

TUNING FOR QUALITY UNTUK UJI AKURASI MESIN PENERJEMAH STATISTIK (MPS) BAHASA INDONESIA - BAHASA DAYAK KANAYATN

Muhammad Hasbiansyah¹, Herry Sujaini², Novi Safriadi³.

Program Studi Teknik Informatika Fakultas Teknik Universitas Tanjungpura¹²³

e-mail: m.hasbiansyah@gmail.com¹, herry_sujaini@yahoo.com², bangnops@gmail.com³

Abstrak—Bahasa merupakan alat komunikasi yang penting bagi manusia, karena dengan menggunakan bahasa kita dapat mengetahui dan mengetahui informasi yang dibutuhkan, dengan bahasa juga manusia dapat mengekspresikan diri, menyampaikan kritik dan pendapat, pikiran serta keinginannya. Namun tidak semua orang memiliki bahasa yang sama, sehingga hal ini dapat menghalangi dalam melakukan pertukaran informasi. Oleh karena itu, saat ini salah satu teknologi yang sedang dikembangkan yaitu mesin penerjemah untuk mengatasi masalah penerjemahan bahasa. Namun kualitas dari hasil terjemahan yang dihasilkan masih mengandung keterbatasan. Hasil terjemahan yang optimal dapat diperoleh dengan menggunakan konsep penerjemahan statistik. Mesin Penerjemah Statistik (*Statistical Machine Translation*) merupakan sebuah pendekatan mesin penerjemah dengan hasil terjemahan yang dihasilkan atas dasar model statistik yang parameter-parameternya diambil dari hasil analisis korpus paralel. Tujuan yang ingin dicapai dalam penelitian ini adalah melakukan proses *tuning* pada mesin penerjemah statistik bahasa Indonesia ke bahasa Dayak Kanayatn untuk mengetahui pengaruh proses *tuning* terhadap hasil terjemahan. Pengujian dilakukan dengan membandingkan skor akurasi hasil terjemahan sebelum dan setelah proses *tuning*. Penelitian menggunakan korpus paralel sebanyak 3667. Pengujian dilakukan secara otomatis menggunakan *Bilingual Evaluation Understudy* (BLEU). Hasil dari pengujian adalah terdapat peningkatan skor BLEU sebesar 3,04%. Berdasarkan hasil pengujian, tingkat akurasi terjemahan setelah proses *tuning* pada mesin penerjemah statistik bahasa Indonesia ke bahasa Dayak Kanayatn dengan mengubah bobot parameter-parameter *phrase translation*, *language model*, *distortion* dan *word penalty* dapat meningkatkan nilai akurasi hasil terjemahan.

Kata Kunci— *Tuning for quality*, mesin penerjemah statistik, BLEU, bahasa Indonesia, bahasa Dayak Kanayatn

I. PENDAHULUAN

Bahasa merupakan alat komunikasi yang penting bagi manusia, karena dengan menggunakan bahasa kita dapat mengetahui dan mengetahui informasi yang dibutuhkan. Dengan bahasa juga manusia dapat mengekspresikan diri, menyampaikan kritik dan pendapat, pikiran serta keinginannya. Bahasa juga merupakan alat yang digunakan untuk sosialisasi dan adaptasi sosial antara manusia satu dengan manusia lainnya agar kehidupannya semakin berkembang. Pada tahun 2012, penelitian berlanjut dengan mengambil sampel di 70 lokasi di wilayah Maluku dan

Papua. Hasil dari penelitian itu, jumlah bahasa dan sub bahasa di seluruh Indonesia mencapai 546 bahasa dan kemungkinan jumlahnya akan terus bertambah karena penelitian yang dilakukan belum selesai [1].

Sama seperti bahasa daerah lain nya, bahasa Dayak Kanayatn juga memiliki fungsi yang sama, yaitu sebagai lambang kebanggaan daerah, lambang identitas daerah dan alat penghubung antarwarga masyarakat daerah. Namun fungsi dari bahasa daerah tersebut saat ini mengalami banyak penurunan. Salahsatu upaya yang dilakukan untuk mempertahankan penggunaan bahasa daerah, yaitu dengan menjadikan salah satu mata pelajaran muatan lokal pada tingkat sekolah dasar. Namun belum memadai untuk menjadi solusi dari permasalahan. Ada beberapa faktor yang menjadi penyebab. Faktor kedwibahasaan atau kemultibahasaan yang berkembang di masyarakat, perbedaan kelompok umur, pendidikan, wilayah permukiman, jenis kelamin. tidak menutup kemungkinan penyebab pergeseran bahasa tidak hanya dipengaruhi oleh faktor diatas, akan tetapi juga dapat disebabkan oleh adanya perbedaan etnis atau adat istiadat penutur bahasa dari mana mereka berasal [2].

Berdasarkan faktor-faktor yang telah dijabarkan, untuk menghindari terjadinya kepunahan bahasa daerah, salah satu caranya adalah dengan mesin penerjemah statistik. Mesin penerjemah statistik (*Statistical Machine Translation*) merupakan sebuah pendekatan mesin penerjemah dengan hasil terjemahan yang dihasilkan atas dasar model statistik yang parameter-parameternya diambil dari hasil analisis korpus paralel.

Namun kualitas hasil terjemahan dengan konsep penerjemahan statistik juga belum sepenuhnya baik, banyak cara yang dapat dilakukan untuk meningkatkan kualitas hasil terjemahan dari sebuah mesin penerjemah statistik, salah satu cara yang dapat dilakukan adalah dengan proses *tuning for quality*.

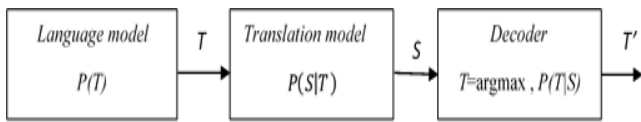
Berdasarkan uraian tersebut, maka akan dilakukan penelitian untuk membangun dan mengimplementasikan mesin penerjemah statistik bahasa Indonesia ke bahasa Dayak Kanayatn. Selanjutnya, akan dilakukan proses *tuning for quality* untuk mengetahui pengaruh proses *tuning for quality* tersebut terhadap kualitas hasil terjemahan.

II. URAIAN PENELITIAN

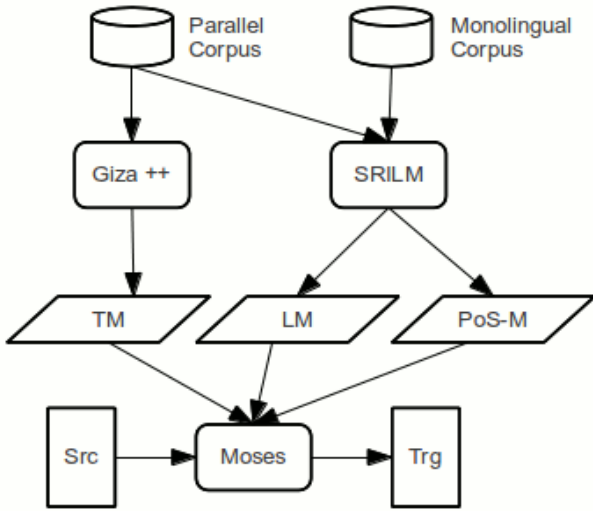
A. Mesin Penerjemah Statistik

Mesin penerjemah statistik merupakan salah satu jenis mesin penerjemah dengan menggunakan pendekatan statistik. Terdapat 3 komponen yang terlibat dalam proses penerjemahan dari satu bahasa ke bahasa lain pada mesin penerjemah statistik, yaitu *language model*, *translation*

model, dan decoder.



Gambar 1. Komponen mesin penerjemah statistik [3]



Gambar 2. Arsitektur mesin pnerjemah statistik Moses [4]

Sumber data utama yang dipergunakan adalah parallel corpus dan monolingual corpus. Proses training terhadap parallel corpus menggunakan GIZA++ menghasilkan translation model (TM). Proses training terhadap bahasa target pada parallel corpus ditambah dengan monolingual corpus bahasa target menggunakan SRILM menghasilkan language model (LM), sedangkan PoS model (PoS-M) dihasilkan dari bahasa target pada parallel corpus yang setiap katanya sudah ditandai dengan PoS. TM, LM dan PoS-M digunakan untuk menghasilkan decoder Moses. Selanjutnya Moses digunakan sebagai mesin penerjemah untuk menghasilkan bahasa target dari input kalimat dalam bahasa sumber.

B. Moses

Moses adalah salah satu Mesin Penerjemah Sattistik yang memungkinkan untuk menerjemahkan secara otomatis setiap pasangan bahasa. Moses digunakan untuk melatih model statistik teks terjemahan dari bahasa sumber ke bahasa sasaran. Saat melakukan penerjemahkan bahasa, Moses membutuhkan korpus dalam dua bahasa, bahasa sumber dan bahasa sasaran. Moses dirilis di bawah lisensi LGPL (Lesser General Public License) dan tersedia sebagai kode sumber dan binari untuk Windows dan Linux. Perkembangannya didukung oleh proyek EuroMatrix, dengan pendanaan oleh European Commission [5].

C. Korpus

Korpus didefinisikan sebagai koleksi atau sekumpulan contoh teks tulis atau lisan dalam bentuk data yang dapat dibaca dengan menggunakan seperangkat mesin dan dapat diberi catatan berupa berbagai bentuk informasi linguistik [6]. Kriteria korpus yang telah menjadi kesepakatan banyak ahli, yakni: dapat dibaca dengan menggunakan seperangkat mesin, berupa teks otentik, digunakan sebagai sampel dan mewakili bahasa atau variasi bahasa tertentu.

D. Automatic Evaluation

Sitem evaluasi otomatis yang populer saat ini adalah BLEU (Bilingual Evaluation Understudy). BLEU adalah sebuah algoritma yang berfungsi untuk mengevaluasi kualitas dari sebuah hasil terjemahan yang telah diterjemahkan oleh mesin dari satu bahasa alami ke bahasa lain. BLEU mengukur modified n-gram precision score antara hasil terjemahan otomatis dengan tejemahan rujukan dan menggunakan konstanta yang dinamakan brevity penalty.

Nilai BLEU didapat dari hasil perkalian antara brevity penalty dengan rata-rata geometri dari modified precision score. Semakin tinggi nilai BLEU, maka semakin akurat dengan rujukan. Sangat penting untuk diketahui bahwa semakin banyak terjemahan rujukan per kalimatnya, maka akan semakin tinggi nilainya. Untuk menghasikan nilai BLEU yang tinggi, panjang kalimat hasil terjemahan harus mendekati panjang dari kalimat referensi dan kalimat hasil terjemahan harus memiliki kata dan urutan yang sama dengan kalimat referensi. Rumus BLEU sebagai berikut [7]:

$$BP_{BLEU} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases} \tag{2.1}$$

$$P_n = \frac{\sum C_e \text{ corpus } n\text{-gram} \sum C_r \text{ count}_{clip}(n\text{-gram})}{\sum C_e \text{ corpus } n\text{-gram} \sum C_r \text{ count}(n\text{-gram})}$$

$$BLEU = BP_{BLEU} \cdot e^{\sum_{n=1}^N w_n \log P_n} \tag{2.2}$$

Keterangan:

- BP = brevity penalty
- c = jumlah kata dari hasil terjemahan otomatis
- r = jumlah kata rujukan
- P_n = modified precision score
- w_n = 1/N (standar nilai N untuk BLEU adalah 4)
- p_n = jumlah n-gram hasil terjemahan yang sesuai dengan rujukan dibagi jumlah n-gram hasil terjemahan

III. HASIL DAN DISKUSI

A. Data Penelitian

Data penelitian berupa buku biografi dengan judul Chairul Tanjung Si Anak Singkong. Buku biografi tersebut selanjutnya diolah menjadi korpus teks paralel bahasa Indonesia dan bahasa Dayak Kanayatn. Adapun jumlahnya yaitu 3667 pasangan kalimat korpus paralel bahasa Indonesia dan bahasa Dayak Kanayatn. Dengan jumlah 7495 token untuk korpus bahasa Indonesia dan 8886 token untuk bahasa Dayak Kanayatn.

B. Implementasi Mesin Penerjemah Statistik Jawa ke Bahasa Indonesia

1. Implementasi SRILM

Model bahasa digunakan sebagai sumber pengetahuan berbasis teks dengan nilai-nilai probabilistik. Penelitian ini menggunakan n-gram sebagai language model. Model bahasa dibangun dengan tools SRILM. Model bahasa akan menghasilkan output dengan format file *.lm. Berikut merupakan tabel model bahasa yang dihasilkan oleh SRILM pada mesin penerjemah statistik bahasa Indonesia-bahasa Dayak Kanayatn.

```
\data\
ngram 1=8888
ngram 2=38487
ngram 3=2853

\1-grams:
-2.758733 ahe -0.4582965
-2.904671 ayukng -0.2597015
-----
\2-grams
-1.992702 agi' ahe -0.1748312
-0.4792891 ahe pun -0.2520465
-----
\3-grams
-1.071469 nana' agi' minta
-1.923558 nang aku kanal
```

Gambar 3. Tabel model bahasa dengan bahasa Dayak *Kanayatr* sebagai bahasa target

2. Implementasi Giza++ Untuk Pemodelan Translasi

Model translasi digunakan untuk memasang teks input dalam bahasa sumber dengan teks output dalam bahasa target. Model translasi dibangun dengan *tools* Giza++. Proses pemodelan translasi oleh Giza++ menghasilkan dokumen *vocabulary corpus*, *word alignment* dan tabel model translasi. Dokumen-dokumen tersebut terdapat dalam folder “train” yang didalamnya terdapat 4 file yaitu “corpus, giza.dy-id, giza.id-dy dan model”.

```
1 UNK 0
2 , 3642
3 ka' 1785
4 man 1671
5 nang 1342
```

Gambar 4. Dokumen *vocabulary corpus* bahasa Dayak *Kanayatr*

Angka 1 sampai 5 pada dokumen *vocabulary corpus* merupakan *uniq id* untuk setiap data token, sedangkan angka disebelah kanan token menunjukkan frekuensi kemunculan. *Vocabulary corpus* yang dihasilkan mesin penerjemah bahasa Indonesia – bahasa Dayak *Kanayatr* terdiri dari 7495 token untuk korpus bahasa Indonesia dan 8886 token untuk bahasa Dayak *Kanayatr*.

```
# Sentence pair (2) source length 17 target length
17 alignment score : 1.91958e-07

saudara-saudaraku , pimpinan dan seluruh karyawan
perusahaan di bawah naungan ct corp yang saya cintai
dan banggaan

NULL ({} ) saudara-saudaraku ({} 1 {} ) , ({} 2 {} )
pimpinan ({} 3 {} ) man ({} 4 {} ) samua ({} 5 {} ) karyawan
({} 6 {} ) parusahaan ({} 7 {} ) ka' ({} 8 {} ) babah ({} 9 {} )
naungan ({} 10 {} ) ct ({} 11 {} ) corp ({} 12 {} ) nang ({}
13 {} ) aku ({} 14 {} ) nyintai' ({} 15 {} ) man ({} 16 {} )
mangga'atn ({} 17 {} )
```

Gambar 5. Dokument *alignment* bahasa Indonesia - bahasa Dayak *Kanayatr*

Dokumen *alignment* Bahasa Jawa–bahasa Indonesia terdapat tiga baris kalimat. Baris pertama berisi letak kalimat target (2) dalam korpus, panjang kalimat sumber (17), panjang kalimat target (17) dan skor *alignment*. Baris kedua merupakan bahasa sumber dan baris ketiga merupakan *alignment* kalimat bahasa target terhadap kalimat bahasa sumber. Kata “samua” ({} 5 {}) memiliki makna bahwa kata “samua” pada kalimat bahasa target, di-align ke kata kelima pada kalimat bahasa sumber yaitu “seluruh”.

```
rasa ketakutan ||| rasa gali' ||| 1 0.0952381 1
0.214286
rekan terdekat ||| ayukng tasamak ||| 1 0.0377358 1
0.571429
selalu ada dalam jiwa ||| salalu ada dalamp jiwa ||| 1
0.436471 1 0.270184
selama beberapa hari ||| salama sangahe ari ||| 1
0.645863 1 0.59931
selepas menjabat ||| laka' manjabat ||| 1 0.0411985 1
0.0572917
```

Gambar 6. Tabel model translasi mesin penerjemah bahasa Indonesia-bahasa Dayak *Kanayatr*

Proses pemodelan translasi oleh Giza++ akan menghasilkan tabel model translasi yang terdiri dari tabel kata yang berisi sekumpulan kata-kata yang telah dipadankan antara bahasa sumber dengan bahasa target yang memiliki nilai probabilitas.

C. Pengujian Hasil Terjemahan Mesin Translasi

Pengujian hasil translasi dilakukan dengan cara pengujian otomatis dari mesin penerjemah. Pengujian otomatis dari mesin penerjemah menghasilkan keluaran berupa nilai akurasi yang dihasilkan oleh BLEU (*Bilingual Evaluation Understudy*). Hasil pengujian ini nantinya akan menjadi parameter untuk membandingkannya dengan hasil pengujian setelah dilakukan proses *tuning*.

Langkah pada pengujian otomatis, korpus yang akan diuji terlebih dahulu melalui langkah translasi otomatis yang akan memberikan output berupa korpus dalam bahasa target yang telah diterjemahkan oleh mesin. Korpus uji yang digunakan pada tahap ini berjumlah 3667.

Setelah membuat output berupa hasil translasi otomatis dari mesin penerjemah, langkah selanjutnya adalah mendapatkan skor dari output dengan cara membandingkan output tersebut dengan korpus manual bahasa target yang telah dibuat sebelumnya.

```
hasbi@hasbi:~$
~/NLP/mosesdecoder/scripts/generic/multi-bleu.perl
~/NLP/data9/id-dy.lowercased.dy <
~/NLP/tuning/1,1,1,0.dy
BLEU = 89,47, 96.1/92.4/89.0/85.7 (BP=0.986, ratio
=0.986, hyp_len=57361, ref_len=58150)
```

Gambar 6. Tampilan skor BLEU dari 1,1,1,0.dy

Berdasarkan Gambar 6 diperoleh nilai awal dari 1,1,1,0.dy sebesar 89,47%.

D. Melakukan Proses Tuning For Quality

Setelah mendapatkan skor BLEU awal dari korpus uji, maka langkah selanjutnya adalah melakukan proses tuning for quality. Proses tuning for quality yang dilakukan pada file *moses.ini* yang terdapat pada folder *train* dan subfolder *model*. Berikut proses yang dilakukan, meliputi:

• Mengubah bobot dari parameter-parameter,

```

lanjut x moses.ini x
# Input factors
[input-factors]
0

# mapping steps
[mapping]
0 T 0

[distortion-limit]
6

# feature functions
[feature]
UnknownWordPenalty
WordPenalty
PhrasePenalty
PhraseDictionaryMemory name=TranslationModel0 num-features=4 path=/
home/hasbi/NLP/data9/train/model/phrase-table.gz input-factor=0 output-
factor=0
LexicalReordering name=LexicalReordering0 num-features=6 type=wbe-msd-
bidirectional-fe-allff input-factor=0 output-factor=0 path=/home/hasbi/
NLP/data9/train/model/reordering-table.wbe-msd-bidirectional-fe.gz
Distortion
KENLM lazyken=0 name=LMO factor=0 path=/home/hasbi/NLP/data9/dayak.lm
order=3

# dense weights for feature functions
[weight]
UnknownWordPenalty0= 1
WordPenalty0= -1.9
PhrasePenalty0= 0.1
TranslationModel0= 0.2 0.2 0.2 0.2
LexicalReordering0= 0.3 0.3 0.3 0.3 0.3 0.3
Distortion0= 0.7
LMO= 0.8
    
```

Gambar 7. Tampilan Mengubah Bobot Parameter

• Membuat output setelah mengubah parameter-parameter

```

hasbi@hasbi:~$
FeatureFunction: UnknownWordPenalty0 start: 0 end: 0
line=WordPenalty
FeatureFunction: WordPenalty0 start: 1 end: 1
line=PhrasePenalty
FeatureFunction: PhrasePenalty0 start: 2 end: 2
line=PhraseDictionaryMemory name=TranslationModel0 num-features=4 path=/
home/hasbi/NLP/data9/train/model/phrase-table.gz input-factor=0 output-f
actor=0
FeatureFunction: TranslationModel0 start: 3 end: 6
line=LexicalReordering name=LexicalReordering0 num-features=6 type=wbe-m
sd-bidirectional-fe-allff input-factor=0 output-factor=0 path=/home/hasbi
/NLP/data9/train/model/reordering-table.wbe-msd-bidirectional-fe.gz
Initializing Lexical Reordering Feature..
FeatureFunction: LexicalReordering0 start: 7 end: 12
line=Distortion
FeatureFunction: Distortion0 start: 13 end: 13
line=KENLM lazyken=0 name=LMO factor=0 path=/home/hasbi/NLP/data9/dayak.
lm order=3
FeatureFunction: LMO start: 14 end: 14
Loading the LM will be faster if you build a binary file.
Reading /home/hasbi/NLP/data9/dayak.lm
-----5---10---15---20---25---30---35---40---45---50---55---60---65---70---
75---80---85---90---95---100
*****
Loading UnknownWordPenalty0
Loading WordPenalty0
Loading PhrasePenalty0
Loading LexicalReordering0
Loading table into memory...done.
Loading Distortion0
Loading LMO
Loading TranslationModel0
Start loading text phrase table. Moses format : [3.901] seconds
Reading /home/hasbi/NLP/data9/train/model/phrase-table.gz
-----5---10---15---20---25---30---35---40---45---50---55---60---65---70---
75---80---85---90---95---100
*****
    
```

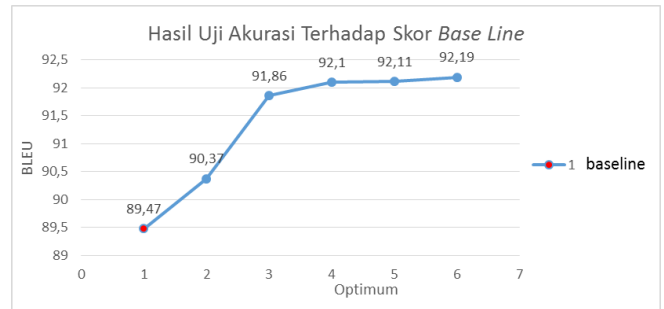
Gambar 8. Tampilan Membuat Output Terjemahan

E. Pengujian Ulang Hasil Terjemahan Mesin Translasi

Langkah berikutnya adalah menguji kembali hasil terjemahan mesin translasi bahasa Indonesia-Bahasa Dayak *Kanayatn* yang telah melewati proses *tuning*. Langkah pengujian dilakukan dengan cara melakukan pengujian otomatis yang akan memberikan output berupa korpus dalam bahasa target yang telah diterjemahkan oleh mesin.

1. Pengujian Otomatis

Pengujian dilakukan dengan membagi ke dalam beberapa tahap. Pengujian dilakukan dengan cara membandingkan nilai BLEU hasil terjemahan otomatis dari mesin penerjemah Bahasa Indonesia-bahasa Dayak *Kanayatn* sebelum dan setelah melewati proses *tuning*. Pengujian pertama menghasilkan skor BLEU sebesar 89,47 dan dijadikan *baseline*. Pengujian kedua menghasilkan skor BLEU sebesar 90,37. Pengujian ketiga menghasilkan skor BLEU sebesar 91,86. Pengujian keempat menghasilkan skor BLEU sebesar 92,10. Pengujian kelima menghasilkan skor BLEU sebesar 92,11 dan pengujian keenam menghasilkan skor BLEU sebesar 92,19.



Gambar 9. Grafik Peningkatan Skor BLEU Terhadap Skor *Baseline*

```

hasbi@hasbi:~$
~/NLP/mosesdecoder/scripts/generic/multi-bleu.perl
~/NLP/data9/id-dy.lowercased.dy <
~/NLP/data9/lanjut/0.1,0.8,0.7,-1.9.dy
BLEU = 92,19, 96.3/93.5/90.9/88.4 (BP=1.000, ratio
=1.000, hyp_len=58126, ref_len=58150)
    
```

Gambar 10. Tampilan nilai BLEU setelah proses *tuning*

Berdasarkan Gambar 9 diketahui sebelum dilakukan proses *tuning*, skor BLEU pada korpus uji 3667 sebesar 89,47 dan setelah dilakukan proses *tuning* didapat skor BLEU sebesar 92,19. Terdapat peningkatan nilai BLEU sebesar 3,04% dilihat dari perbandingan sebelum dan sesudah dilakukan proses *tuning*.

IV. KESIMPULAN

A. Kesimpulan

Berdasarkan hasil analisis dan pengujian, maka kesimpulan yang dapat diambil sebagai berikut.

1. Mesin penerjemah statistik dapat diimplementasikan untuk menerjemahkan bahasa Indonesia-bahasa Dayak *Kanayatn*.
2. Berdasarkan hasil pengujian, proses *tuning* dapat meningkatkan nilai akurasi terjemahan mesin penerjemah bahasa Indonesia-bahasa Dayak *Kanayatn*. Sebelum dilakukan proses *tuning*, skor BLEU pada korpus uji 3667 sebesar 89,47 dan setelah dilakukan proses *tuning* didapat skor BLEU sebesar 92,19. Terdapat peningkatan nilai BLEU sebesar 3,04% dilihat dari perbandingan sebelum dan sesudah dilakukan proses *tuning*.

B. Saran

Beberapa saran yang dapat diberikan sebagai pengembangan dari penelitian ini adalah sebagai berikut.

1. Perlu dilakukan penelitian lanjutan untuk melakukan analisis dalam menghasilkan terjemahan bahasa Indonesia-bahasa Dayak *Kanayatn* dengan mempertimbangkan hubungan antar frase dalam kalimat.
2. Perlu adanya teknik penanganan OOV (*Out of Vocabulary*) untuk menerjemahkan kata-kata yang tidak terdapat dalam korpus sehingga dapat meningkatkan kualitas terjemahan mesin penerjemah statistik.
3. Perlu dilakukan pengecekan ulang terhadap korpus teks paralel untuk mencegah kesalahan penulisan.
4. Perlu penambahan jumlah korpus untuk meningkatkan kualitas terjemahan mesin penerjemah statistik.
5. Pada proses pengujian perlu dilakukan kombinasi antara empat parameter dengan bobot dua angka di belakang koma sejak awal pengujian.

DAFTAR PUSTAKA

- [1] Akuntono, Indra. 2012. Jumlah Ragam Bahasa di Indonesia. Jakarta : Dikti.
- [2] Triyono, Sulis. 2006. Pembahasan Hasil Penelitian: Pergeseran Bahasa Daerah Akibat Kontak Bahasa Melalui Pembauran. *LITERA* Vol 5 No 1.
- [3] Manning, Christopher D., Schutze, Hinrich. 2000. *Foundations Of Statistical Natural Language Processing*. London : The MIT Press Cambridge Massachusetts.
- [4] Sujaini, Herry., Negara, Arif Bijaksana Putra. 2015. *Analysis of Extended Word Similarity Clustering based Algorithm on Cognate Language*. Gujarat: ESRSA Publications Pvt. Ltd.
- [5] Koehn, Philipp. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session, Prague, Czech Republic.
- [6] McEnery, T., et al. 2006. *Corpus-Based Language Studies: An Advanced Resource Book*. Oxon: Routledge.
- [7] Tanuwijaya, Hansel., Manurung, Hisar Maruli. 2009. Penerjemahan Inggris-Indonesia Menggunakan Mesin Penerjemah Statistik Dengan Word Reordering dan Phrase Reordering. Jakarta, *Jurnal Ilmu Komputer dan Informasi* Vol 2 No 1.