**Jurnal TEKNOINFO**, Vol. 15, No. 1, 2021, 13-19, ISSN: 2615-224X
DOI : 10.33365/jti.v15i1.889

# PENGARUH PENAMBAHAN KORPUS PARALEL PADA MESIN PENERJEMAH STATISTIK BAHASA INDONESIA KE BAHASA LAMPUNG DIALEK NYO

**Zaenal Abidin[1)], Permata [2)]**

*[1,2]Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia*
*[1,2]Jl. Z.A. Pagaralam No.9-11 Labuhan Ratu, Bandar Lampung*
*Email: [1]permata@teknokrat.ac.id, [2]zabin@teknokrat.ac.id*

**Abstract**

Lampung Province, a province located on the island of Sumatera. People in Lampung province use two main regional languages, namely Lampung dialect of Api and Nyo. For the immigrants in Lampung province, they have difficulty in communicating with the indigenous people of Lampung. As an alternative, both immigrants and the indigenous people of Lampung speak Indonesian. The research potential of translating Indonesian into Lampung is widely open. This research is focused on building a model for translating Indonesian into the Lampung language dialect of Nyo with Statistical Machine Translation approach and the tool used for it, is Moses. The research begins with (1) making a parallel corpus of Indonesian and its translation in the Lampung language dialect of Nyo, then making mono corpus of the Lampung language dialect of Nyo, (2) building a language model, (3) building a translation model, (4) combining language models and models translation using the Moses decoder to be able to translate from Indonesian to Lampung language dialect of Nyo. The research experiment was focused on observing the effect of adding a parallel corpus from 1000 sentences, 2000 sentences and 3000 sentences. This research uses 100 test sentences that are not in the parallel corpus or mono corpus. The results showed that the accuracy value when using 1000 sentences, 2000 sentences and 3000 sentences was the accuracy value of the bilingual evaluation under-study, respectively, namely 25.91%, 43.18% and 45.26%.

**Keyword:** Corpus parallel Indonesian - Lampung; Dialect of Nyo; Statistical Machine Translation; Bilingual evaluation under-study.

**Abstrak**

Provinsi Lampung merupakan salah satu provinsi yang terletak di pulau Sumatera. Masyarakat provinsi Lampung menggunakan dua bahasa daerah yaitu bahasa Lampung dialek Api dan bahasa Lampung dialek Nyo. Untuk para pendatang di provinsi Lampung memiliki permasalahan dalam komunikasi dengan penduduk asli Lampung. Sebagai alternatinya, baik pendatang dan penduduk asli Lampung berkomunikasi dengan bahasa Indonesia. Potensi penelitian penerjemahan bahasa Indonesia ke Lampung terbuka lebar. Penelitian ini difokuskan untuk membangun model penerjemahan bahasa Indonesia ke dalam dialek bahasa Lampung Nyo dengan pendekatan Mesin Statistik Terjemahan dan alat yang digunakan untuk itu adalah *Moses*. Penelitian diawali dengan (1) membuat korpus paralel bahasa Indonesia dan terjemahannya dalam dialek bahasa Lampung bahasa Nyo, kemudian membuat mono korpus bahasa Lampung dialek Nyo, (2) membangun model bahasa, (3) membangun terjemahan model, (4) memadukan model bahasa dan model terjemahan menggunakan *Decoder Moses* untuk dapat menerjemahkan dari bahasa dialek bahasa Indonesia ke bahasa Lampung Nyo. Eksperimen penelitian difokuskan untuk mengamati pengaruh penambahan korpus paralel dari 1000 kalimat, 2000 kalimat dan 3000 kalimat. Penelitian ini menggunakan 100 kalimat tes yang tidak berada dalam korpus paralel atau korpus mono. Hasil penelitian menunjukkan bahwa nilai akurasi saat menggunakan 1000 kalimat, 2000 kalimat dan 3000 kalimat merupakan nilai akurasi evaluasi bilingual yang diteliti berturut-turut yaitu 25.91%, 43.18% dan 45.26%.

**Kata Kunci:** *Korpus paralel Indonesia-Lampung, Dialek Nyo, Statistical Machine Translation, Bilingual evaluation under-study.*

## 1. Pendahuluan

Provinsi Lampung adalah salah satu provinsi yang terletak di pintu gerbang masuk menuju Pulau Sumatera. Provinsi Lampung memiliki kekayaan budaya, salah satunya adalah bahasa Lampung dan aksara Lampung. Secara umum di Provinsi Lampung terdapat dua dialek utama yaitu dialek api dan dialek nyo. Pemerintah Provinsi Lampung memiliki perhatian besar terhadap bahasa Lampung. Berbagai upaya terus dilakukan pemerintah provinsi untuk melestarikan dan memelihara bahasa Lampung. Pemerintah Lampung melalui Peraturan Gubernur Nomor 39 Tahun 2014 tentang Mata Pelajaran Bahasa dan Literasi Lampung menetapkan bahwa bahasa Lampung merupakan muatan lokal wajib di tingkat satuan pendidikan dasar hingga menengah dan didukung dengan ketersediaan buku teks mulai dari SD, SMP. dan SMA, beserta kamus bahasa Lampung. Bahasa Lampung baik dialek api maupun dialek nyo digunakan masyarakat Lampung untuk berkomunikasi sehari-hari baik di lingkungan keluarga maupun di acara-acara adat. Bahasa Lampung termasuk dalam kelas Austronesia dalam rumpun bahasa Melayu Polinesia. Dua dialek utama tersebut adalah dialek api dan dialek nyo yang mengacu pada kata "apa" [1].

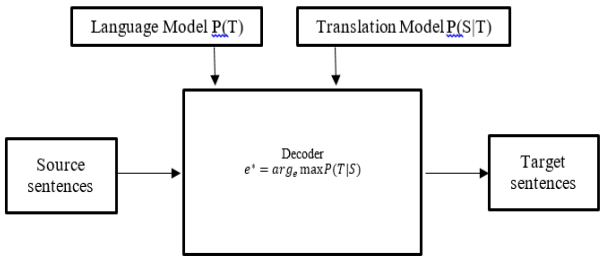
Beberapa penelitian terkait bahasa Lampung telah dilakukan oleh para peneliti. Penelitian penerjemahan bahasa Lampung dalam dialek api menggunakan metode *Neural Machine Translation* (NMT) tanpa mekanisme *Attention* [2] dan metode *Neural Machine Translation* (NMT) dengan mekanisme *Attention* [3]. Penelitian tentang *Statistical Machine Translation* (SMT) pada kalimat api dialek Lampung [4]. Sedangkan penelitian bahasa Lampung dari aspek tuturan terlebih dahulu dilakukan [5]. Sebagian besar penelitian SMT dilakukan di Universitas Tanjungpura, termasuk SMT dari Bugis Wajo ke Indonesia [6] dan SMT dari Indonesia ke Sunda [7]. Eksperimen SMT Indonesia-Jepang [8], Pengembangan SMT Indonesia-Jepang dengan Terjemahan Lemma dan Proses Posting Tambahan [9] dan SMT Inggris-Bengla [10]. Penelitian ini bertujuan untuk membangun model bahasa Indonesia SMT - Lampung dialek nyo dan melihat pengaruh penambahan korpus paralel terhadap akurasi terjemahan dengan skor *Bilingual Evaluasi Under-study* (BLEU). Kontribusi dan kebaruan dalam penelitian ini adalah (1) penelitian ini merupakan penelitian pertama SMT untuk bahasa Indonesia - Lampung dialek nyo, (2) model SMT untuk bahasa Indonesia - Lampung dialek nyo, (3) terjemahan bahasa Indonesia ke bahasa Lampung dialek nyo akan dilakukan secara komputasi.

Pada bagian II, penjelasan tentang transalasi mesin statistik dan Evaluasi Otomatis. Pada bagian III, penjelasan tentang metode penelitian. Pada bagian IV, penjelasan hasil dan analisis. Di Bagian V tentang kesimpulan penelitian. Berkaitan dengan penelitian ini adalah penelitian sebelumnya tentang *Neural Machine Translation* untuk bahasa Lampung dialek api dengan mekanisme atensi [2] dan tanpa mekanisme atensi [3], penelitian tentang statistik mesin terjemahan untuk dialek api bahasa Lampung [4].

## 2. Statistical Machine Translation

### 2.1. Model Bahasa, Model Terjemahan dan *Decoder*

Secara umum SMT memiliki arsitektur dengan beberapa komponen yaitu model bahasa, model terjemahan dan *decoder*. Model bahasa digunakan untuk menemukan kefasihan terjemahan. Model terjemahan berfungsi untuk menemukan ketepatan terjemahan. Sedangkan *decoder* berfungsi untuk mencari teks dalam bahasa target yang memiliki nilai probabilitas tertinggi dengan mempertimbangkan model bahasa dan model terjemahan [11]. Arsitektur SMT ditunjukkan pada gambar 1 di bawah ini.




**Gambar 1.** Aritektur SMT [11]

Mesin terjemahan statistik adalah mesin yang menggunakan pendekatan statistik dalam mengolah terjemahan bahasa sumber ke dalam bahasa target. Pendekatan statistik yang digunakan adalah konsep probabilitas. Konsep probabilitas dalam mesin penerjemahan statistik mengasumsikan bahwa setiap kalimat T merupakan kemungkinan terjemahan dari kalimat S dalam bahasa sumber [12]. Melalui pendekatan bahwa teks terjemahan yang didasarkan pada distribusi probabilitas P (T | S) dapat dilakukan dengan Teorema Bayes, yaitu:

$$\boldsymbol{P(T|S) = \frac{P(S|T)}{P(S)} P(T)} \quad \textbf{(1)}$$

Bagian-bagian yang merupakan elemen kunci dalam model bahasa adalah probabilitas rangkaian kata yang ditulis sebagai P (w_1, w_2,…, w_n) atau P (w_ {1, n)). Model bahasa menetapkan probabilitas ke serangkaian n kata dengan rata-rata distribusi probabilitas. Urutannya bisa berupa frase atau kalimat dan probabilitasnya dapat diperkirakan dari kumpulan dokumen yang besar. Salah satu contoh pendekatan model bahasa adalah model n-gram. Model bahasa n-gram merupakan salah satu jenis model bahasa probalilistik untuk memprediksi item berikutnya dalam urutan pada bentuk (n-1). Probabilitas

bersyarat dapat dihitung dari jumlah frekuensi n-gram:

$$\boldsymbol{P(w_i|w_{i-(n-1)}, \dots, w_{i-1}) = \frac{count(w_{i-(n-1)}, w_{i-1}, \dots, w_i)}{count(w_{i-(n-1)}, \dots, w_{i-1})}} \quad \textbf{(2)}$$

Berikut adalah contoh model n-gram yaitu:

1. Unigram (1-gram): $P(w_1), P(w_2), \dots, P(w_n)$ .
2. Bigram (2-gram): $P(w_1), P(w_2|w_1), \dots, P(w_n|w_{n-1})$
3. Trigram (3-gram): $P(w_{1,n}) = P(w_1), P(w_2|w_1), P(w_3|w_{1,2}) .. P(w_{n-2}|w_{n-1})$

Model terjemahan digunakan untuk mencocokkan teks masukan dalam bahasa sumber dengan teks keluaran dalam bahasa target. Pada mesin penerjemahan statistik terdapat dua model terjemahan, yaitu model terjemahan berbasis kata (*word-based translation model*) dan model terjemahan berbasis frase (model terjemahan berbasis frase) [12]. *Decoder* bertugas menemukan teks dalam bahasa target yang memiliki probabilitas terbesar dengan mempertimbangkan model terjemahan dan faktor model bahasa [12]. Perhitungan $\boldsymbol{\hat{T}}$ (hasil terjemahan) dapat dituliskan sebagai berikut:

$$\boldsymbol{\hat{T} = arg_T max P(T|S) = arg_T \max \frac{P(S|T)}{P(S)} . P(T)}$$

$$\boldsymbol{\hat{T} = arg_T \max P(S|T) . P(T)} \quad \textbf{(3)}$$

**2.2 Evaluasi Otomatis**

Evaluasi hasil terjemahan dilakukan dengan membandingkan kalimat hasil terjemahan dengan kalimat acuan menggunakan *Bilingual Evaluation Understudy* (BLEU). BLEU merupakan algoritma yang berfungsi untuk mengevaluasi kualitas hasil terjemahan yang telah diterjemahkan mesin dari bahasa sumber ke bahasa tujuan. BLEU mengukur skor presisi berbasis statistik yang telah dimodifikasi antara hasil terjemahan, secara otomatis, dengan terjemahan referensi menggunakan konstanta yang disebut hukuman singkat (BP) [13].

$$\text{BP}_{\text{BLEU}} = \begin{cases} 1 & \text{jika } c > r \\ e^{(1-\frac{r}{c})} & \text{jika } c \leq r \end{cases}$$

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\,gram \in C} Count_{clip}(n\,gram)}{\sum_{C' \in \{Candidates\}} \sum_{n\,gram' \in C'} Count_{clip}(n\,gram')} \quad \textbf{(4)}$$

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

BP adalah simbol penalti singkat, c adalah jumlah kata dari hasil terjemahan mesin, r adalah panjang terjemahan referensi efektif. pn adalah rata-rata geometris dari presisi n-gram yang dimodifikasi. Nilai N yang digunakan adalah N = 4 dan $\boldsymbol{w_n = \frac{1}{N}}$ [14].

**2.3 Karakteristik Bahasa Lampung**

Bahasa Lampung termasuk dalam kelas rumpun bahasa Austronesia selain Polinesia Melayu. Bahasa Lampung memiliki dua dialek utama yang hidup berdampingan dan keduanya aktif digunakan oleh setiap pengguna dialek tersebut. Dialek tersebut adalah dialek api dan dialek nyo yang mengacu pada kata 'apa' [1]. Bahasa Lampung memiliki struktur tata bahasa yang mirip dengan bahasa Indonesia. Ini berisi subjek, predikat, objek, deskripsi, dan lainnya. Kalimat dalam bahasa Lampung juga mirip dengan kalimat bahasa Indonesia, ada kalimat tunggal, kalimat majemuk, kalimat tanya, kalimat perintah, kalimat tanya, kalimat berita dan lain-lain.

Bahasa Lampung memiliki struktur tata bahasa yang mirip dengan bahasa Indonesia. Dalam bahasa Lampung ada mata pelajaran, predikat, benda, informasi dan lain-lain. Kalimat dalam bahasa Lampung juga mirip dengan kalimat bahasa Indonesia, ada kalimat tunggal, kalimat majemuk, kalimat tanya, kalimat perintah, kalimat tanya, kalimat berita dan lain-lain. Bagian ini menyajikan berbagai kalimat bahasa Lampung beserta terjemahannya dalam bahasa Inggris.

• Kalimat tunggal: Misalnya dalam bahasa Lampung, 'Burhan lapah mit sekula' berarti 'Burhan pergi ke sekolah'.

• Kalimat majemuk: Misalnya dalam bahasa Lampung 'Burhan lapah mit sekula walau badan ni mak sihat' artinya 'Burhan bersekolah padahal badannya tidak sehat'.

• Kalimat imperatif: Misalnya dalam bahasa Lampung, 'Mejong pai!' Berarti 'Duduk dulu!'.

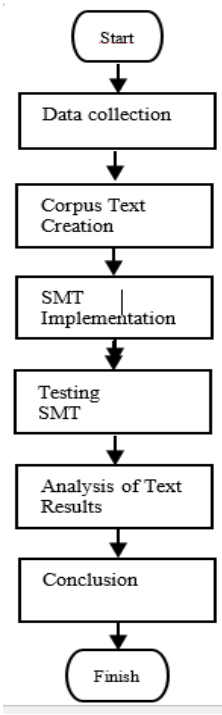
• Kalimat tanya: Misalnya dalam bahasa Lampung, 'Ulah api sanak lunik miwang teghus?' Berarti 'Mengapa anak-anak menangis sepanjang waktu?'

• Kalimat Berita: Misalnya dalam bahasa Lampung 'Indui becawa, nyak mak haga lijung mit Jakarata' artinya 'Bunda bilang, saya gak mau ke Jakarta'.

• Kalimat sempurna: Misalnya dalam bahasa Lampung 'Nyak ngebattu ulun tuhani di ghani sunday' artinya 'Saya bantu orang tua saya di hari Minggu'.

## 3. Metodologi Penelitian

Metodologi penelitian yang digunakan dalam penelitian penerjemahan dari bahasa Indonesia ke bahasa Lampung dialek nyo dijelaskan pada Gambar 2 berikut:



**Gambar 2.** Diagram Alir Penelitian

### 3.1 Pengumpulan Data

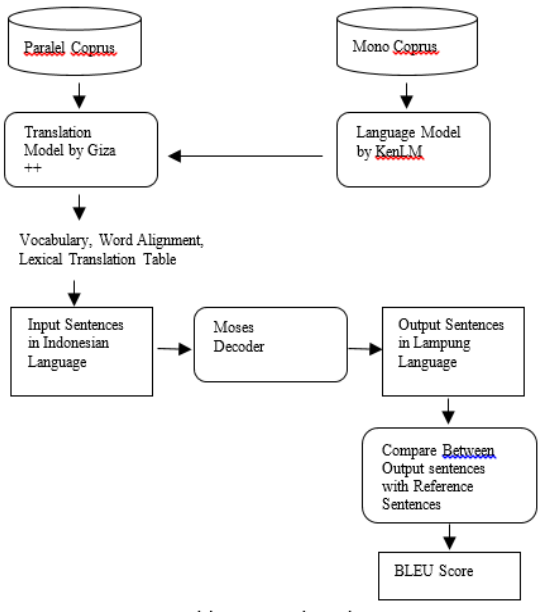
Data yang digunakan dalam penelitian ini adalah dokumen teks bahasa Indonesia dialek bahasa Lampung dokumen teks nyo yang bersumber dari buku teks bahasa Lampung yang digunakan di sekolah dasar dan menengah kemudian diolah menjadi teks korpus. Materi penelitian berupa pasangan 3000 korpus paralel Bahasa Indonesia dan Lampung Dialek Nyo, 3000 Kalimat Mono Corpus Bahasa Lampung Dialek Nyo dan 100 Kalimat Ujian Bahasa Indonesia.

### 3.2 Pembuatan Teks Korpus

Dokumen teks yang telah dikumpulkan selanjutnya dibuat menjadi korpus paralel dan monolingual. Dokumen tersebut terdiri dari 3000 pasang Dialek Indonesia - Lampung paralel nyo dan 3000 Dialek Lampung monolingual nyo. Korpus tersebut kemudian disimpan dalam format .lg untuk korpus bahasa Lampung dan .id untuk korpus bahasa Indonesia.

### 3.3 Implementasi SMT

Arsitektur sistem SMT bahasa Indonesia ke bahasa Lampung ditunjukkan pada Gambar 3.




**Gambar 3.** Arsitektur SMT Indonesia-Lampung

Gambar 3 menunjukkan desain arsitektur SMT Indonesia - Lampung. Arsitektur terdiri dari beberapa tahapan yaitu tahap pertama pembuatan korpus paralel dan mono corpus, pemodelan bahasa Lampung dengan KenLM, pemodelan terjemahan dengan GIZA ++, proses *decoding*, tahap terakhir adalah pengujian dan evaluasi hasil terjemahan.

Arsitektur SMT Bahasa Indonesia - Lampung Dialek Nyo yang dibangun dalam penelitian ini terdiri dari (1) kumpulan paralel 3000 pasang kalimat nyo bahasa Indonesia - Lampung, (2) 3000 kalimat mono korpus Lampung dalam dialek nyo (3) menghasilkan model bahasa dengan GIZA ++. GIZA ++ untuk menyelaraskan kata kopus paralel kami [15], (4) menghasilkan model terjemahan dengan KenLM, KenLM disertakan dalam *Moses* dan default dalam perangkat *Moses* [15]. Model bahasa (LM) digunakan untuk memastikan keluaran yang lancar, sehingga dibangun dengan bahasa target. Dokumentasi KenLM memberikan penjelasan lengkap tentang opsi baris perintah, tetapi berikut ini akan membangun model bahasa 3-gram yang sesuai [15]. (5) menggabungkan model bahasa dan model terjemahan dengan *Moses Decoder*. Tugas *Moses Decoder* adalah menemukan kalimat dengan skor tertinggi dalam bahasa target (menurut model terjemahan) yang sesuai dengan kalimat sumber yang diberikan [15]. (6) melakukan uji coba dengan *Moses*, (7) Menilai hasil terjemahan melalui skor BLEU.

### 3.4 Menguji SMT Dialek Bahasa Indonesia - Lampung Nyo

Pengujian hasil implementasi SMT Bahasa Indonesia - Lampung dilakukan dengan menggunakan 100 kalimat uji. Kalimat percobaan tidak ada dalam korpus paralel. Pengujian dilakukan dengan menggunakan alat *Moses* dengan melihat nilai BLEU yang diperoleh.

Skenario percobaan SMT dilakukan tiga kali. (1) Pada skenario eksperimen pertama dengan kondisi 1000 kalimat korpus paralel dan 3000 mono korpus, (2) pada skenario kedua eksperimen dengan kondisi 2000 kalimat korpus paralel dan 3000 mono korpus, (3) pada skenario eksperimen ketiga dengan kondisi 3000 kalimat korpus paralel dan 3000 mono korpus. Dalam setiap skenario pengujian dilakukan dengan 100 kalimat pengujian.

## 4. Analisis dan Hasil

### 4.1. Implementasi Dialek Nyo Bahasa Indonesia - Lampung

Pelaksanaan SMT Indonesia - Lampung Bahasa dialek nyo diawali dengan penyusunan korpus paralel dan monokorpus. Materi penelitian ini adalah 3000 kalimat korpus paralel Indonesia - Lampung dan 3000 kalimat mono korpus bahasa Lampung. Dalam *Moses* Manual [15], ada tiga aktivitas yang dilakukan dalam tahap persiapan korpus, yaitu (1) tokenisasi: Artinya harus ada spasi di antara (mis.) Kata dan tanda baca (2) truecasing: Kata-kata awal di setiap kalimat diubah menjadi casing yang paling memungkinkan. Ini membantu mengurangi ketersebaran data. (3) pembersihan: Kalimat panjang dan

kalimat kosong dihilangkan karena dapat menyebabkan masalah pada *training*, dan kalimat yang jelas tidak selaras dihapus.

Tahap selanjutnya adalah membangun model bahasa. Model bahasa (LM) digunakan untuk memastikan keluaran yang lancar, sehingga dibangun dengan bahasa target [15] (dalam hal ini Dialek Bahasa Lampung dari Nyo). Akhirnya kita sampai pada acara utama - melatih model terjemahan. Untuk melakukan ini, kami menjalankan wordalignment (menggunakan GIZA ++), ekstraksi frase dan penilaian, membuat tabel *reordering lexicalised* [15]. Tahap terakhir dari SMT adalah *decoding*. *Decoding* menguji 100 kalimat percobaan dalam Bahasa Indonesia untuk mendapatkan 100 kalimat Bahasa Lampung Dialek Nyo. *Decoding* menggunakan *Moses*.

### 4.2. Implementasi KenLM untuk Model Langauge

Pada penelitian SMT Indonesia-Lampung, model bahasa digunakan sebagai sumber informasi nilai probabilitas distribusi kata yang dibangun dengan menggunakan metode n-gram. Contoh model Bahasa ditunjukkan pada gambar 4 di bawah ini.

```
\data\
ngram 1=3861
ngram 2=13755
ngram 3=17282

\1-grams:
-3.564323       selamat   -0.08786054
-2.9411547      tukuk     -0.19878922
-2.4330168      nyo       -0.22172873
……
\2-grams:
-1.0940454      selamat </s>      0
-0.67361087     tukuk </s>        0
```

**Gambar 4.** Contoh Tabel Model Bahasa Indonesia-Lampung [15]

### 4.3. Implementasi Giza ++ untuk Model Terjemahan

Pada penelitian SMT Indonesia - Lampung, model terjemahan yang dibangun dari korpus paralel Bahasa Indonesia - Lampung menggunakan Giza ++. Model terjemahan berfungsi untuk memasangkan teks masukan dari bahasa Indonesia ke teks keluaran bahasa Lampung. Hasil model penerjemahan berupa file dokumen korpus kosakata, perataan kata dan tabel model leksikal. Contoh kosakata ditunjukkan pada gambar 5 di bawah ini

```
1       UNK     0
2       ikam    343
3       ?       278
4       nyo     97
5       mak     81
6       ijo     81
7       no      75
8       ago     74
9       wat     67
10      sekam   67
..      …       ...
```

**Gambar 5.** Contoh kosakata korpus bahasa Lampung [15]

### 4.4. Penerapan Decoding

Setelah model bahasa dan model terjemahan dibangun, langkah selanjutnya adalah menyatukan kedua model tersebut menggunakan decoder moses. Pengujian secara otomatis menggunakan 100 kalimat pengujian dalam skenario pertama. Hasil dari skenario pertama akan digunakan sebagai dasar untuk skenario kedua dan ketiga. Pengujian otomatis akan menghasilkan nilai evaluasi *bilingual* (BLEU). Nilai BLEU akan digunakan sebagai parameter perbandingan antara skenario pertama, skenario kedua dan skenario ketiga.

**Gambar 6.** Contoh sintaks untuk *Decoding Moses* [15]

### 4.5. Hasil Eksperimen

Dalam penelitian ini skenario pertama digunakan sebagai

```
~/mosesdecoder/bin/moses -f ~/working/mert-work/moses.ini

~/mosesdecoder/bin/moses-f ~/RISET/ttraining3/TRAINING3/mert-work/moses.ini
<~/RISET/ttraining3/testing3/testing4A.id>
~/RISET/ttraining3/testing3/Hasiltesting4A.lg

~/mosesdecoder/scripts/generic/multi-bleu.perl -lc
~/RISET/ttraining3/testing3/Acuantesting4A.lg <
~/RISET/ttraining3/testing3/Hasiltesting4A.lg
```

base line. Skenario pertama menggunakan 1000 kalimat korpus paralel bahasa Indonesia - Lampung dan 3000 kalimat mono corpus bahasa Lampung. Skenario kedua menggunakan 2000 kalimat parallel corpus bahasa Indonesia - Lampung dan 3000 kalimat mono corpus bahasa Lampung dan skenario ketiga menggunakan 3000 kalimat parallel corpus Indonesia - Lampung dan 3000 kalimat mono corpus bahasa Lampung. Hasil skenario ditunjukkan pada tabel 1 di bawah ini.

**Tabel 1.** Skor Bleu Untuk Skenario Yang Berbeda

| **BLUE Score** | **Result of Statistical Machine Translation** | | |
|---|---|---|---|
| | **First Scenario** | **Second Scenario** | **Third Scenario** |
| Indonesian – Lampung Dialect of Nyo | 25.91 % | 43.18 % | 45.26 % |

Fakta yang diperoleh dari tabel 1 bahwa skenario pertama, kedua dan ketiga berupa penambahan jumlah korpus paralel berpotensi meningkatkan akurasi penerjemahan TPS Bahasa Indonesia-Lampung. Eksperimen dilakukan dalam tiga skenario. Skenario pertama menghasilkan 25.91%. Skenario kedua menghasilkan 43,18% dan skenario ketiga menghasilkan

skor BLEU sebesar 45,26%. Seperti pada tabel 2 di bawah ini akan diberikan sampel terkait dengan hasil skenario ketiga.

**Tabel 2.** Contoh Hasil Skenario Ketiga

| **Indonesian Input Sentences** | **Reference Senteces in Lampung Language dialect of *nyo*** | **Result from SMT Lampung dialect of Nyo** |
|---|---|---|
| saya tidak mencuci baju di sungai | ikam mak muppeh kaway di way balak | ikam mak **nginyau kawai** di **wai** balak |
| tono memberi makanan ikan di kolam ani | tono ngejuk kanen punyeu di kulam ani | tono ngejuk kanen punyeu di kulam ani |
| anak yang disuntik itu tidak sehat | sanak si disuttik ino mak sihat | sanak si disuttik ino sihat |
| dia tidur saja | yo pedem gaweh | yo pedem gaweh |
| mereka menangkap maling di dusun ini | tiyan ninjuk maling di anek ijo | tiyan **nakkep** maling di anek ijo |

Informasi dari tabel 2 menunjukkan lima sampel kalimat tes yang digunakan dalam penelitian ini dari skenario ketiga. Kolom pertama pada tabel 2 adalah kalimat tes bahasa Indonesia, kolom kedua adalah kalimat referensi untuk terjemahan kalimat tes dalam bahasa Lampung dialek nyo dan kolom ketiga adalah hasil dari SMT.
Pada contoh kalimat pertama skenario ketiga terlihat hasil terjemahannya berbeda dengan referensi. SMT memberi 'nginyau' untuk kata 'cuci' dalam bahasa Indonesia. Pada contoh kalimat kedua dan keempat, SMT memberikan hasil yang sama dengan referensinya. Pada contoh kalimat ketiga, TPS kurang dalam menerjemahkan kalimat masukan terutama pada kata 'tidak'. Pada contoh kalimat kelima skenario ketiga menunjukkan hasil terjemahannya berbeda dengan referensi. SMT memberi 'nakkep' untuk kata 'menangkap' dalam bahasa Indonesia.

## 5. Kesimpulan dan Keterlanjutan Penelitian

Penerjemahan Bahasa Indonesia ke Bahasa Lampung Dialek Nyo dapat dilakukan dengan menggunakan metode Mesin Terjemahan Statistik. Penambahan jumlah korpus paralel berdampak pada peningkatan akurasi hasil terjemahan Bahasa Indonesia ke Bahasa Lampung Dialek Nyo berdasarkan nilai BLEU. Pada penggunaan 1000 corpus parallel mendapatkan nilai BLEU sebesar 25,91%, sedangkan pada penggunaan korpus paralel 2000 mendapatkan nilai BLEU sebesar 43,18% sedangkan pada korpus paralel sebesar 3000 memperoleh nilai BLEU sebesar 45,26%. Penelitian lanjutan tentang SMT Bahasa Indonesia dan Lampung dapat dikembangkan melalui (1) mengamati perubahan jumlah mono korpus, (2) menambahkan POS tag ke korpus paralel, (3) meningkatkan kualitas korpus paralel dan mono korpus.

## Ucapan Terimakasih

Publikasi ini merupakan hasil hibah penelitian dosen pemula dari Kementerian Riset dan Teknologi / Badan Riset dan Inovasi Nasional (Kemenristek / BRIN) dengan nomor kontrak 078/SP2H/LT/DRPM/2020, 839/SP2H/LT/MONO/LL2/2020,046/UTI/LPPM/E.1.3/ VII/2020.

## Daftar Pustaka

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] R. Darwis, H. Sujaini, and R. D. Nyoto, "Peningkatan Mesin Penerjemah Statistik dengan Menambah Kuantitas Korpus Monolingual (Studi Kasus : Bahasa Indonesia - Sunda)," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 1, p. 27, 2019, doi:

10.26418/justin.v7i1.27254.

[8] H. S. Simon and A. Purwarianti, "Experiments on Indonesian-Japanese statistical machine translation," *Proceeding - IEEE Cybern. 2013 IEEE Int. Conf. Comput. Intell. Cybern.*, pp. 80–84, 2013, doi: 10.1109/CyberneticsCom.2013.6865786.

[9] M. A. Sulaeman and A. Purwarianti, "Development of Indonesian-Japanese statistical machine translation using lemma translation and additional post-process," *Proc. - 5th Int. Conf. Electr. Eng. Informatics Bridg. Knowl. between Acad. Ind. Community, ICEEI 2015*, no. i, pp. 54–58, 2015, doi: 10.1109/ICEEI.2015.7352469.

[10] A. H. Imam, M. R. Mahmud Arman, S. H. Chowdhury, and K. Mahmood, "Impact of corpus size and quality on English-Bangla statistical machine translation system," *14th Int. Conf. Comput. Inf. Technol. ICCIT 2011*, no. Iccit, pp. 566–571, 2011, doi: 10.1109/ICCITechn.2011.6164853.

[11] 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.

[12] 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.

[13] A. Hermanto, T. Adji, and N. A. Setiawan, "Recurrent neural network language model for English-Indonesian Machine Translation: Experimental study.," in *ICSITech*, 2015, pp. 132–136.

[14] 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.

[15] P. Koehn, *Statistical Machine Translation SystemUser Manual and Code Guide*. 2019.