



## Penggunaan Bahasa Indonesia sebagai *Pivot Language* pada Mesin Penerjemah Madura-Sunda dengan Metode *Transfer* dan *Triangulation*

Herry Sujaini<sup>1</sup>

<sup>1</sup>Jurusan Informatika, Fakultas Teknik, Universitas Tanjungpura

<sup>1</sup>hs@untan.ac.id

### Abstract

This paper is an attempt to focus on investigating the pivot (bridge) language technique, where the *pivot* language used to improve Statistical Machine Translation (SMT) quality. In this case, Indonesian is used as a *pivot* language, where each available corpus can be used to support the Madurese-Sundanese language pair. Experiments that have been carried out using the parallel corpus of the Indonesian-Madurese and Indonesian-Sundanese languages are equal to 5K and 6K sentences respectively, while the monolingual corpus used Malay, Sundanese and Indonesian each at 10K, 10K and 100K sentences. This study compares the results of applying the *Triangulation* and Transfer methods using Indonesian as a *pivot* language. The results of the research proved that the *Triangulation* method has better acceleration when compared to the Transfer method. From the results of the experiments conducted, the *Triangulation* method increased the average Indonesian pivot-based SMT testing by 6.18% for Madura-Sundanese SMT and 7.27% for Madurese-Sundanese SMT.

Keywords: statistical machine translation, *pivot* language, Indonesian, Madurese-Sundanese

### Abstrak

Makalah ini merupakan upaya untuk fokus pada investigasi teknik bahasa *pivot* (penghubung), di mana bahasa penghubung digunakan untuk meningkatkan kualitas Mesin Penerjemah Statistik (MPS) untuk bahasa bersumber daya rendah, yaitu bahasa Madura-Sunda. Dalam hal ini, bahasa Indonesia digunakan sebagai bahasa penghubung, di mana korpus yang tersedia masing-masing dapat digunakan untuk mendukung pasangan bahasa Madura-Sunda. Eksperimen yang dilakukan menggunakan korpus paralel bahasa Indonesia-Madura dan Indonesia-Sunda masing-masing sebesar sebesar 5K dan 6K kalimat, sedangkan korpus monolingual yang digunakan adalah bahasa Melayu, Sunda dan Indonesia masing-masing sebesar 10K, 10K, dan 100K kalimat. Penelitian ini membandingkan hasil penerapan metode *Triangulation* dan *Transfer* menggunakan bahasa Indonesia sebagai bahasa *pivot*. Hasil penelitian yang dilakukan menunjukkan bahwa metode *Triangulation* memiliki akurasi yang lebih baik, jika dibandingkan dengan metode Transfer. Dari hasil percobaan yang dilakukan, metode *Triangulation* rata-rata meningkatkan akurasi MPS berbasis *pivot* bahasa Indonesia sebesar 6,18% untuk MPS Madura-Sunda dan 7,27% untuk MPS Sunda-Madura.

Kata kunci: Mesin Penerjemah Statistik, bahasa penghubung, bahasa Indonesia, Madura-Sunda

© 2019 Jurnal RESTI

### 1. Pendahuluan

Tujuan dari Mesin Penerjemah Statistik (MPS) adalah untuk menerjemahkan urutan bahasa sumber menjadi bahasa target dengan menilai hasil terbaik dari kemungkinan hasil terjemahan dari kalimat sumber dan kalimat target serta memperhatikan nilai terbaik dari struktur kalimat hasilnya. Kuantitas yang cukup besar dari korpus paralel yang selaras mempengaruhi fungsi dan kinerja sistem MPS. Di sisi lain, mengumpulkan data paralel dwi bahasa (*parallel corpus*) dalam

praktiknya menjadi masalah karena biaya tinggi dan keterbatasan ruang lingkup yang sebagai akibatnya dapat membatasi penelitian terkait dan aplikasinya. Oleh karena itu, kelangkaan data paralel untuk banyak pasangan bahasa adalah salah satu masalah utama dalam MPS [1].

Korpus paralel untuk dwi bahasa daerah di Indonesia jarang ditemukan, termasuk bahasa-bahasa yang banyak penuturnya, seperti bahasa Melayu, Jawa, Sunda, Madura, Batak dan lain-lain. Bahkan untuk

bahasa sumber daya yang baik, seperti yang termasuk dalam Europarl [2], yang mencakup bahasa debat di Parlemen Eropa, kinerja MPS menurun secara signifikan ketika diterapkan pada domain yang sedikit berbeda. Oleh karena itu, dengan perubahan domain, kinerja kehilangan efisiensinya. Solusi umum untuk kurangnya data paralel adalah menggunakan teknik bahasa *pivot* [3]. Teknik ini digunakan untuk menghasilkan MPS saat korpus dwibahasa untuk bahasa sumber dan target tidak memadai. Masalah ini menjadi signifikan ketika ada bahasa dengan sumber daya NLP (*Natural Language Processing*) yang tidak efisien untuk dapat menyediakan data pada sistem MPS. Namun, ada sumber daya yang cukup antara mereka dan beberapa bahasa lainnya. Meskipun diklaim bahwa bahasa perantara tidak mengarah pada peningkatan dalam kasus umum, ide ini dapat digunakan sebagai metode sederhana untuk memperkaya kinerja terjemahan bahkan untuk sistem yang sudah ada [4].

Dalam sebuah penelitian yang dilakukan oleh Utiyama dan Isahara [5], penggunaan bahasa *pivot* melalui terjemahan frasa dan terjemahan kalimat telah dilakukan. Selain itu, Wu dan Wang [6] membahas tiga metode untuk strategi *pivot* dalam temuan mereka termasuk metode penerjemahan frase, yaitu *Transfer*, *Triangulation* dan *Synthetic corpus*. Beberapa peneliti menyelidiki sistem MPS dengan teknik bahasa *pivot*. Misalnya Babych dkk. [7] menggunakan bahasa Rusia sebagai *pivot* untuk menerjemahkan dari bahasa Ukraina ke bahasa Inggris. Perbandingan mereka mengungkapkan bahwa adalah mungkin untuk mencapai kualitas terjemahan yang lebih baik dengan pendekatan bahasa *pivot*. Habash dan Hu [8] membandingkan dua pendekatan untuk sistem penerjemah bahasa Arab-Cina dengan pendekatan *direct translation* melalui Bahasa Inggris sebagai bahasa *pivot*. Temuan penelitian mereka menunjukkan bahwa menggunakan bahasa Inggris sebagai bahasa *pivot* dalam kedua pendekatan mengungguli terjemahan langsung dari bahasa Arab ke bahasa Mandarin.

Dalam penelitian lain, Bakhshaei et al. [9] menggunakan bahasa Inggris sebagai bahasa penghubung saat menerjemahkan dari bahasa Persia ke bahasa Jerman dan menyimpulkan bahwa menggunakan teknik *pivot* dalam kombinasi level-frase mengungguli sistem terjemahan langsung. Selanjutnya, Al-Hunaity et al. [10] mempresentasikan perbandingan antara strategi strategi komunitas dalam sistem MPS Denmark-Arab. Menurut temuan mereka, diilustrasikan bahwa *pivoting* kalimat mengambil alih frasa yang berputar ketika korpora paralel umum tidak tersedia. Nakov dan Ng [11] mencoba untuk mengeksplorasi kesamaan antara bahasa yang miskin sumber daya dan bahasa yang kaya sumber daya untuk tugas terjemahan. Paul et al. [12] mendiskusikan kriteria yang harus dipertimbangkan untuk pemilihan bahasa *pivot* yang

baik. Penggunaan teknik pengolahan sumber daya tersegmentasi dipertunjukkan oleh Kunchukuttan dkk. [13]. Dabre dkk. [14] menggunakan *multiple decoding paths* (MDP) untuk mengatasi keterbatasan korpora berukuran kecil.

Penggunaan bahasa Indonesia sebagai *pivot language* pernah dibahas oleh beberapa peneliti sebelumnya. Suryani dkk. [15] menggunakan bahasa Indonesia sebagai *pivot language* untuk menerjemahkan bahasa Inggris ke bahasa Sunda dan Jawa, eksperimen menunjukkan bahwa teknik yang digunakan menghasilkan akurasi terjemahan yang relatif rendah. Ningtyas dkk. [16] yang menguji penggunaan bahasa Indonesia sebagai *pivot language* pada mesin penerjemah statistik Bahasa Inggris - Bahasa Melayu Sambah memperoleh hasil peningkatan nilai akurasi sebesar 5,015% BLEU (*bilingual evaluation understudy*) dari nilai pengujian semula yang didapatkan dari mesin penerjemah tanpa *pivot*.

Seperti umumnya mesin penerjemah antar dua bahasa, terjemahan antar bahasa daerah seperti bahasa Madura ke Sunda dibangun secara langsung dengan menggunakan korpus paralel bahasa Madura-Sunda yang tentu saja sangat sulit untuk didapatkan. Untuk bahasa-bahasa yang memiliki sumber daya rendah seperti ini, peneliti-peneliti sebelumnya berusaha membangun mesin penerjemah dengan pendekatan bahasa *pivot*. Metode yang digunakan dengan bahasa *pivot* tersebut perlu diteliti lebih jauh hasilnya jika diterapkan pada bahasa-bahasa daerah di Indonesia. Dalam makalah ini dibahas bagaimana bahasa *pivot* tersebut diimplementasikan untuk membangun MPS yang memiliki sumber daya rendah bahasa Madura-Sunda dan sebaliknya. Metode yang digunakan adalah *Triangulation Pivoting* yang membentuk model terjemahan baru dari dua model terjemahan, yaitu model terjemahan Madura ke Indonesia dan Indonesia ke Sunda untuk MPS bahasa Madura ke Sunda dan model terjemahan bahasa Sunda ke Indonesia dan Indonesia ke Madura untuk MPS bahasa Madura ke Sunda. Metode tersebut akan dibandingkan dengan metode lain yaitu metode *Transfer*.

## 2. Metode Penelitian

### 2.1. Mesin Penerjemah Statistik

Paradigma MPS sebagai elemen yang paling penting adalah gagasannya, yaitu probabilitas kalimat sumber dan target dapat menemukan terjemahan terbaik. Paradigma yang sering digunakan MPS pada model *loglinear* adalah MPS berbasis frase, MPS berbasis frase hirarkis, dan MPS berbasis ngram. Dalam percobaan ini menggunakan sistem MPS berbasis frase dengan kerangka kerja entropi maksimum [17].

$$\hat{y}_1 = \arg \max P(x|y) \quad (1)$$

Model MPS berbasis frase adalah contoh pendekatan *noisy-channel*, di mana dapat disajikan hipotesis terjemahan  $t$  sebagai kalimat target saat diberikan  $s$  sebagai kalimat sumber. Dengan menggunakan Aturan Bayes dari (2), rumusan pada (1) untuk terjemahan yang paling mungkin terjadi dapat dituliskan seperti yang ditunjukkan pada (3).

$$P(t/s) = P(t) * P(s/t) = P(s) \quad (2)$$

$$\arg \max P(t/s) = \arg \max P(t) * P(s/t) \quad (3)$$

dengan  $t$  merupakan kalimat target dan  $s$  merupakan kalimat sumber.  $P(t)$  merupakan *language model* dan  $P(s/t)$  merupakan *translation model*. Operasi *argmax* adalah proses pencarian yang dilakukan oleh *decoder* yang merupakan bagian dari sistem MPS.

Proses penerjemahan melibatkan pengelompokan kalimat sumber ke dalam frase sumber  $s$ , menerjemahkan setiap frasa sumber menjadi frasa target  $t$ , dan menyusun ulang frasa target ini untuk menghasilkan kalimat target  $\hat{t}$ .

### 2.2. Metode *Triangulation Pivoting*

Kumpulan data berkualitas tinggi tidak selalu tersedia untuk melatih sistem MPS. Salah satu cara yang mungkin untuk mengatasi kebuntuan ini adalah dengan menggunakan bahasa ketiga sebagai jembatan yang di dalamnya terdapat bilingual sumber-*pivot* dan *pivot*-target yang berkualitas tinggi. Strategi berbasis *pivot* yang digunakan untuk sistem MPS dapat diklasifikasikan ke dalam tiga kategori yaitu metode *Transfer*, *Triangulation* dan *Synthetic corpus* [6].

Strategi *pivot* terjemahan frasa pada penelitian ini adalah dengan membuat tabel terjemahan frasa Madura ke Indonesia dan Indonesia ke Sunda untuk MPS bahasa Madura ke Sunda dan model terjemahan Sunda ke Indonesia dan Indonesia ke Madura untuk MPS bahasa Madura ke Sunda.

Dalam teknik ini, frasa  $md$  dalam tabel frase sumber-*pivot* terhubung ke  $id$ , dan frasa ini dikaitkan dengan frasa dalam tabel frase *pivot*-target. Kami menautkan frasa  $md$  dan  $sd$  dalam tabel frase baru untuk sumber-target. Untuk skoring pasangan frase dari tabel-frase baru, anggap  $P(id|md)$  sebagai skor dari frasa Madura-Indonesia dan  $P(sd|id)$  sebagai skor dari frasa Indonesia-Sunda, maka skor dari frase pasangan baru  $md$  dan  $sd$ ,  $P(sd|md)$ , dalam tabel frase Madura-Sunda dihitung:

$$P(sd|md) = \sum_{id} P(sd, id|md) \quad (4)$$

$id$  adalah variabel tersembunyi yang merupakan singkatan dari frasa bahasa *pivot*.

$$P(sd|md) = \sum_{id} P(sd|id, md)P(id|md) \quad (5)$$

Asumsikan,  $md$  dan  $sd$  bersifat independen, untuk setiap  $e$ :

$$P(sd|md) \approx \sum_{id} P(sd|id)P(id|md) \quad (6)$$

Jika disederhanakan, penjumlahan pada semua frase  $e$  digantikan oleh nilai maksimal, maka Persamaan (6) akan menjadi dengan:

$$P(sd|md) \approx \max_e P(sd|id)P(id|md) \quad (7)$$

Begitu juga sebaliknya, untuk MPS Sunda-Madura didapatkan persamaan :

$$P(md|sd) \approx \max_e P(md|id)P(id|sd) \quad (8)$$

Menerapkan model terjemahan pada korpus bilingual kecil akan menghasilkan kinerja sistem terjemahan yang buruk. Oleh karena itu, penyebab dari kinerja yang buruk adalah data yang kecil. Bertujuan untuk meningkatkan kinerja ini, kita dapat memanfaatkan korpus paralel sumber *pivot* dan *pivot* tambahan. Selain itu, lebih dari satu bahasa *pivot* dapat digunakan untuk memperkaya kualitas kinerja terjemahan. Bahasa *pivot* yang berbeda dapat menangkap fenomena bahasa yang berbeda dan dapat meningkatkan kualitas terjemahan dengan menambahkan pasangan frase sumber-target yang berkualitas.

Jika disertakan bahasa *pivot*  $k$ , model *pivot*  $k$  dapat diperkirakan. Interpolasi linier digunakan untuk menggabungkan semua model yang dihasilkan ini dengan model standar yang dilatih dengan sumber target korpus.

### 2.3. Data dan Metode Pengujian

Korpus merupakan sekumpulan data yang berupa kalimat-kalimat dalam bahasa tertentu. Eksperimen yang dilakukan menggunakan korpus paralel bahasa Indonesia-Madura sebesar 5K kalimat dan bahasa Indonesia-Sunda sebesar 6K kalimat. Korpus monolingual bahasa Melayu, Sunda dan Indonesia masing-masing sebesar 10K, 10K, dan 100K kalimat. Korpus monolingual untuk bahasa Indonesia jauh lebih mudah dikumpulkan daripada dua bahasa lainnya. Tabel 1 menyajikan statistik korpus, yang telah digunakan dalam penelitian berulang kali, termasuk sumber dan informasi bahasa target di setiap arah.

Data pengujian diambil dari luar korpus paralel yang disediakan berupa 100 kalimat bahasa Sunda yang diterjemahkan ke bahasa Madura secara manual.

Tabel 1. Statistik Korpus

Terjemahan	Id-Md	Id-Sd
Kalimat	5.000	6.000
Kata	Sumber	62.090
	Target	61.748
		72.080

Proses *cleaning*, *tokenisasi*, dan *lowercase* terhadap korpus paralel yang sudah dipersiapkan dilakukan sebelum digunakan lebih lanjut dalam eksperimen. Model Bahasa yang digunakan pada sistem baseline adalah model trigram bahasa Indonesia, Sunda, dan Madura yang dilatih menggunakan toolkit SRILM [18].

Korpus paralel Indonesia-Madura dan Indonesia-Sunda kemudian dilatih dengan GIZA++ [19] untuk memperoleh *word alignments*, *phrase table*, *language model*, dan *model combination weights*. Selanjutnya digunakan MOSES yang merupakan tools untuk implementasi MPS. MOSES [20] digunakan untuk melatih model terjemahan terjemahan dari bahasa sumber ke bahasa target. Adapun untuk mengevaluasi kinerja sistem, penelitian ini menggunakan metode BLEU[21].

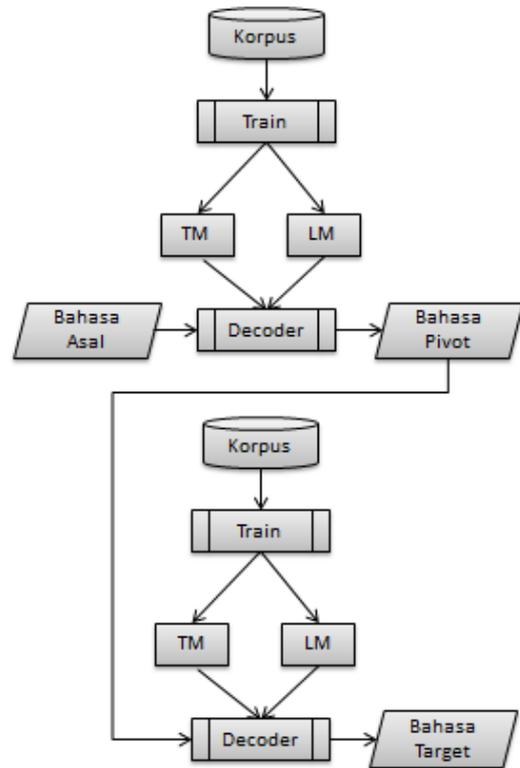
**3. Hasil dan Pembahasan**

Untuk menerapkan metode *Triangulation* pada penelitian ini, sistem MPS Madura-Sunda dengan menggunakan bahasa Indonesia sebagai bahasa *pivot* menggunakan dua tabel frase (Madura-Indonesai dan Indonesia-Sunda). Dalam metode *pivot* frase, berdasarkan kedua tabel tersebut, dibentuk tabel frase Madura-Sunda. Selanjutnya, algoritma pencocokan yang mengidentifikasi pasangan kalimat paralel di antara tabel digunakan. Perlakuan yang sama dilakukan untuk sistem MPS Sunda-Madura dengan menggunakan bahasa Indonesia sebagai bahasa *pivot* menggunakan tabel frase Sunda-Indonesia dan Indonesia-Madura. Arsitektur pengujian dengan metode *Transfer* dan *Triangulation* dapat dilihat pada Gambar 1 dan 2. Dari gambar tersebut terlihat bahwa kedua metode sama-sama menggunakan dua korpus, akan tetapi memperlakukan kedua korpus tersebut dengan cara berbeda.

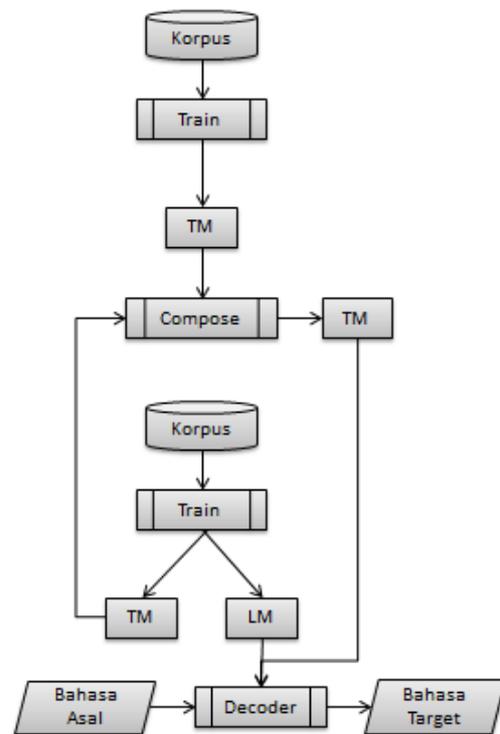
Untuk MPS Madura-Sunda dengan metode *Transfer*, korpus paralel Madura-Indonesia dilatih dengan GIZA++ untuk menghasilkan translation model (TM), sedangkan korpus monolingual Indonesia dilatih dengan menggunakan SRILM untuk menghasilkan language model (LM) bahasa Indonesia. Kedua model tersebut digunakan pada *decoder* (MOSES) untuk menerjemahkan bahasa madura ke bahasa Indonesia. Korpus paralel Indonesia-Sunda dilatih dengan GIZA++ untuk menghasilkan *translation model* (TM), sedangkan korpus monolingual Sunda dilatih dengan menggunakan SRILM untuk *menghasilkan language model* (LM) bahasa Sunda. Kedua model tersebut digunakan pada *decoder* (MOSES) untuk menerjemahkan bahasa Indonesia luaran dari *decoder* Madura Indonesia ke bahasa Sunda.

Untuk MPS Sunda-Madura dengan metode *Transfer*, korpus paralel Sunda-Indonesia dilatih dengan GIZA++ untuk menghasilkan translation model (TM), sedangkan korpus monolingual Indonesia dilatih dengan menggunakan SRILM untuk menghasilkan LM bahasa Indonesia. TM dan LM tersebut digunakan pada *decoder* (MOSES) untuk menerjemahkan bahasa Sunda ke bahasa Indonesia. Korpus paralel Indonesia-Madura dilatih dengan GIZA++ untuk menghasilkan TM, sedangkan korpus monolingual Madura dilatih dengan menggunakan SRILM untuk menghasilkan language

model (LM) bahasa Madura. Kedua model tersebut digunakan pada *decoder* (MOSES) untuk menerjemahkan bahasa Indonesia luaran dari *decoder* Sunda Indonesia ke bahasa Madura.



Gambar 1. Arsitektur Metode Transfer



Gambar 2. Arsitektur Metode Triangulation

Pembuatan MPS Madura-Sunda dengan metode *Triangulation* menggunakan korpus paralel Madura-Indonesia dan Indonesia-Sunda yang dilatih dengan GIZA++ untuk menghasilkan *translation model* (TM) masing-masing pasangan bahasa. Kedua TM tersebut (Madura-Indonesia dan Indonesia-Sunda) disusun kembali dengan melihat frase yang sama pada bahasa Indonesia sebagai bahasa *pivot* dimana nilai probabilitasnya dihitung dengan menggunakan persamaan (7). Hasilnya berupa TM Madura-Sunda yang akan dipakai oleh *decoder* (MOSES), sedangkan LM modelnya menggunakan monolingual bahasa Sunda yang dilatih dengan SRILM.

Pembuatan MPS Sunda-Madura dengan metode *Triangulation* menggunakan korpus paralel Sunda-Indonesia dan Indonesia-Madura yang dilatih dengan GIZA++ untuk menghasilkan *translation model* (TM) masing-masing pasangan bahasa. Kedua TM tersebut (Sunda-Indonesia dan Indonesia-Madura) disusun kembali dengan melihat frase yang sama pada bahasa Indonesia sebagai bahasa *pivot* dimana nilai probabilitasnya dihitung dengan menggunakan persamaan (8). Hasilnya berupa TM Sunda-Madura yang akan dipakai oleh *decoder* (MOSES), sedangkan LM modelnya menggunakan monolingual bahasa Madura yang dilatih dengan SRILM.

Tabel 2 mengilustrasikan hasil dari sistem terjemahan dengan metode *Transfer* dan *Triangulation* MPS Madura-Sunda dua arah melalui Bahasa Indonesia sebagai bahasa perantara. Hasil ini didapat dengan menggunakan LM yang dilatih murni dari bahasa target tanpa tambahan kalimat serta dari pengujian dengan menambahkan korpus monolingual pada bahasa target masing-masing sebesar 10K untuk bahasa Madura dan Sunda serta 100K untuk bahasa Indonesia.

Tabel 2. Nilai BLEU MPS dengan Bahasa *Pivot*

Metode	LM murni (%)		LM Tambahan (%)	
	Md-Sd	Sd-Md	Md-Sd	Sd-Md
<i>Transfer</i>	19,44	18,02	23,07	19,90
<i>Triangulation</i>	20,81	19,67	21,41	20,97
Peningkatan	7,05	9,16	5,31	5,38

Untuk data set pelatihan LM murni, nilai BLEU dari sistem terjemahan dengan metode *Transfer* adalah 19,44% untuk MPS Madura-Sunda dan 18,02% untuk MPS Sunda-Madura. Sedangkan pada pengujian yang menggunakan monolingual tambahan, nilai BLEU dari sistem terjemahan dengan metode *Transfer* adalah 23,07% untuk MPS Madura-Sunda dan 19,90% untuk MPS Sunda-Madura. Pada data set pelatihan LM murni, nilai BLEU dari sistem terjemahan dengan metode *Triangulation* adalah 20,81% untuk MPS Madura-Sunda dan 19,67% untuk MPS Sunda-Madura. Sedangkan pada pengujian yang menggunakan monolingual tambahan, nilai BLEU dari sistem terjemahan dengan metode *Transfer* adalah 21,41%

untuk MPS Madura-Sunda dan 20,97% untuk MPS Sunda-Madura.

Jika dibandingkan dengan metode *Transfer*, MPS Madura-Sunda dengan monolingual murni dapat meningkatkan akurasi MPS sebesar 7,05%, sedangkan dengan monolingual tambahan dapat meningkatkan akurasi MPS sebesar 5,31%. Berarti metode *Triangulation* rata-rata meningkatkan akurasi MPS berbasis *pivot* bahasa Indonesia sebesar 6,18% untuk MPS Madura-Sunda. Demikian juga untuk MPS Sunda-Madura, metode *Triangulation* rata-rata meningkatkan akurasi MPS berbasis *pivot* menggunakan bahasa Indonesia sebesar 7,27%. Ini menunjukkan bahwa, penggunaan metode *Triangulation* dapat memanfaatkan sumber-sumber yang tersedia dengan lebih baik.

Kelemahan dari bahasa *pivot* adalah bahwa setiap langkah retranslasi menghasilkan kemungkinan kesalahan dan ambiguitas, karena dengan menggunakan bahasa *pivot* berarti melibatkan dua langkah terjemahan. Jika dilihat dari arsitektur metode *Transfer* (Gambar 1), jelas terlihat bahwa suatu kalimat dari bahasa asal diterjemahkan sebanyak dua kali, yaitu ke bahasa *pivot*, baru kemudian hasilnya diterjemahkan lagi ke bahasa target. Metode *Triangulation* menghasilkan terjemahan yang lebih baik karena hanya menggunakan satu langkah terjemahan. Seperti yang terlihat pada arsitektur (Gambar 2), yang digabung adalah model bahasanya, kemudian model ini yang digunakan pada *decoder* mesin penerjemah Madura-Sunda dan sebaliknya.

#### 4. Kesimpulan

Dalam artikel ini, dibandingkan dua metode terjemahan bahasa *pivot* umum yang terdiri dari *pivot* level frase untuk MPS bahasa berdaya sumber rendah Madura-Sunda dua arah dengan menggunakan bahasa Indonesia sebagai bahasa bahasa penghubung. Temuan-temuan dari percobaan yang dilakukan mengungkapkan bahwa menggunakan bahasa Indonesia sebagai bahasa penghubung dengan menggunakan metode *Triangulation* memiliki akurasi yang lebih baik, jika dibandingkan dengan metode *Transfer*. Dari hasil percobaan yang dilakukan, metode *Triangulation* rata-rata meningkatkan akurasi MPS berbasis *pivot* bahasa Indonesia sebesar 6,18% untuk MPS Madura-Sunda dan 7,27% untuk MPS Sunda-Madura.

Hasil penelitian ini dapat dijadikan sebagai acuan untuk penelitian-penelitian berikutnya dalam memilih metode mesin penerjemah statistik dengan pendekatan bahasa *pivot*. Pendekatan ini sangat berguna untuk membangun mesin penerjemah antar bahasa daerah di Indonesia yang memang memiliki sumber daya sangat rendah. Pada penelitian-penelitian selanjutnya, justifikasi hasil penelitian ini akan menjadi semakin kuat jika dilakukan dengan data berupa korpus yang

lebih besar.

### Daftar Rujukan

- [1] Babych, B., Hartley, H., Sharoff, S., Mudraya, O., 2007. Assisting translators in indirect lexical transfer. *Proceedings of ACL 2007: the 45th Annual Meeting of the Association of Computational Linguistics*.
- [2] Koehn, P., 2005. Europarl: A parallel corpus for statistical machine translation. *Proceedings of AAMT: the 10th Machine Translation Summit*. Phuket, Thailand, pages 79–86.
- [3] Kholy, A.E., Habash, N., Leusch, G., Matusov, E., Sawaf, H., 2013. Language Independent Connectivity Strength Features for Phrase Pivot Statistical Machine Translation. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Sofia, Bulgaria, Agustus 2013.
- [4] Matusov, E., Leusch, G., Banchs, R. E., Bertoldi, N., Dechelotte, D., Federico, M., Ney, H. 2008. System Combination for Machine Translation of Spoken and Written Language. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(7), 1222-1237.
- [5] Utiyama, M., Isahara, H., 2007. A comparison of pivot methods for phrase-based statistical machinetranslation. *Proceedings of HLT-NAACL 2007, the Human Language Technology Conference of the North American Chapters of the Association of Computational Linguistics*. pages 484–491.
- [6] Wu, H., Wang, H., 2007. Pivot language approach for phrase-based statistical machine translation. *Proceedings of ACL: the 45th Annual Meeting of the Association of Computational Linguistics*. pages 856–863.
- [7] Babych, B., Hartley, A., Sharoff, S., 2008. Translating from under-resourced languages: Comparing direct transfer against pivot translation. *Proceedings of MT Summit XI*.
- [8] Habash, N., Hu, J. 2009. Improving Arabic-Chinese statistical machine translation using English as pivot language. *Proceedings of the Fourth Workshop on Statistical Machine Translation - StatMT 09*.
- [9] Somaye, B., Khadivi, S., Riahi, N., 2010. Farsi-german statistical machine translation through bridge language. *Proceedings of IST, 5th International Symposium on Telecommunications*.
- [10] Al-Hunaity, M., Maegaard, B., Hansen, D. H., 2010. Using English as a Pivot Language to Enhance Danish-Arabic Statistical Machine Translation. In K. Choukri, O. Rambow, B. Maegaard, & I. Alkharashi (Eds.), *Proceedings of the workshop in LR and HLT for Semitic Languages* Valletta, Malta: European Language Resources Association.
- [11] Nakov, P., Ng, H.T., 2012. Improving statistical machine translation for a resource-poor language using related resource-rich languages. *JAIR: Journal of Artificial Intelligence Research* pages 179–222.
- [12] Paul, M., Finch, A., Sumita, E., 2013. How to Choose the Best Pivot Language for Automatic Translation of Low-Resource Languages. *ACM Transactions on Asian Language Information Processing*, 12(4), 1-17.
- [13] Kunchukuttan, A., Pudupully, R., Chatterjee, R., Mishra, A., Tacharyya, P.B., 2014. The iit bombay smt system for icon2014 tools contest. *Proceedings of ICON 2014, the Natural Language Processing Tools Contest*.
- [14] Dabre, R., Cromieres, F., Kurohashi, S., & Bhattacharyya, P., 2015. Leveraging Small Multilingual Corpora for SMT Using Many Pivot Languages. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- [15] Suryani, A. A., Ariesanti, I., Yohanes, B. W., Subair, M., Budiwati, S. D., Rintyarna, B. S., 2016. Enriching English into Sundanese and Javanese translation list using pivot language. *2016 International Conference on Information & Communication Technology and Systems (ICTS)*.
- [16] Ningtyas, D. W., Sujaini, H., & Safriadi, N., 2018. Penggunaan Pivot Language pada Mesin Penerjemah Statistik Bahasa Inggris ke Bahasa Melayu Sambas. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 173.
- [17] Berger, A.L., Pietra, S.D., Vincent, J., Pietra, D., 1996. Amaximumentropyapproach to natural language processing. *Computational Linguistics* 22(1):39–71.
- [18] Stolcke, A., Zheng, J., Wang, W., Abrash, V., 2011. SRILM at sixteen: Update and outlook, Automatic Speech Recognition and Understanding (ASRU), *2011 IEEE Workshop*, Waikoloa.
- [19] Och, F. J., Ney, H., 2003. A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, 1(29), pp. 19-51, 2003.
- [20] Koehn, P., 2010. Statistical machine translation, New York: *Cambridge University Press*.
- [21] Papineni, K., Roukos, S., Ward, T., Zhu, W.-J., 2002. BLEU: A Method For Automatic Evaluation of Machine Translation, *Proceedings of the 40th Annual Meeting of the Association of Computational Linguistics (ACL)*, Pennsylvania.