hasil ada di tautan(<https://drive.google.com/drive/folders/1eIQcgZCXSYtaoVoqSbp0w9JcFgWGBjM?usp=sharing>).

Pada MTOOV pada prinsipnya menggunakan kalimat majemuk guna melihat kemampuan SMT dalam menerjemahkan kalimat bahasa Indonesia yang tidak sederhana tapi dengan tetap mempertahankan tanpa OOV. Hasil eksperimen, menggunakan 25 kalimat majemuk dengan *out of vocabulary* (OOV) disingkat menjadi MTOOV, menunjukkan adanya pengaruh pada hasil penerjemahan SMT sehingga peningkatan akurasi hanya sebesar 21.18 % dari strategi 1 menuju strategi 2 akan tetapi dari strategi 2 menuju strategi 3 hasilnya lebih tinggi yaitu menambah 29.78 % akurasinya. Artinya penambahan jumlah korpus paralel berkontribusi positif.

Hasil pengamatan yang diberikan pada tautan tersebut menunjukkan pada strategi pertama SMT memberikan hasil penerjemahan berupa kalimat yang berisi campuran antara kata bahasa Indonesia dan kata dalam bahasa Lampung dialek *api* walaupun demikian SMT mampu memberikan hasil berupa kalimat. Sebuah contoh hasil SMT diberikan pada tabel 6 di bawah ini. Pada strategi 2, SMT secara umum memberikan hasil sepenuhnya

kalimat berbahasa Lampung *api* walaupun susunan kata yang dibentuk masih tertukar secara posisi dan masih ada kata yang berbahasa Indonesia sedangkan pada strategi 3 hasil SMT lebih baik dari strategi 2.

Tabel 6. Contoh hasil MTOOV

Kalimat Uji (TDOOV)	tuan-tuan yth memancing walau anak-anak nya tidak ada yang menjaga di pantai
Hasil terjemahan acuan oleh penutur bahasa Lampung <i>Api</i>	kuti ghumpok ngawil walau sanak-sanak ni mak wat sai ngejaga di pantai
<i>Output strategi 1</i>	tuan-tuan yth memancing walau anak-anak di mak ngedok sai menjaga di pantai
<i>Output strategi 2</i>	sanak-sanak kuti ghumpok memancing kipak ngejaga ni mak ngedok sai di pantai
<i>Output strategi 3</i>	kuti ghumpok ngawil walau sanak-sanak ni mak ngedok sai ngejaga di pantai

3.3 Analisis Hasil MDOOV

Hasil eksperimen pada SMT bahasa Indonesia ke bahasa Lampung dialek *api* yang disajikan pada tabel 2 menunjukkan informasi peningkatan akurasi dari strategi 1, strategi 2 menuju strategi 3. Detail hasil ada di tautan(<https://drive.google.com/drive/folders/1eIQcgZCXSytAoVqSbp0w9JcFgWGBjM?usp=sharing>).

Pada MDOOV pada prinsipnya menggunakan kalimat majemuk guna melihat kemampuan SMT dalam menerjemahkan kalimat bahasa Indonesia yang tidak sederhana dan dengan sengaja menambahkan sebuah kata OOV. Hasil eksperimen, menggunakan 25 kalimat majemuk dengan *out of vocabulary* (OOV) disingkat menjadi MDOOV, menunjukkan hampir tidak adanya pengaruh pada penambahan jumlah paralel korpus dari hasil penerjemahan SMT yang hanya meningkat sebesar 2.28 % dari strategi 1 menuju strategi 2 akan tetapi dari strategi 2 menuju strategi 3 hasilnya lebih tinggi yaitu menambah 43.72 % akurasi. Artinya penambahan jumlah korpus paralel berkontribusi positif walaupun terdeteksi adanya OOV. Penambahan OOV memberikan kontribusi positif pada *language model* yang dihasilkan.

Hasil pengamatan yang diberikan pada tautan tersebut menunjukkan pada strategi pertama SMT memberikan hasil penerjemahan berupa kalimat yang berisi campuran antara kata bahasa Indonesia dan kata dalam bahasa Lampung dialek *api* walaupun demikian SMT mampu memberikan hasil berupa kalimat. Sebuah contoh hasil SMT diberikan pada tabel 7 di bawah ini. Pada strategi 2, SMT secara umum memberikan hasil sepenuhnya kalimat berbahasa Lampung *api* walaupun demikian kata penambahan jumlah korpus paralel tidak berkontribusi positif dan juga dipengaruhi OOV sedangkan pada strategi 3 hasil SMT lebih baik dari strategi 2.

Tabel 7. Contoh hasil MDOOV

Kalimat Uji (MDOOV)	tuan-tuan yth memancing di laut teluk kiluan walau anak-anak nya tidak ada yang menjaga di pantai
Hasil terjemahan acuan oleh penutur bahasa Lampung Api	kuti ghumpok ngawil di lawok teluk kiluan walau sanak-sanak ni mak wat sai ngejaga di pantai
Output strategi 1	tuan-tuan yth memancing dipa-dipa laut teluk kiluan walau anak-anak di mak ngedok sai menjaga di pantai
Output strategi 2	kuti ghumpok teluk memancing di laok kiluan kipak ngejaga sanak-sanak ni mak ngedok sai di pantai
Output strategi 3	kuti ghumpok ngawil di lawok teluk kiluan walau sanak-sanak ni mak ngedok sai ngejaga di pantai

4. KESIMPULAN

Efek peningkatan jumlah korpus paralel pada eksperimen mesin penerjemah menggunakan metode SMT menunjukkan peningkatan hasil yang positif berdasarkan perolehan pada tabel 2 dan tabel 3. Secara umum pada hasil eksperimen di strategi 1 diperoleh hasil penerjemahan yang berupa kalimat walaupun berisi campuran antara bahasa Indonesia dan bahasa Lampung dialek *api*, pada strategi 2 diperoleh hasil penerjemahan yang berupa kalimat bahasa Lampung dialek *api* walaupun masih tertukar posisi kata pada tempat seharusnya dan masih ditemukan sedikit kata berbahasa Indonesia, sedangkan pada strategi 3 memberikan hasil yang terbaik dibandingkan pada strategi 1 dan 2.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kementerian Riset dan Teknologi / Badan Riset dan Inovasi Nasional (KEMENRISTEK/BRIN) yang telah memberi dukungan pendanaan / financial terhadap penelitian ini melalui skema Penelitian Dosen Pemula (PDP) tahun 2020 dengan nomor kontrak 023/UTI/LPPM/E.1.3/VII/2020.

DAFTAR PUSTAKA

- [1] Z. Abidin, "Penerapan Neural Machine Translation untuk Eksperimen Penerjemahan secara Otomatis pada Bahasa Lampung – Indonesia," *Pros. Semin. Nas. Metod. Kuantitatif 2017*, no. 978, pp. 53–68, 2017.
- [2] Z. Abidin, A. Sucipto, and A. Budiman, "Penerjemahan Kalimat Bahasa Lampung-Indonesia Dengan

- Pendekatan Neural Machine Translation Berbasis Attention Translation of Sentence Lampung-Indonesian Languages With Neural Machine Translation Attention Based,” *J. Kelitbangan*, vol. 06, no. 02, pp. 191–206, 2018.
- [3] P. Permata and Z. Abidin, “Statistical Machine Translation Pada Bahasa Lampung Dialek Api Ke Bahasa Indonesia,” *Media Inform. Budidarma*, vol. 4, no. 3, pp. 519–528, 2020, doi: 10.30865/mib.v4i3.2116.
- [4] S. Ningsih and S. Saniati, “Eksperimen Pengenalan Ucapan Aksara Lampung Dengan CMU Sphinx 4,” *J. Teknoinfo*, vol. 12, no. 1, p. 33, 2018, doi: 10.33365/jti.v12i1.40.
- [5] 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.
- [6] 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.*, 2016, doi: 10.1109/ICITSI.2015.7437678.
- [7] Megaria, “Afiks Pembentuk Adjektiva dalam Bahasa Lampung Dialek A Logat Belalau (Analisis Morfologis),” *J. LOKABASA*, vol. 4, no. 2, pp. 195–201, 2013.
- [8] H. Tanuwijaya and H. Manurung, Maruli, “Penerjemah Dokumen Inggris-Indonesia Menggunakan Mesin Penerjemah Statistik Dengan Word Reordering dan Phrase Reordering,” *J. Ilmu Komput. dan Inf.*, vol. 2, no. 1, pp. 17–24, 2009.
- [9] H. Sujaini, “Meningkatkan Peran Model Bahasa dalam Mesin Penerjemah Statistik (Studi Kasus Bahasa Indonesia-Dayak Kanayatn),” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 3, no. 2, p. 51, 2017, doi: 10.23917/khif.v3i2.4398.
- [10] T. Apriani, H. Sujaini, and N. Safriadi, “Pengaruh Kuantitas Korpus Terhadap Akurasi Mesin Penerjemah Statistik Bahasa Bugis Wajo Ke Bahasa Indonesia,” *J. Sist. dan Teknol. Inf.*, vol. 1, no. 1, pp. 1–6, 2016.
- [11] C. Adiputra, Krisna and Y. Arase, “Performance of Japanese-to-Indonesian Machine Translation on Different Models,” *Proc. 23rd Annu. Meet. Linguist. Process. Soc.*, no. C, pp. 7–10, 2017.
- [12] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “BLEU: a Method for Automatic Evaluation of Machine Translation,” *Proc. 40th Annu. Meet. Assoc. Comput. Linguist.*, pp. 311–318, 2002, doi: 10.1002/andp.19223712302.
- [13] F. Ariyani, “Distribusi Verba Berfrefiks (N-) Pada Bahasa Lampung dalam Kitab Kuntara Raja Niti dan Buku Ajar. Ranah: Jurnal Kajian Bahasa 3,” *Ranah J. Kaji. Bhs.*, vol. 3, no. 2, pp. 124–134, 2014, doi: <https://doi.org/10.26499/rnh.v3i2.43>.
- [14] P. Bhattacharyya, *Machine Translation*. Boca Raton: Taylor & Francis Group, 2015.
- [15] D. Jurafsky and J. Marti, H, *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, vol. 3, no. 2. Prentice-Hall, Inc., 2008.