

# 階層的フレーズベース翻訳におけるピボット翻訳手法の応用

三浦 明波<sup>1,a)</sup> Graham Neubig<sup>1,b)</sup> Sakriani Sakti<sup>1,c)</sup> 戸田 智基<sup>1,d)</sup> 中村 哲<sup>1,e)</sup>

概要：統計的機械翻訳において、特定の言語対で十分な文量の対訳コーパスが得られない場合、中間言語を用いたピボット翻訳が有効な解法の一つである。フレーズベース翻訳のためのピボット翻訳手法が複数考案されている一方、英語と日本語のような語順が大きく異なる言語間では文の構造を考慮することも重要である。本稿では、フレーズベース翻訳で有用性の知られているピボット翻訳手法を、階層構造を表現できる階層的フレーズ翻訳に応用し、階層構造に基づく機械翻訳におけるピボット翻訳の影響を示す。

## 1. はじめに

統計的機械翻訳 (Statistical Machine Translation: SMT[1]) では、学習に使用する対訳コーパスが大規模になるほど、高精度な訳出結果が得られることが知られている [2]。一方、英語を含まない言語対などを考慮すれば、多くの言語対について、大規模な対訳コーパスを直ちに用意することは困難である。このような、容易に対訳コーパスを取得できない言語対においても、既存の言語資源を有効に用いて機械翻訳を実現できれば、真の多言語翻訳が可能となり、国際社会への貢献が期待できる。

特定の言語対で十分な文量の対訳コーパスが得られない場合、中間言語を用いたピボット翻訳が有効な解法の一つである [3][4][5][6]。中間言語を用いる方法も様々であるが、2つの機械翻訳システムが与えられていて、一方の出力言語と他方の入力言語が一致しているならば、それらをパイプライン処理する逐次的ピボット翻訳 (Cascade Translation) 手法が先ず考えられる [3]。より高度なピボット翻訳の手法としては、複数の機械翻訳システムで学習されたモデルデータから、目的の言語対のモデルデータを合成して機械翻訳を行うテーブル合成手法 (Triangulation) もある [7]。これらの手法は特に、統計翻訳の主流であるフレーズベース機械翻訳 (Phrase-Based Machine Translation: PBMT[8]) について数多く提案され、検証されてきた。

しかし、英語と日本語、英語と中国語といった語順の大きく異なる言語間では、句の対応のみでなく文の構造を考慮

することも重要である。PBMT の利点を活かしつつ、文の階層構造を表現できるように拡張した階層的フレーズベース翻訳 (Hierarchical Phrase-Based Machine Translation: Hiero[9]) では語順の異なる言語対において、翻訳の精度向上が期待できる。

本稿では、先行研究によって PBMT で有用性の知られているピボット翻訳手法が、異なる枠組みの SMT でも有効であるかどうかを検証することを目的とする。そこで、Hiero においてピボット翻訳を応用し、階層構造に基づく機械翻訳におけるピボット翻訳の影響を調査する。

比較実験のため、PBMT と Hiero で、国連文書を元にした多言語コーパス [10] の様々な組合せで、逐次的ピボット翻訳と、テーブル合成に基づくピボット翻訳を行い、両者の翻訳結果を評価した。

## 2. 機械翻訳手法

機械翻訳において、原言語文を目的言語文に翻訳する際、語彙選択と並び替えを正確に行う必要がある。これらの問題を解決すべく、様々な翻訳手法が提案されており、本節で代表的な手法について紹介する。

### 2.1 フレーズベース機械翻訳

Koehn らによる PBMT[8] は統計的機械翻訳で最も代表的な手法である。PBMT モデルを学習する際に、原言語文と目的言語文の対訳データから単語アライメントを取り、アライメント結果をもとに複数の単語からなるフレーズを抽出し、各フレーズ対応にスコア付けを行う。ただし、スコア付けは、各フレーズの翻訳確率と単語の翻訳確率、そしてペナルティを考慮して行われる。

PBMT を用いて翻訳を行う際に、フレーズ対応の辞書を利用して、原言語のフレーズを目的言語のフレーズへと変

<sup>1</sup> 奈良先端科学技術大学院大学 情報科学研究科

Nara Institute of Science and Technology

a) miura.akiba.lr9@is.naist.jp

b) neubig@is.naist.jp

c) ssakti@is.naist.jp

d) tomoki@is.naist.jp

e) s-nakamura@is.naist.jp

換し、並び替える。様々な訳出候補が存在する中で、次式に基づき、与えられた原言語の文  $f$  を、最も確率の高い目的言語文  $\hat{e}$  に翻訳する。

$$\hat{e} = \arg \max_e Pr(e|f) \quad (1)$$

$$= \arg \max_e \sum_{m=1}^M \lambda_m h_m(e, f) \quad (2)$$

ここで、 $h_m(e, f)$  は素性関数であり、 $\lambda_m$  はその重みである。素性関数として用いられるものには、双方向のフレーズ翻訳確率、双方向の語彙翻訳確率、単語ペナルティ、フレーズペナルティ、言語モデル、並び替えモデルなどがある。 $\lambda_m$  は、パラメータ調整用データにおける BLEU スコア [11] などの自動評価尺度が最大となるように設定される [12]。

PBMT は、翻訳対象である 2 言語間の対訳データさえ用意すれば容易に学習し、高速な翻訳を行うことが可能であり、多くの研究や実用システムで利用されている。しかし、文の構造を考慮しない手法であるため、単語の並び替えが効果的に行えない傾向にある。高度な並び替えモデルを導入することは可能であるが [13][14]、長距離の並び替えは未だ困難であり、ピボット翻訳で用いることは容易ではない。

## 2.2 階層的フレーズベース翻訳

PBMT の汎用性を残しつつ、並び替えの精度を高めた方法として、Hiero[9] が Chiang によって提案されている。PBMT のように単語列を扱うだけでなく、下記のような変数を含めたルールも利用する。

$$[X_1] \text{ visit } [X_2] \rightarrow [X_1] \text{ は } [X_2] \text{ を訪れる}$$

$X_1$  や  $X_2$  に該当する単語列を先ず翻訳し、ルールの  $X_1$  とルールの  $X_2$  の代わりに代入していくことで翻訳を行う。つまり、適用する各ルールと、 $X_1, X_2$  に当てはまる様々な候補の確率とルール自体の確率、さらに言語モデルを考慮することで、最適な翻訳文を選択する。従って、並び替えを個別のモデルで行うのではなく、並び替えと語彙選択を同時に行うルールを利用することで、PBMT よりも高い並び替え精度を得ることが可能となる。その反面、翻訳に多くの時間を要し、各モデルのサイズも大きくなる傾向がある。

## 3. ピボット翻訳手法

PBMT におけるピボット翻訳手法が数多く考案されており、本節では代表的なピボット翻訳手法について紹介する。例として、中間言語として英語を用いた、フランス語から中国語へのピボット翻訳について説明を行うが、任意の言語組について応用可能であり、英語以外の言語を中間言語に用いることも当然可能である。

### 3.1 逐次的ピボット翻訳

逐次的なピボット翻訳 [3] による、仏中翻訳の様子を図 1 に示す。この方式では先ず、仏英、英中それぞれの言語対で、対訳コーパスを用いて翻訳システムを構築する。そしてフランス語の入力文を英語へ翻訳し、英語の訳文を中国語に翻訳することで、結果的にフランス語から中国語への翻訳が可能となる。この手法は機械翻訳の入力と出力のみを利用するため、PBMT である必然性はなく、任意の機械翻訳システムを組み合わせることができる。優れた 2 つの機械翻訳システムがあれば、そのまま高精度なピボット翻訳が期待できることや、既存のシステムを使い回せること、実現が非常に容易であることが利点と言える。逆に、最初の翻訳システムの翻訳誤りが次のシステムに伝播し、加算的な誤りによって精度が落ちることは欠点となる。

また、この図では最も確率が高くなる英語の 1 文のみに訳しているが、仏英、英中それぞれの翻訳段階で  $n$ -best の文に訳し、探索の幅を広げたマルチセンテンス方式も存在する [4]。

### 3.2 コーパス翻訳方式

直接的にコーパスを翻訳するピボット翻訳方式 [3] による、仏中翻訳の様子を図 2 に示す。この方式では先ず、仏英、英中のうちの片側、英中翻訳の対訳コーパスのみを用いて翻訳システムを構築する。そして仏英翻訳の対訳コーパスの英語側の全文を英中翻訳にかけることで、中国語に翻訳されたコーパスが得られ、元のフランス語のコーパスと合わせることで仏中の対訳が得られる。これによって得られた仏中の疑似対訳コーパスを用いて、SMT のモデルを学習することが可能となる。対訳コーパスの翻訳時に、少しの翻訳誤りが含まれていても、統計モデルの学習に大きく影響しなければ、高精度な訳出が期待できる。既存のシステムから新しい学習データやシステムを作り直すことになるため、一度疑似対訳コーパスを作ってしまうと、それ以降の学習過程が通常の統計翻訳と同等であることは長所となる。

De Gispert らは、スペイン語を中間言語としたカタルーニャ語と英語のピボット翻訳で、逐次的ピボット翻訳とコーパス翻訳方式によるピボット翻訳の比較実験 [3] を行った。その結果、これらの手法間で有意な差は示されなかった。

### 3.3 テーブル合成方式

PBMT では、対訳コーパスによってフレーズ対応を学習してスコア付けした翻訳モデルを、フレーズテーブルと呼ばれる形式で格納する。フレーズテーブルを合成することで仏中のピボット翻訳を行う様子を図 3 に示す。

Cohn らによるフレーズテーブル合成手法 [7] では、PBMT の翻訳確率を格納した仏英、英独のフレーズテーブル  $T_{FE}, T_{EG}$  から、仏独の翻訳確率を推定してフレーズテー

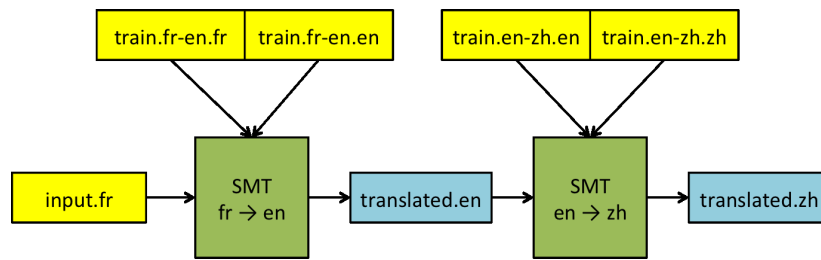


図 1 逐次的ピボット翻訳

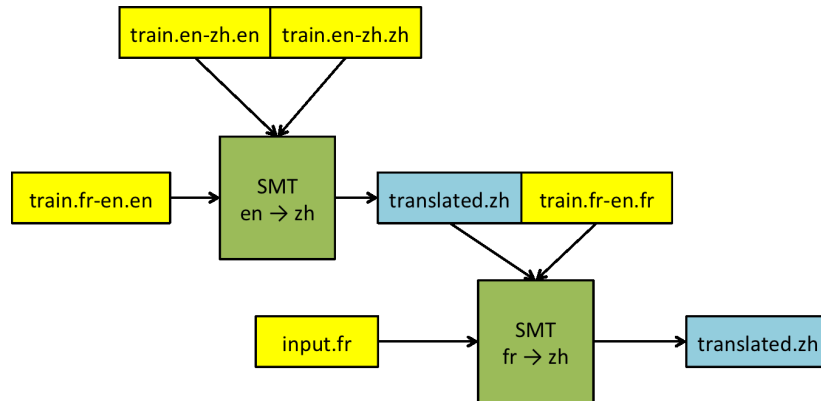


図 2 コーパス翻訳方式

ブル  $T_{FG}$  を合成する.  $T_{FG}$  を作成するには, フレーズ翻訳確率  $\phi(\cdot)$  と語彙翻訳確率  $p_\omega(\cdot)$  を用いて, 次式のように翻訳確率の推定を行う.

$$\phi(\bar{f}|\bar{g}) = \sum_{\bar{e} \in T_{FE} \cap T_{EG}} \phi(\bar{f}|\bar{e}) \phi(\bar{e}|\bar{g}) \quad (3)$$

$$\phi(\bar{g}|\bar{f}) = \sum_{\bar{e} \in T_{FE} \cap T_{EG}} \phi(\bar{g}|\bar{e}) \phi(\bar{e}|\bar{f}) \quad (4)$$

$$p_\omega(\bar{f}|\bar{g}) = \sum_{\bar{e} \in T_{FE} \cap T_{EG}} p_\omega(\bar{f}|\bar{e}) p_\omega(\bar{e}|\bar{g}) \quad (5)$$

$$p_\omega(\bar{g}|\bar{f}) = \sum_{\bar{e} \in T_{FE} \cap T_{EG}} p_\omega(\bar{g}|\bar{e}) p_\omega(\bar{e}|\bar{f}) \quad (6)$$

ただし,  $\bar{f}, \bar{e}, \bar{g}$  はそれぞれフランス語, 英語, ドイツ語のフレーズであり,  $\bar{e} \in T_{FE} \cap T_{EG}$  は英語のフレーズ  $\bar{e}$  が  $T_{FE}, T_{EG}$  の双方に含まれていることを示す. この手法では, 翻訳確率の推定を行うために全フレーズ対応の組合せを求めて算出する必要があるため, 大規模なフレーズテーブルの合成には長い時間を要するが, 既存のモデルデータから精度の高い訳出が期待できる.

Utiyama らは, 英語を中間言語とした複数の言語対で, 逐次的ピボット翻訳とテーブル合成に基づくピボット翻訳で比較実験を行った [4]. その結果, フレーズテーブル合成方式では,  $n=1$  の単純な逐次的ピボット翻訳や,  $n=15$  のマルチセンテンス方式よりも高い BLEU スコアが得られたと報告している.

## 4. 実験

本研究では, 前節で述べたピボット翻訳手法を Hiero に対して応用し, 比較実験を行った.

### 4.1 実験手順

前節で紹介したピボット翻訳手法のうち, 実現が非常に容易で比較しやすい 3.1 節の逐次的ピボット翻訳と, PBMT で高い実用性が示された 3.3 節のテーブル合成方式を Hiero で応用した場合に, どのような結果が得られるかを調査する評価実験を行った. PBMT モデルの構築には Moses[15], Hiero モデルの構築には Travatar[16] を利用する. Moses PBMT, Travatar Hiero のそれぞれについて以下のように翻訳の学習と評価を行い, ピボット翻訳手法の違いによる翻訳精度を比較した. ここでは前節と同じく, 英語を中間言語とした, フランス語と中国語のピボット翻訳を例に説明を行う.

#### Direct (直接翻訳)

仏英, 英中, 仏中の言語対それぞれについて対訳コーパスから翻訳モデルを学習, 最適化, 評価

#### Cascade (逐次的ピボット翻訳)

仏英, 英中で学習された翻訳モデルでパイプライン処理を行い, 仏中翻訳の評価

#### Triangulation (テーブル合成方式)

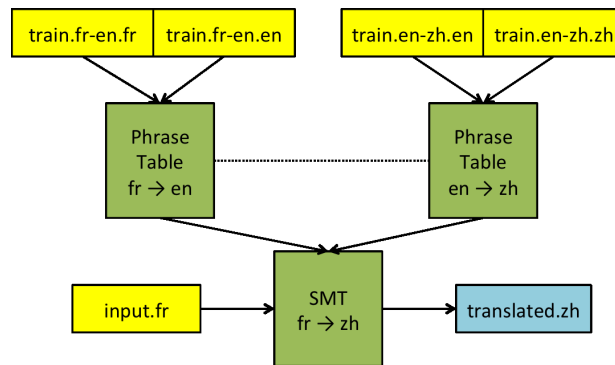


図 3 テーブル合成方式

仏英、英中で学習された翻訳モデルから翻訳確率の推定により仏中翻訳モデルを学習、最適化、評価

Direct では、目的の言語対の対訳コーパスを用意して、通常の方法で直接 SMT の学習と翻訳を行う。Cascade は、3.1 節で述べた逐次のピボット翻訳に相当し、PBMT や Hiero といった枠組みに関わらず、2つの機械翻訳システムをそのまま逐次的に実行可能である。Triangulation では、3.3 節で述べた手法で双方向のフレーズ翻訳確率と語彙翻訳確率を推定し、仏中の翻訳モデルを作成する。

Moses PBMT では翻訳モデルとしてフレーズテーブルの各レコードにフレーズ対応と翻訳確率が格納されているため、数式 (3)-(6) に基づき確率推定を行い、フォーマットに従い仏中翻訳のフレーズテーブルを作成した。Moses の標準設定ではフレーズ対応のみでなく、各フレーズに対して並び替え確率を付与する語彙化並び替えモデル [17] も学習するが、並び替えモデルのテーブル合成手法は現時点では確立されていない。そのため、今回の実験では論点を絞るため、並び替えの素性として歪みペナルティのみを用いる。

Travatar Hiero では、単語列としてのフレーズ対応ではなく、非終端記号を含んだルール対応を用いるが、それぞれのルール対応について PBMT と同様、双方向のフレーズ翻訳確率と語彙翻訳確率をルールテーブルに格納している。本実験では数式 (3)-(6) の  $\bar{f}, \bar{e}, \bar{g}$  をそのまま非終端記号を含めた単語列と見なして、同様にルールテーブルの合成を行った。

上記では仏英中で説明を行ったが、同様に複数の組合せでピボット翻訳を行い、評価した。一部、英語以外の言語を中間言語として用いた評価も行っている。

#### 4.2 実験条件

本稿の実験では、対訳コーパスとして、MultiUN 国連文書コーパス [10] を用いた。これは主として英語を元に、フランス語、中国語、アラビア語、ロシア語、スペイン語、

ドイツ語に訳された多言語コーパスである。コーパスを構成する文章は人手によってアライメントされているが、単位のアライメントは自動的に推定されている。本研究では、文量の関係でスペイン語とドイツ語を除外した 5ヶ国語で、1対1対応の文のみを利用し、更に信頼度が 0.5 以下の文を破棄することで、ある程度の文アライメント精度を確保した。このようなフィルタリングを行っても、300 万文以上からなる大規模なコーパスが得られたが、全文を利用すると、Hiero モデルの学習時に巨大なモデルデータが生成され、現実的な時間でルールテーブルの合成が行えない都合上、翻訳モデルと言語モデルの学習には 50 万文のみを用いた。また、最適化と評価のために、それぞれ 1500 文ずつを別途に抜き出して用いた。表 1 に実験で用いたデータの詳細を示す。

Dataset	Lang	Words	Sentences	Average
				Sentence Length
Train	En	13.2M	500k	26.3
	Fr	15.7M	500k	31.3
	Zh	12.4M	500k	24.8
	Ar	11.6M	500k	23.2
	Ru	11.9M	500k	23.9
Dev	En	37.9k	1.5k	25.3
	Fr	44.9k	1.5k	29.9
	Zh	35.0k	1.5k	23.4
	Ar	33.2k	1.5k	22.2
	Ru	34.5k	1.5k	23.0
Test	En	38.5k	1.5k	25.7
	Fr	45.2k	1.5k	30.2
	Zh	36.0k	1.5k	24.0
	Ar	33.6k	1.5k	22.2
	Ru	34.7k	1.5k	23.2

表 1 実験に用いた対訳コーパスのデータ内訳

中国語の単語分割には KyTea[18] を用いた。目的言語の

言語モデルには Moses PBMT, Travatar Hiero で共通して, KenLM[19] で 5-gram モデルの学習を行った. また, 対訳データ間のアライメントを取るツールとして GIZA++[20] を用いた. Moses や Travatar では, 並び替えモデルの学習を行わない以外は, デフォルトの設定を用いて実験を行った. また, それぞれのピボット翻訳手法の評価で共通して, 自動評価尺度には BLEU[11] を用いており, 最適化では BLEU スコアが最大となるよう調整を行った.

### 4.3 実験結果

4.1 で説明した手順のうち, 対訳コーパスを用いた学習と最適化の後に, 直接翻訳 (Direct) を行い評価した結果を表 2 に示す. また, 様々な言語と機械翻訳方式の組合せについて Direct, Triangulation, Cascade の各方式で翻訳を行い評価した結果を表 3 に示す. 表 3 のうち, Direct の評価のみ, 中間言語を用いていないため, Pivot の列に示された言語は無関係であることに注意されたい. Triangulation と Cascade の評価のうち, BLEU スコアが高くなっているものを各行で太字にしてある.

Lang 1	Lang 2	BLEU Score [%]			
		Moses		Hiero	
		→	←	→	←
En	Ar	43.03	52.47	37.22	47.82
En	Fr	53.58	54.68	50.33	49.56
En	Ru	46.21	53.59	41.03	49.66
En	Zh	33.87	40.20	34.91	40.80
Ar	Zh	31.54	30.29	29.84	28.93
Fr	Ru	41.65	47.43	34.70	43.38
Fr	Zh	29.77	35.38	28.05	34.36
Ru	Zh	32.46	30.64	30.78	30.50

表 2 言語対と機械翻訳手法の違いによる翻訳精度の比較

先行研究により, PBMT のピボット翻訳手法では Triangulation が Cascade よりも高い精度の訳出を行えることが示されており, このことは表 3 でも再現されている. しかし, この表においては, 英語を中間言語とした場合の Hiero のピボット翻訳手法では, Triangulation が Cascade よりも低くなっている. また, 仏英中や中英仏の Cascade では, Direct よりもスコアが高いという結果も得られた.

まず, Cascade で Direct よりも高いスコアが得られた原因としては, 今回の実験で用いた多言語コーパスが英語を中心にして翻訳されているという性格上, 英語を含む言語対では質の高い対訳コーパスが得られるが, 英語を含まない言語対では, 比較的質の低い対訳コーパスが得られた可能性が考えられる. このことは, 表 2 により, 言語対の一方を英語以外のどの言語に固定しても, 他方を英語にした言語対のスコアが, それ以外の組合せよりも高くなること

から確認できる. そして仏英, 英仏の翻訳精度が非常に高いため, 直接フランス語と中国語の対訳で学習したものよりも精度の高い訳出が行われたものと考えられる.

一方, 英語を中間言語とした場合, Hiero の Triangulation で Cascade より低いスコアが得られた原因としては, テーブル合成方式における, 翻訳確率の近似方法が PBMT に比べて問題となることが考えられる. 4.1 で述べたように, Hiero の Triangulation でも PBMT と同様に数式 (3)-(6) を用いて, ピボット側で一致するルールの対応から翻訳確率の推定を行う. しかし, 実際はルール中の変数にどのような役割のフレーズが当てはまるかは不明であり, 1 つの変数  $X$  のみを用いる Hiero 文法の制約が十分でないと考えられる. 例えば,

$$\begin{aligned} a X b \text{ (フランス語)} &\rightarrow X c \text{ (英語)} \\ X c \text{ (英語)} &\rightarrow d X e \text{ (中国語)} \end{aligned}$$

という 2 つのルール対応から,

$$a X b \text{ (フランス語)} \rightarrow c X d \text{ (中国語)}$$

というルール対応を作って確率的にスコアを算出するのは, 英語側で変数  $X$  にどのようなフレーズが代入されるか分からないため, 不適切なルール対応が合成されるおそれがある. このような問題を解決するためには, 句構造のラベルを付与したルールを用いる [21] アプローチなどが考えられる.

表 3 の下 4 行では, 英語以外の中間言語を用いた場合のピボット翻訳の評価を示しており, 限られた組合せであるが, Hiero の Triangulation でも Cascade より高いスコアを示している. 検証を行うには, より多くの組合せを考慮する必要があるが, この表からも, 中間言語の選択がピボット翻訳の性能に影響を及ぼすことが分かり, 更なる調査を要する.

また, 表 2 や表 3 から, 中英, 英中翻訳では Hiero が PBMT の精度を上回り, それ以外の言語対では Hiero が PBMT の精度を下回る傾向が見られる. この原因としては, 並び替えを多く要しない言語対では PBMT の並び替えモデルでおおよそ十分であり, Hiero ではモデルサイズの制約で最大フレーズ長が 5 に制限され, Moses PBMT の最大フレーズ長である 7 より短いため, 長いフレーズの用語が多用される国連コーパスでは PBMT の方が有効であることが考えられる. この点についても, 今後詳細な調査を行う予定である.

## 5. おわりに

本研究では, ピボット翻訳を階層的フレーズベース翻訳に適用し, その有効性を検証した. その結果, PBMT で安定して高い精度を実現するテーブル合成方式は Hiero にお

Source	Pivot	Target	MT Method	BLEU Score [%]		
				Direct	Triangulation	Cascade
Ar	En	Zh	Moses	31.54	<b>29.40</b>	28.78
			Hiero	29.84	28.41	<b>29.11</b>
Fr	En	Zh	Moses	29.77	<b>29.31</b>	29.16
			Hiero	28.05	27.57	<b>29.64</b>
Ru	En	Zh	Moses	32.46	<b>30.67</b>	30.25
			Hiero	30.78	29.32	<b>30.10</b>
Zh	En	Ar	Moses	30.29	<b>28.82</b>	28.27
			Hiero	28.93	26.22	<b>27.62</b>
Zh	En	Fr	Moses	35.38	<b>35.21</b>	35.16
			Hiero	34.36	32.26	<b>35.23</b>
Zh	En	Ru	Moses	30.64	<b>30.12</b>	29.55
			Hiero	30.50	27.82	<b>29.88</b>
En	Fr	Zh	Moses	33.87	<b>32.13</b>	31.09
			Hiero	34.91	<b>32.79</b>	30.57
Zh	Fr	En	Moses	40.20	<b>36.52</b>	35.37
			Hiero	40.80	<b>34.94</b>	34.28
En	Zh	Fr	Moses	53.58	<b>45.29</b>	41.21
			Hiero	50.33	<b>43.79</b>	35.78
Fr	Zh	En	Moses	54.68	<b>45.22</b>	41.12
			Hiero	49.56	<b>43.51</b>	35.16

表 3 ピボット翻訳手法の違いによる翻訳精度の比較

いて必ずしも逐次のピボット翻訳より高い精度を実現するとは限らないことが分かった。

今後の課題としては、Hiero のモデルで、ルールテーブル合成方式の翻訳精度を高めることを試みる予定である。これを実現する方法として、文の構造に則した制約を考慮するモデル学習手法や、より洗練された確率推定法が考えられる。また、同じ言語対であっても、ピボットとして用いる中間言語によって差が出ることも PBMT を用いた研究で示されており [22]、言語対を固定して中間言語を変えた場合の Hiero における影響を調査することも予定している。

#### 参考文献

- [1] Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, Vol. 19, pp. 263–312, 1993.
- [2] Christopher Dyer, Aaron Cordova, Alex Mont, and Jimmy Lin. Fast, easy, and cheap: construction of statistical machine translation models with mapreduce. In *Proc. WMT*, pp. 199–207, 2008.
- [3] Adrià de Gispert and José B. Mariño. Catalan-english statistical machine translation without parallel corpus: Bridging through spanish. In *Proc. of LREC 5th Workshop on Strategies for developing machine translation for minority languages*, 2006.
- [4] Masao Utiyama and Hitoshi Isahara. A comparison of pivot methods for phrase-based statistical machine translation. In *Proc. NAACL*, pp. 484–491, 2007.
- [5] Jörg Tiedemann. Character-based pivot translation for under-resourced languages and domains. In *EACL12*, pp. 141–151, 2012.
- [6] Xiaoning Zhu, Zhongjun He, Hua Wu, Conghui Zhu, Haifeng Wang, and Tiejun Zhao. Improving pivot-based statistical machine translation by pivoting the co-occurrence count of phrase pairs. In *Proc. EMNLP*, 2014.
- [7] Trevor Cohn and Mirella Lapata. Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proc. ACL*, pp. 728–735, June 2007.
- [8] Phillip Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase-based translation. In *Proc. HLT*, pp. 48–54, 2003.
- [9] David Chiang. Hierarchical phrase-based translation. *Computational Linguistics*, Vol. 33, No. 2, pp. 201–228, 2007.
- [10] Andreas Eisele and Yu Chen. MultiUN: A Multilingual Corpus from United Nation Documents. In *Proc. of the Seventh conference on International Language Resources and Evaluation*, pp. 2868–2872, 2010.
- [11] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proc. ACL*, pp. 311–318, 2002.
- [12] Franz Josef Och. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, pp. 160–167, 2003.
- [13] Michel Galley and Christopher D. Manning. A simple and effective hierarchical phrase reordering model. In *Proc. EMNLP*, pp. 848–856, 2008.
- [14] Isao Goto, Masao Utiyama, Eiichiro Sumita, Akihiro

- Tamura, and Sadao Kurohashi. Distortion model considering rich context for statistical machine translation. In *Proc. ACL*, pp. 155–165, August 2013.
- [15] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. Moses: Open source toolkit for statistical machine translation. In *Proc. ACL*, pp. 177–180, 2007.
- [16] Graham Neubig. Travatar: A forest-to-string machine translation engine based on tree transducers. In *Proc. ACL Demo Track*, pp. 91–96, 2013.
- [17] Philipp Koehn, Amittai Axelrod, Alexandra Birch Mayne, Chris Callison-Burch, Miles Osborne, and David Talbot. Edinburgh system description for the 2005 IWSLT speech translation evaluation. In *Proc. IWSLT*, 2005.
- [18] Graham Neubig, Yosuke Nakata, and Shinsuke Mori. Pointwise prediction for robust, adaptable Japanese morphological analysis. In *Proc. ACL*, pp. 529–533, 2011.
- [19] Kenneth Heafield. KenLM: faster and smaller language model queries. In *Proc. WMT*, July 2011.
- [20] Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, Vol. 29, No. 1, pp. 19–51, 2003.
- [21] Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. What’s in a translation rule? In *Proc. HLT*, pp. 273–280, 2004.
- [22] Michael Paul, Hirofumi Yamamoto, Eiichiro Sumita, and Satoshi Nakamura. On the importance of pivot language selection for statistical machine translation. In *Proc. NAACL*, pp. 221–224, June 2009.