

統計的手法を用いた英日パターン翻訳

江木孝史 *1 村上仁一 *2 村田真樹 *2 徳久雅人 *2

*1 鳥取大学 工学部 知能情報工学科

*2 鳥取大学大学院 工学研究科 情報エレクトロニクス専攻

*1,*2 {s082008, murakami, murata, tokuhisa} @ ike.tottori-u.ac.jp

1 はじめに

パターン翻訳は、人手で作成した大量の文パターンを用意し翻訳文を出力する方法である。通常文パターン辞書の作成は人手で行うため開発に時間がかかる。しかしパターン翻訳の長所として、入力文が文パターンに適合した場合は翻訳精度の高い文が得られる。

本実験では人手によらず、文パターンをプログラムで自動作成する。そしてパターンマッチングを用いて自動的に英日方向の統計翻訳を行う。実験の結果から、用いた tri-gram のスコアが高い翻訳文において、翻訳精度の高い文が得られた。

尚、過去に西村らはパターン翻訳と句に基づく統計翻訳を組み合わせた手法(以降、先行手法と述べる)で日英翻訳を行い翻訳精度を求めている [1]。そこで考察として先行手法と比較した際の利点、欠点について報告する。また句に基づく統計翻訳との精度比較を行い結果を報告する。

2 英日パターン翻訳

一般的な英日パターン翻訳の手順を以下に示す。

手順1 英日文パターン辞書と英日対訳単語辞書を用意する。

手順2 入力文と英語文パターンを照合する。

手順3 対応する日本語文パターンと英日対訳単語辞書を用いて翻訳文を生成する。

翻訳例と単語辞書の一部を表1, 表2に示す。

表1 英日パターン翻訳の例

入力文	The fire started in the kitchen .
英語文パターン	The X1 started in the X2 .
日本語文パターン	X1 は X2 から 出た。
翻訳文	火は 台所 から 出た。

表2 英日対訳単語辞書の例

fire	火
kitchen	台所

3 GIZA++

GIZA++ とは英日方向、日英方向の対訳文について最尤な単語の対応を得るツールである [2]。対訳文を用いて IBM モデル 1~5 を学習し単語レベルの確率を得る。本研究では英日対訳単語辞書の作成に使用する。

表3に例を示す。

表3 GIZA++ を用いた英日方向の単語辞書の例

fire	火	0.09
kitchen	台所	0.44

4 データベース

実験には電子辞書から抽出した単文データベースと重文・複文データベースを用いる [3]。

4.1 データベースの注意点

- 実験に使用した単文データは日本語文が単文であるため、英語文に重文・複文が含まれる場合がある。
- 実験に使用した重文・複文データは日本語文が重文・複文であるため、英語文に単文が含まれる場合がある。

5 提案手法

GIZA++ を用いて英日文パターンをプログラムで自動作成する。

以下に手順を示す。

5.1 英日対訳単語辞書の作成手順

手順1 GIZA++ を用いて英日方向、日英方向の単語対応を調べ、英日単語辞書、日英単語辞書を作成す

る。作成した英日単語辞書は表3に示すように左から英単語、対応する日本語単語、英単語が日本語単語に翻訳される確率が与えられている。

手順2 手順1で作成した英日単語辞書と日英単語辞書の翻訳確率を掛け合わせ一定の閾値 (α) 以上の確率を持つ単語を英日対訳単語辞書として登録する。

5.2 英日文パターン辞書の作成手順

英日対訳学習文と英日対訳単語辞書を用いて英日文パターン辞書を作成する。

以下に英日文パターン辞書作成の手順を示す。

手順1 英語学習文の各単語と英日対訳単語辞書の単語を照合する。一致したら英単語に対応する日本語単語と日本語学習文の単語を照合する。

手順2 日本語文の照合に成功したら一致した英日対訳学習文の単語を変数に置き換える。

図1に英日文パターン辞書作成の手順を示す。

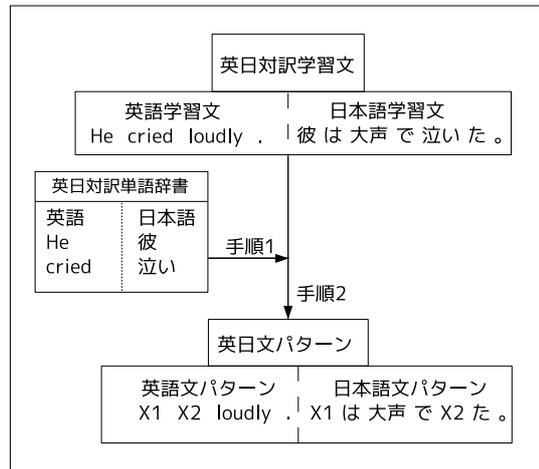


図1 英日文パターン辞書の作成手順

5.3 翻訳文生成の手順

翻訳文生成の手順を以下に示す。

手順1 入力文を読み込み、入力文に一致する英語文パターンを選択する。

手順2 一致する英語文パターンが存在する場合、変数部に対応する英単語を英日対訳単語辞書を用いて日本語単語に翻訳する。

手順3 英語文パターンに対応する日本語文パターンの変数部を英日対訳単語辞書を用いて日本語単語に置き換える。

手順4 複数の翻訳候補が出力された場合、tri-gram モデルを用いて翻訳候補の絞込みを行う。tri-gram の学習データは英日対訳学習文の日本語学習文を用いる。候補文に対し tri-gram のスコアが存在しない場合はペナルティーとして -1,000 を付与する。

手順5 tri-gram スコアが最大となる翻訳候補文を翻訳文として出力する。

本実験では翻訳結果を分類するために翻訳文の tri-gram スコアに閾値を設ける。tri-gram 学習データと閾値の詳細は 6.3 節で述べる。

図2に翻訳文の生成手順を示す。

表 4 各ランクの自動評価結果

	BLEU	NIST
A ランク (379 文)	0.5664	6.9185
B ランク (247 文)	0.2993	4.4478
C ランク (292 文)	0.2228	3.9444
D ランク (2,334 文)	0.0686	3.0176
A~D ランク (3,252 文)	0.1683	3.8956

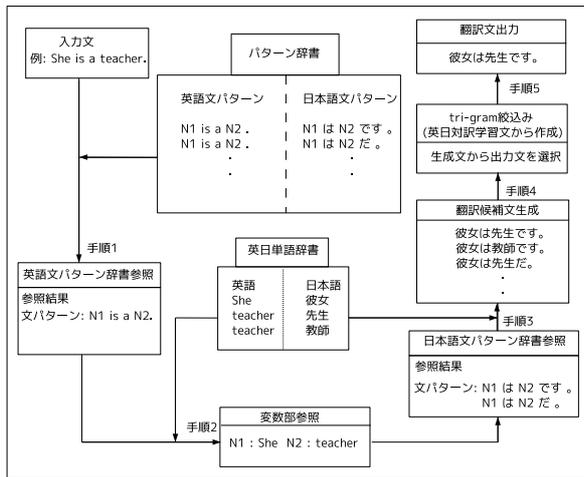


図 2 翻訳文の生成手順

6 実験条件

本実験では単文データベースから入力テスト文 10,000 文と英日対訳学習文 100,000 文を用いる。同様に重文・複文データベースから入力テスト文 10,000 文と英日対訳学習文 100,000 文を用いる。

6.1 単文データ

単文テストデータ 10,000 に対し英日パターン翻訳を行う。

英日パターン辞書の作成には 4.1 節の手順 2 で示す $\alpha=0.05$ (17,128 語登録) を用いる。英日対訳学習文 100,000 文に対し 86,204 文の英日パターンを得た。また、パターン翻訳には $\alpha=0.005$ (76,202 語登録) の英日対訳単語辞書を用いる。

6.2 重文・複文データ

重文・複文テストデータ 10,000 に対し英日パターン翻訳を行う。

英日パターン辞書の作成には $\alpha=0.05$ (16,110 語登録) を用いる。英日対訳学習文 100,000 文に対し 87,674 文のパターン英日パターンを得た。また、ターン翻訳には $\alpha=0.01$ (49,704 語登録) の英日対訳単語辞書を用いる。

6.3 tri-gram 学習データ

tri-gram 学習データには英日対訳学習文 100,000 文を使用する。翻訳文の tri-gram スコアに閾値を設け、-1,000 以上 0 以下の場合 A ランク、-2,000 以上 -1,001 以下の場合 B ランク、-3,000 以上 -2,001 以下の場合 C ランク、-3,001 以下の場合 D ランクとする。また、以下の場合は翻訳結果を出力しない。

- 入力文が英語文パターンに一致しない
- 5.3 節の手順 2 において変数部に対応する単語が英日対訳単語辞書に登録されていない

上記の条件を満たす場合は翻訳結果を出力しない。

6.4 ベースラインシステム

ベースラインシステムには Moses を用いる [4]。本実験ではパラメータチューニングは行っていない。また、形態素解析には Mecab を使用する [5]。

7 実験結果

7.1 単文テストデータの自動評価結果

7.1.1 英日パターンに照合した文のみの評価

入力文 10,000 文に対し、3,252 文の翻訳文を得た。翻訳文 3,252 文に対して A ランクは 379 文、B ランクは 247 文、C ランクは 292 文、D ランクは 2,334 文であった。各ランクの BLEU と NIST の評価結果を表 4 に示す。

7.1.2 A ランクとベースラインの比較

単文において、パターン翻訳を用いて翻訳された A ランク 379 文と、ベースラインシステムで翻訳された 379 文の BLEU 値の比較結果を表 5 に示す。

表 5 A ランクの翻訳文とベースラインの比較

	BLEU	NIST
ベースライン	0.5434	6.7248
提案手法	0.5664	6.9185

表 5 の結果から、提案手法はベースラインと比較して BLEU 値が 2.3% 向上している。B~D ランクはベースラインと比較して BLEU 値の低下が認められたが、単文におけるパターン翻訳は A ランクの翻訳文に関して高い翻訳精度が得られた。

7.1.3 10000 文での自動評価結果

提案手法は表 4 の A ランクの翻訳文 (379 文) と、ベースラインから得た A ランクを除く 9,621 文の翻訳文を合わせた 10,000 文を評価に用いる。そしてベースラインと比較する。

表 6 ベースラインと提案手法の比較

	BLEU	NIST
ベースライン	0.1375	3.7743
提案手法	0.1381	3.7798

表 6 において、ベースラインと比較して提案手法は BLEU 値が 0.06% 向上した。

7.2 単文テストデータの人手評価結果

表 4 の A~D ランクに対し、ランダムに抽出した 50 文を用いて人手による比較を行う。評価基準を以下に示す。

- 提案手法○：提案手法がベースラインよりも良い
- 提案手法×：提案手法がベースラインに劣っている
- 差なし：翻訳品質に差がない
- 同一出力：提案手法、ベースラインの出力が同じ

以下に単文における人手評価の結果を示す。

表 7 人手評価の結果

ランク	提案手法○	提案手法×	差なし	同一出力
A ランク	9/50	0/50	19/50	22/50
B ランク	6/50	11/50	24/50	9/50
C ランク	7/50	9/50	33/50	1/50
D ランク	1/50	10/50	39/50	0/50

表 7 より、A ランクは提案手法○が多く、提案手法の有効性が認められた。B~D ランクは提案手法×の割合が多く、提案手法の有効性は認められなかった。

7.2.1 翻訳例

以下に A~D の提案手法○と提案手法×の翻訳例を示す。

A ランクの翻訳例

表 8 A ランクにおける提案手法○の翻訳例

入力文	A pendulum swings back and forth .
英語文パターン	A X1 swings back and forth .
日本語文パターン	X1 は 左右 に 振動 する。
正解文	振り子は 左右 に 振動 する。
ベースライン	振り子は 左右 に して いる。
提案手法	振り子は 左右 に 振動 する。

表 8 において提案手法は正解文と同じ翻訳文が得られている。しかしベースラインは"swings"が翻訳されておらず、文として不適切である。

B ランクの翻訳例

表 9 B ランクにおける提案手法○の翻訳例

入力文	She called him all the bad names .
英語文パターン	X1 called X2 all the bad names .
日本語文パターン	X1 は X2 をくそみそに言った。
正解文	彼女は さんざん 彼の悪口を言った。
ベースライン	彼女は 彼を をけなした。
提案手法	彼女は 彼をくそみそに言った。

表 9 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは”を”が連続して翻訳されており、文として不適切である。

表 10 B ランクにおける提案手法×の翻訳例

入力文	The telephone is out of order .
英語文パターン	X1 X2 X3 out of order .
日本語文パターン	X1 X2 は 狂いがきて X3 .
正解文	電話がこわれている。
ベースライン	その電話は故障している。
提案手法	その電話は狂いがきている。

表 10 においてベースラインは正解文と同等の翻訳文が得られている。しかし提案手法は日本語文として意味が通じない不適切な翻訳がされている。

C ランクの翻訳例

表 11 C ランクにおける提案手法○の翻訳例

入力文	His pistol went off accidentally .
英語文パターン	His X1 went off accidentally .
日本語文パターン	その男の X1 が暴発した。
正解文	彼のピストルが暴発した。
ベースライン	彼のピストルは誤って 出発した。
提案手法	その男のピストルが暴発した。

表 11 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは”went off”が”出発”と翻訳されており、文として不適切である。

表 12 C ランクにおける提案手法×の翻訳例

入力文	He is bent with age .
英語文パターン	X1 X2 X3 with X4 .
日本語文パターン	X1 は X4 に X3 てい X2 .
正解文	彼は年で腰が曲がっている。
ベースライン	彼は年のせいで腰が 曲がっている。
提案手法	彼は時代に曲がっていて。

表 12 においてベースラインは正解文と同等の翻訳文が得られている。しかし提案手法は”age”が”時代”として翻訳されており、日本語文として意味が通じない。

D ランクの翻訳例

表 13 D ランクにおける提案手法○の翻訳例

入力文	The streets are wet after the rain .
英語文パターン	The X1 X2 X3 X4 the X5 .
日本語文パターン	X1 は X5 X4 X3 て X2 .
正解文	雨のあとで街路がぬれている。
ベースライン	通りにはを潤したを 提供してくれた。
提案手法	通りは大雨後ぬれている。

表 13 において提案手法は日本語文として意味が通じる翻訳文が得られている。しかしベースラインは助詞が連続して翻訳されており、文として不適切である。

表 14 D ランクにおける提案手法×の翻訳例

入力文	Tell me the exact time .
英語文パターン	X1 X2 the X3 X4 .
日本語文パターン	X1 が X3 X4 を X2 た。
正解文	正確な時刻を教えてください。
ベースライン	正確な時刻を教えてください。
提案手法	話しが掴め時間をくれた。

表 14 においてベースラインは正解文と同じ翻訳文が得られている。しかし提案手法は不適切な翻訳がされている。

7.3 単文データのまとめ

自動評価結果から BLEU 値、NIST 値共に向上し、提案手法の有効性が示された。

人手評価結果から、A ランクの翻訳文はベースラインよりも優れている文が多いと言える。一方で B~D ランクの翻訳文はベースラインに劣っている文が多く、A ランクと B~D ランクでは翻訳精度に差がある。A ランクにおいて、人手評価からも提案手法の有効性が示された。

7.4 重文・複文テストデータの自動評価結果

7.4.1 パターンに照合した文のみの評価

入力文 10,000 文に対し、823 文の翻訳文を得た。翻訳文 823 文に対して A ランクは 408 文、B ランクは 31 文、C ランクは 16 文、D ランクは 368 文であった。評価結果を表 15 に示す。

表 15 各ランクの翻訳文における自動評価結果

	BLEU	NIST
A(408 文)	0.5662	7.7735
B(31 文)	0.4717	4.8247
C(16 文)	0.3517	3.6805
D(368 文)	0.0710	2.3499
A~D(823 文)	0.3630	5.6752

7.4.2 A ランクとベースラインの比較

重文・複文において、パターン翻訳を用いて翻訳された A ランク 408 文と、ベースラインシステムで翻訳された 408 文の BLEU 値の比較結果を表 16 に示す。

表 16 A ランクの翻訳文とベースラインの比較

	BLEU	NIST
ベースライン	0.5348	7.4597
提案手法	0.5662	7.7735

表 16 の結果から、提案手法はベースラインと比較して BLEU 値で 3.1% 向上している。また、B、C ランクも BLEU 値の向上が認められた。D ランクはベースラインと比較して BLEU 値の低下が認められたが、重文・複文におけるパターン翻訳は A~C ランクの翻訳文に関して高い翻訳精度が得られた。

7.4.3 10000 文での自動評価結果

提案手法は表 15 の A ランクの翻訳文 (408 文) と、ベースラインから得た A ランクを除く 9,592 文の翻訳文を合わせた 10,000 文を評価に用いる。そしてベースラインと比較する。

表 17 ベースラインと提案手法の比較

	BLEU	NIST
ベースライン	0.0987	3.1100
提案手法	0.0999	3.1267

表 17 において、ベースラインと比較して提案手法は BLEU 値が 0.12% 向上した。

7.5 重文・複文テストデータの手手評価結果

表 15 の翻訳文に対し、7.2 節の評価基準を用いて人手評価を行う。尚、重文・複文は単文に比べ英日文パターンに照合した文数が少ないためランダムに抽出した 10 文を用いて人手による比較を行う。

以下に重文・複文における人手評価の結果を示す。

表 18 人手評価結果

ランク	提案手法○	提案手法×	差なし	同一出力
A ランク	3/10	0/10	0/10	7/10
B ランク	4/10	0/10	5/10	1/10
C ランク	2/10	2/10	5/10	1/10
D ランク	0/10	2/10	8/10	0/10

表 18 より、A~B ランクは提案手法○が多く、提案手法の有効性が認められた。C~D ランクは提案手法×の割合が多く、提案手法の有効性は認められなかった。

7.5.1 翻訳例

以下に提案手法○と提案手法×の翻訳例を示す。

表 19 A ランクにおける提案手法○の翻訳例

入力文	The waves break on the rocks .
英語文パターン	The X1 break on the X2 .
日本語文パターン	X1 が X2 に 当たって 砕ける。
正解文	波が岩に当たって砕ける。
ベースライン	波が岩に当たった。
提案手法	波が岩に当たって砕ける。

表 19 において提案手法は正解文と同じの意味を持つ翻訳結果が得られている。一方ベースラインの翻訳は意味として通じるが、“砕ける”が翻訳されておらず、提案手法の翻訳文に比べて劣っているとと言える。

表 20 B ランクにおける提案手法○の翻訳例

入力文	A man was struck dead by lightning .
英語文パターン	A X1 was struck dead by lightning .
日本語文パターン	一人の X1 が 雷に 打たれて 死んだ。
正解文	人が雷に打たれて死んだ。
ベースライン	一人の男がは雷に打たれて死んだ。
提案手法	一人の男が雷に打たれて死んだ。

表 20 において提案手法は正解文と同等の意味を持つ出力が得られている。しかしベースラインの出力文は助詞が連続して翻訳されており、文として不適切である。

表 21 C ランクにおける提案手法○の翻訳例

入力文	He demonstrated that the world is round .
英語文パターン	X1 demonstrated that X2 X3 X4 round .
日本語文パターン	X1 X4 X3 が 丸い ということ X2 証明した。
正解文	彼は地球は丸い ということを 実証した。
ベースライン	世界は である ことが わかった のではないかと 心配している。
提案手法	彼は世界が丸い ということを 証明した。

表 21 において提案手法は正解文と同等の出力が得られている。しかしベースラインは文として意味が通じず不適切である。

表 22 C ランクにおける提案手法×の翻訳例

入力文	It is ten minutes ' walk from here to the station .
英語文パターン	It is X1 X2 ' walk X3 X4 to the X5 .
日本語文パターン	X4 X3 X5 まであるいて X1 X2 です。
正解文	ここから 駅まで 歩いて 10 分 です。
ベースライン	ここから 駅までは 歩いて 10 分 です。
提案手法	ここから 駅まであるいて 0 分 です。

表 22 においてベースラインは正解文と同等の出力が得られている。しかし提案手法は数字が適切に翻訳されず、文として意味が通じない。

表 23 D ランクにおける提案手法×の翻訳例

入力文	She is a good writer .
英語文パターン	X1 X2 a X3 X4 .
日本語文パターン	X1 X2 異国 ふう に X3 X4 だ。
正解文	彼女は 文章 を 書く のが 上手 だ。
ベースライン	彼女は 立派 な 作家 だ。
提案手法	彼女は 異国 ふう に 上手 作家 だ。

表 23 においてベースラインの翻訳文は正解文と異なるが日本語として自然な翻訳である。しかし提案手法は文として意味が通じず日本語文として不適切である。

7.6 重文・複文データのまとめ

自動評価結果から BLEU 値、NIST 値共に向上し、提案手法の有効性が示された。

人手評価結果から、A、B ランクの翻訳文はベースラインよりも優れている文が多いと言える。一方で C ランクは提案手法○と×が同数、D ランクの翻訳文はベースラインの翻訳に劣っている文が多く、A、B ランクと C、

D ランクでは翻訳精度に差がある。A、B ランクに関して人手評価からも提案手法の有効性が示された。

8 考察

8.1 パターン翻訳の翻訳精度

8.1.1 単文における翻訳精度の考察

単文において、A ランクの翻訳文の精度が向上した理由として、A ランクの翻訳文は 5.3 節の手順 4 で示すペナルティーが付与されなかった。よって、A ランクは高い翻訳精度が得られたと考えられる。

8.1.2 重文・複文における翻訳精度の考察

重文・複文において、A~C ランクの翻訳精度が向上した理由として、単文に比べて英日文パターン辞書と英日対訳単語辞書に用いた単語数が少なかったことが挙げられる。特に英日文パターン辞書の作成段階で、文パターン中に変数化された箇所が少なく、文字情報が多く残されていた可能性がある。よって入力文が英語文パターンと一致した際に正解文と同等の翻訳文が多く得られたと考えられる。

8.2 日英方向におけるパターン翻訳との比較

過去に西村らは 1 節の先行手法を用いて日英翻訳実験を行った。単文テストデータ 10,000 文に対し、閾値 0.01(53,185 語登録)の日英対訳単語辞書を用いた日英文パターン辞書と閾値 0.02(34,587 語登録)の日英対訳単語辞書を用いて先行手法を行い、2,103 文が日本語文パターンに照合した。

文パターンに照合した文数を比較すると、先行手法が 2,103 文であるのに対し本実験では 3252 文と英日翻訳の方が 1,000 文以上多く文パターンに照合している。先行手法と本実験に使用した文パターン辞書と対訳単語辞書の閾値および登録単語数が同じでないため単純な比較はできないが、日英翻訳に比べ英日翻訳の方がパターンに照合し易いと考えられる。

この理由として英語と日本語間での表現方法の違いが挙げられる。例えば英文 “I climb the mountain .” は単語を入れ替えて “I the climb mountain .” と表現すると文章として成立しなくなる。しかし日本語文は “私は山に登る。” の単語を入れ替えて “山に私は登る。” と表現しても文章として成立する。このように日本語文は語順が自由であり、複数の表現が生成されるため文パターンに照合しにくい。一方で英文は表現方法が限られているため英語文パターンに照合し易かったと思われる。

8.3 単語レベルのデコーダの検討

統計翻訳の初期は単語に基づく統計翻訳であった。しかし単語に基づく統計翻訳は句に基づく統計翻訳に比べて翻訳精度が低い。そのため近年は句に基づく統計翻訳が主流になっている。

本実験のパターン翻訳は英日対訳単語辞書と対訳パターンが翻訳モデルに、tri-gram が言語モデルに相当すると考えることができる。よって実験で用いたパターン翻訳器は単語に基づく統計翻訳のデコーダとして考えることができる。

9 おわりに

本研究では、文パターンの自動作成と tri-gram による翻訳候補の絞込みを組み合わせた手法から、翻訳文を生成した。実験の結果、パターン翻訳と tri-gram を組み合わせることにより、条件付きながら高い精度の翻訳文が得られることが示された。tri-gram の学習データはモノリナルであるため、今後は参照データに新聞記事コーパスを与え更なる翻訳精度の向上を目指す。

参考文献

- [1] 西村拓哉, 村上仁一, 徳久雅人, 池原悟, “文パターンを用いた統計翻訳”, 言語処理学会第 16 回年次大会, 2010.
- [2] GIZA++, Training of statistical translation models <http://www.fjoch.com/GIZA++>
- [3] 村上仁一, 徳久雅人, “日英対訳データベース作成のための 1 考察”, 言語処理学会第 17 回年次大会, 2011.
- [4] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, “Moses: Open Source Toolkit for Statistical Machine Translation”, Proceedings of the ACL 2007 Demo and Poster Sessions, pages 177180, June 2007.
- [5] MeCab, 日本語形態素解析器 <http://mecab.sourceforge.net/>
- [6] SRILM, The SRI Language Model Toolkit <http://www.speech.sri.com/projects/srilm/>
- [7] NIST, Automatic Evaluation of Machine Translation Quality Using n-gram Co-Occurrence Statistics <http://www.itl.nist.gov/iad/mig/test/mt>
- [8] BLEU, NIST Open MT Scoring <http://www.itl.nist.gov/iad/894.01/tests/mt/>