

# 中間言語情報を記憶するピボット翻訳手法

三浦 明波<sup>†</sup> · Graham Neubig<sup>†,††</sup> · Sakriani Sakti<sup>†</sup> · 戸田 智基<sup>†††</sup> · 中村 哲<sup>†</sup>

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手法の一つである。複数のピボット翻訳手法が考案されている中でも、特に中間言語を介して2つの翻訳モデルを合成するテーブル合成手法で高い翻訳精度を達成可能と報告されている。ところが、従来のテーブル合成手法では、フレーズ対応推定時に用いた中間言語の情報は消失し、翻訳時には利用できない問題が発生する。本論文では、合成時に用いた中間言語の情報も記憶し、中間言語モデルを追加の情報源として翻訳に利用する新たなテーブル合成手法を提案する。また、国連文書による多言語コーパスを用いた実験により、本手法で評価を行ったすべての言語の組み合わせで従来手法よりも有意に高い翻訳精度が得られた。

キーワード：統計的機械翻訳, 多言語翻訳, ピボット翻訳, 同期文脈自由文法, 言語モデル, 対訳コーパス

## Improving Pivot Translation by Remembering the Pivot

AKIVA MIURA<sup>†</sup>, GRAHAM NEUBIG<sup>†,††</sup>, SAKRIANI SAKTI<sup>†</sup>, TOMOKI TODA<sup>†††</sup> and SATOSHI NAKAMURA<sup>†</sup>

In statistical machine translation, the pivot translation approach allows for translation of language pairs with little or no parallel data by introducing a third language for which data exists. In particular, the triangulation method, which translates by combining source-pivot and pivot-target translation models into a source-target model is known for its high translation accuracy. However, in the conventional triangulation method, information of pivot phrases is forgotten, and not used in the translation process. In this research, we propose a novel approach to remember the pivot phrases in the triangulation stage, and use a pivot language model as an additional information source at translation phase. Experimental results on the united nations parallel corpus showed significant improvements in all tested combinations of languages.

**Key Words:** *Statistical Machine Translation, Multilinguality, Pivot Translation, Synchronous Context-Free Grammars, Language Models, Parallel Corpora*

<sup>†</sup> 奈良先端科学技術大学院大学情報科学研究科, Graduate School of Information Science, Nara Institute of Science and Technology

<sup>††</sup> カーネギーメロン大学言語技術研究所, Language Technologies Institute, Carnegie Mellon University

<sup>†††</sup> 名古屋大学情報基盤センター, Information Technology Center, Nagoya University

## 1 はじめに

### 1.1 研究背景

言語は、人間にとって主要なコミュニケーションの道具であると同時に、話者集団にとっては社会的背景に根付いたアイデンティティでもある。母国語の異なる相手と意思疎通を取るためには、翻訳は必要不可欠な技術であるが、専門の知識が必要となるため、ソフトウェア的に代行できる機械翻訳の技術に期待が高まっている。英語と任意の言語間での翻訳で機械翻訳の実用化を目指す例が多いが、英語を含まない言語対においては翻訳精度がまだ実用的なレベルに達していないことが多く、英語を熟知していない利用者にとって様々な言語間で機械翻訳を支障なく利用できる状況とは言えない。

人手で翻訳規則を記述するルールベース機械翻訳 (Rule-Based Machine Translation; RBMT (Nirenburg 1989)) では、対象の2言語に精通した専門家の知識が必要であり、多くの言語対において、多彩な表現を広くカバーすることも困難である。そのため、近年主流の機械翻訳方式であり、機械学習技術を用いて対訳コーパスから自動的に翻訳規則を獲得する統計的機械翻訳 (Statistical Machine Translation; SMT (Brown, Pietra, Pietra, and Mercer 1993)) について本論文では議論を行う。対訳コーパスとは、2言語間で意味の対応する文や句を集めたデータのことを指すが、SMTでは学習に使用する対訳コーパスが大規模になるほど、翻訳結果の精度が向上すると報告されている (Dyer, Cordova, Mont, and Lin 2008)。しかし、英語を含まない言語対などを考慮すれば、多くの言語対において、大規模な対訳コーパスを直ちに取得することは困難と言える。このような、容易に対訳コーパスを取得できないような言語対においても、既存の言語資源を有効に用いて高精度な機械翻訳を実現できれば、機械翻訳の実用の幅が大きく広がることになる。

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 (*Pvt*) を用いたピボット翻訳が有効な手法の一つである (de Gispert and Mariño 2006; Cohn and Lapata 2007; Zhu, He, Wu, Zhu, Wang, and Zhao 2014)。中間言語を用いる方法も様々であるが、一方の目的言語と他方の原言語が一致するような2つの機械翻訳システムを利用できる場合、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 (de Gispert and Mariño 2006)) 手法が容易に実現可能である。より高度なピボット翻訳の手法としては、原言語・中間言語 (*Src-Pvt*) と中間言語・目的言語 (*Pvt-Trg*) の2組の言語対のためにそれぞれ学習されたSMTシステムのモデルを合成し、新しく得られた原言語・目的言語 (*Src-Trg*) のSMTシステムを用いて翻訳を行うテーブル合成手法 (Triangulation (Cohn and Lapata 2007)) も提案されており、この手法で特に高い翻訳精度が得られたと報告されている (Utiyama and Isahara 2007)。

これらの手法は特に、今日広く用いられているSMTの枠組の一つであるフレーズベース機械翻訳 (Phrase-Based Machine Translation; PBMT (Koehn, Och, and Marcu 2003)) について数



図 1 2 組の単語対応から新しい単語対応を推定

多く提案され、検証されてきた。しかし、PBMTにおいて有効性が検証されたピボット翻訳手法が、異なる SMT の枠組でも同様に有効であるかどうかは明らかにされていない。例えば英語と日本語、英語と中国語といった語順の大きく異なる言語間の翻訳では、同期文脈自由文法 (Synchronous Context-Free Grammar; SCFG (Chiang 2007)) のような木構造ベースの SMT によって高度な単語並び替えに対応可能であり、PBMT よりも高い翻訳精度を達成できると報告されている。そのため、PBMT において有効性の知られているピボット翻訳手法が、SCFG による翻訳でも有効であるとすれば、並び替えの問題に高度に対応しつつ直接 *Src-Trg* の対訳コーパスを得られない状況にも対処可能となる。

また、テーブル合成手法では、*Src-Pvt* フレーズ対応と *Pvt-Trg* フレーズ対応から、正しい *Src-Trg* フレーズ対応と確率スコアを推定する必要がある。図 1 に示す例では、個別に学習された (a) の日英翻訳および (b) の英伊翻訳における単語対応から、日伊翻訳における単語対応を推定したい場合、(c) のように単語対応を推定する候補は非常に多く、(d) のように正しい推定結果を得ることは困難である。その上、図 1(c) のように推定された *Src-Trg* の単語対応からは、原言語と目的言語の橋渡しをしていた中間言語の単語情報が分からないため、翻訳を行う上で重要な手がかりとなり得る情報を失ってしまうことになる。このように語義曖昧性や言語間の用語法の差異により、ピボット翻訳は通常の翻訳よりも本質的に多くの曖昧性の問題を抱えており、さらなる翻訳精度の向上には課題がある。

## 1.2 研究目的

本研究では、多言語機械翻訳、とりわけ対訳コーパスの取得が困難である少資源言語対における機械翻訳の高精度化を目指し、従来のピボット翻訳手法を調査、問題点を改善して翻訳精

度を向上させることを目的とする。ピボット翻訳の精度向上に向けて、本論文では2段階の議論を行う。

第1段階目では、従来のPBMTで有効性の知られているピボット翻訳手法が異なる枠組のSMTでも有効であるかどうかを調査する。1.1節で述べたように、PBMTによるピボット翻訳手法においては、テーブル合成手法で高い翻訳精度が確認されているため、木構造ベースのSMTであるSCFGによる翻訳で同等の処理を行うための応用手法を提案する。SCFGとテーブル合成手法によるピボット翻訳が、逐次的ピボット翻訳や、PBMTにおけるピボット翻訳手法よりも高い精度を得られるどうかを比較評価することで、次の段階への予備実験とする<sup>1</sup>。

第2段階目では、テーブル合成手法において発生する曖昧性の問題を解消し、翻訳精度を向上させるための新たな手法を提案する。従来のテーブル合成手法では、図1(c)に示したように、フレーズ対応の推定後には中間言語フレーズの情報が見失われてしまうことを1.1節で述べた。この問題を克服するため、本論文では原言語と目的言語を結び付けていた中間言語フレーズの情報も翻訳モデル中に保存し、原言語から目的言語と中間言語へ同時に翻訳を行うための確率スコアを推定することによって翻訳を行う新しいテーブル合成手法を提案する。通常のSMTシステムでは、入力された原言語文から、目的言語における訳出候補を選出する際、文の自然性を評価し、適切な語彙選択を促すために目的言語の言語モデル（目的言語モデル）を利用する。一方、本手法で提案する翻訳モデルとSMTシステムでは、原言語文に対して目的言語文と中間言語文の翻訳を同時に行うため、目的言語モデルのみではなく、中間言語の言語モデル（中間言語モデル）も同時に考慮して訳出候補の探索を行う。本手法の利点は、英語のように中間言語として選ばれる言語は豊富な単言語資源を得られる傾向が強いため、このような追加の言語情報を翻訳システムに組み込み、精度向上に役立てられることにある<sup>1</sup>。

## 2 統計的機械翻訳

本節では、SMTの基本的な動作原理となる対数線形モデル（2.1節）、SMTの中でも特に代表的な翻訳方式であるフレーズベース機械翻訳（PBMT, 2.2節）と木構造に基づく翻訳方式である同期文脈自由文法（SCFG, 2.3節）、SCFGを3言語以上に対応できるように一般化して拡張された複数同期文脈自由文法（Multi-Synchronous Context-Free Grammar; MSCFG, 2.4節）について説明する。

<sup>1</sup> 本稿の内容の一部は、情報処理学会自然言語処理研究会（三浦, Neubig, Sakti, 戸田, 中村 2014, 2015）およびACL 2015: The 53rd Annual Meeting of the Association for Computational Linguistics (Miura, Neubig, Sakti, Toda, and Nakamura 2015)で報告されている。本稿では、各手法・実験に関する詳細な説明、中国語やアラビア語など語族の異なる言語間での比較評価実験や品詞毎の翻訳精度に関する分析を追加している。

## 2.1 対数線形モデル

SMT の基本的なアイデアは、雑音のある通信路モデル (Shannon 1948) に基いている。ある原言語の文  $\mathbf{f}$  に対して、訳出候補となり得るすべての目的言語文の集合を  $\mathcal{E}(\mathbf{f})$  とする。  $\mathbf{f}$  が目的言語文  $e \in \mathcal{E}(\mathbf{f})$  へと翻訳される確率  $Pr(e|\mathbf{f})$  をすべての  $e$  について計算可能とする。SMT では、  $Pr(e|\mathbf{f})$  を最大化する  $\hat{e} \in \mathcal{E}(\mathbf{f})$  を求める。

$$\hat{e} = \arg \max_{e \in \mathcal{E}(\mathbf{f})} Pr(e|\mathbf{f}) \quad (1)$$

$$= \arg \max_{e \in \mathcal{E}(\mathbf{f})} \frac{Pr(\mathbf{f}|e)Pr(e)}{Pr(\mathbf{f})} \quad (2)$$

$$= \arg \max_{e \in \mathcal{E}(\mathbf{f})} Pr(\mathbf{f}|e)P(e) \quad (3)$$

しかし、このままでは様々な素性を取り入れたモデルの構築が困難であるため、近年では以下のような対数線形モデルに基づく定式化を行うことが一般的である (Och 2003)。

$$\hat{e} = \arg \max_{e \in \mathcal{E}(\mathbf{f})} Pr(e|\mathbf{f}) \quad (4)$$

$$\approx \arg \max_{e \in \mathcal{E}(\mathbf{f})} \frac{\exp(\mathbf{w}^T \mathbf{h}(\mathbf{f}, e))}{\sum_{e'} \exp(\mathbf{w}^T \mathbf{h}(\mathbf{f}, e'))} \quad (5)$$

$$= \arg \max_{e \in \mathcal{E}(\mathbf{f})} \mathbf{w}^T \mathbf{h}(\mathbf{f}, e) \quad (6)$$

ここで、 $\mathbf{h}$  は素性ベクトルであり、翻訳の枠組毎に定められた次元数を持ち、推定された対数確率スコア、導出に伴う単語並び替え、各種ペナルティなどを与える。素性ベクトル中のとりのり重要な要素として、言語モデルと翻訳モデルが挙げられる。言語モデル  $Pr(e)$  は、与えられた文の単語の並びが目的言語においてどの程度自然で流暢であるかを評価するために用いられる。翻訳モデル  $Pr(\mathbf{f}|e)$  は、翻訳文の尤もらしさを規定するための統計モデルであり、対訳コーパスから学習を行う。翻訳モデルは SMT の枠組によって学習・推定方法が異なっており、次節以降で詳細を述べる。 $\mathbf{w}$  は  $\mathbf{h}$  と同じ次元を持っており、素性ベクトルの各要素に対する重み付けを行う。 $\mathbf{w}$  の各要素を最適な値に調整するためには、対訳コーパスを学習用データや評価用データとは別に切り分けた、開発用データを利用し、原言語文の訳出と参照訳（目的言語側の正解訳）との類似度を評価するための自動評価尺度 BLEU (Papineni, Roukos, Ward, and Zhu 2002) などが最大となるようパラメータを求める (Och 2003)。2.2 節以降で説明する各種翻訳枠組も、この対数線形モデルに基いているが、用いる素性はそれぞれで異なる。

## 2.2 フレーズベース機械翻訳

Koehn らによる フレーズベース機械翻訳 (PBMT (Koehn et al. 2003)) は SMT で最も代表的な翻訳枠組である。PBMT の翻訳モデルを学習する際には、先ず対訳コーパスから単語アラインメント (Brown et al. 1993) を学習し、アライン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。

例えば、学習用対訳データから図 2 のような単語対応が得られたとする<sup>2</sup>。得られた単語対応からフレーズの対応を見つけ出して抽出を行う例を図 3 に示す。図のように、与えられた単語対応から抽出されるフレーズ対応の長さは一意に定まらず、複数の長さのフレーズ対応が抽出される。ただし、抽出されるフレーズ対応には、フレーズの内外を横断するような単語対応が存在しないという制約が課され、フレーズの最大長なども制限される。このようにして抽出されたフレーズ対の一覧を元に、フレーズ対や各フレーズの頻度を計算し、PBMT の翻訳モデルが学習される。

PBMT の翻訳モデルは抽出されたフレーズを翻訳の基本単位とし、これによって効率的に慣用句のような連続する単語列の翻訳規則を学習し、質の高い翻訳が可能である。フレーズの区切り方によって、与えられた原言語文から、ある目的言語文へ翻訳されるための導出も複数の候補があり、それぞれの導出で用いられるフレーズ対の確率スコアや並び替えも考慮して最終的な翻訳確率を推定する。式 (6) の対数線形モデルによって確率スコア最大の翻訳候補を探索するが、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。

PBMT は、翻訳対象である 2 言語間の対訳コーパスさえ用意すれば、容易に学習し、高速な翻訳を行うことが可能であるため、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 (Goto, Utiyama, Sumita, Tamura, and Kurohashi 2013)、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。

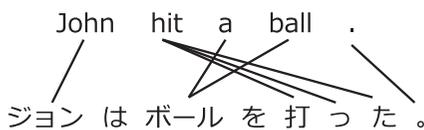


図 2 英-日 単語アラインメント

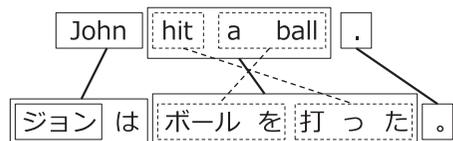


図 3 英-日 フレーズ抽出

<sup>2</sup> 日本語や中国語、タイ語のように、通常の文では単語をスペースで区切らないような言語では、先ず単語分割を行うツールを用いて分かち書きを行う必要がある。

## 2.3 同期文脈自由文法

本節では、木構造に基づく SMT の枠組である同期文脈自由文法 (SCFG (Chiang 2007)) について説明する。SCFG は、階層的フレーズベース翻訳 (Hierarchical Phrase-Based Translation; Hiero (Chiang 2007)) を代表とする様々な翻訳方式で用いられている。SCFG は、以下のような同期導出規則によって構成される。

$$X \longrightarrow \langle \bar{s}, \bar{t} \rangle \quad (7)$$

ここで、 $X$  は同期導出規則の親記号であり、 $\bar{s}$  と  $\bar{t}$  はそれぞれ原言語と目的言語における終端記号と非終端記号からなる記号列である。 $\bar{s}$  と  $\bar{t}$  にはそれぞれ同じ数の非終端記号が含まれ、対応する記号に対して同じインデックスが付与される。以下に日英翻訳における導出規則の例を示す。

$$X \longrightarrow \langle X_0 \text{ of } X_1, X_1 \text{ の } X_0 \rangle \quad (8)$$

Hiero 翻訳モデルのための SCFG の学習手法では、先ず PBMT と同等のアイデアで、対訳コーパスから学習された単語アラインメントを元にフレーズを抽出する。そしてフレーズ対応中の部分フレーズ対応に対しては、非終端記号  $X_i$  で置き換えてよいというヒューリスティックを用いて、多くの SCFG ルールが自動抽出される。例えば、図 2 の単語アラインメントを用いて、以下のような同期導出規則を得ることができる。

$$X \longrightarrow \langle X_0 \text{ hit } X_1 \text{ ., } X_0 \text{ は } X_1 \text{ を打った。} \rangle \quad (9)$$

$$X \longrightarrow \langle \text{John, ジョン} \rangle \quad (10)$$

$$X \longrightarrow \langle \text{a ball, ボール} \rangle \quad (11)$$

また、初期非終端記号  $S$  と初期導出規則  $S \longrightarrow \langle X_0, X_0 \rangle$ 、抽出された上記の導出規則を用いて、以下のような導出が可能である。

$$S \Longrightarrow \langle X_0, X_0 \rangle \quad (12)$$

$$\Longrightarrow \langle X_1 \text{ hit } X_2 \text{ ., } X_1 \text{ は } X_2 \text{ を打った。} \rangle \quad (13)$$

$$\Longrightarrow \langle \text{John hit } X_2 \text{ ., } \text{ジョンは } X_2 \text{ を打った。} \rangle \quad (14)$$

$$\Longrightarrow \langle \text{John hit a ball ., } \text{ジョンは ボール を打った。} \rangle \quad (15)$$

対訳文と単語アラインメントを元に自動的に SCFG ルールが抽出される。抽出された各々のルールには、双方向のフレーズ翻訳確率  $\phi(\bar{s}|\bar{t})$ ,  $\phi(\bar{t}|\bar{s})$ 、双方向の語彙翻訳確率  $\phi_{lex}(\bar{s}|\bar{t})$ ,  $\phi_{lex}(\bar{t}|\bar{s})$ 、ワードペナルティ ( $\bar{t}$  の終端記号数)、フレーズペナルティ (定数 1) の計 6 つのスコアが付与される。

翻訳時には、導出に用いられるルールのスコアと、生成される目的言語文の言語モデルスコアの和を導出確率として最大化するよう探索を行う。言語モデルを考慮しない場合、CKY+法 (Chappelier and Rajman 1998) によって効率的な探索を行ってスコア最大の導出を得ることが可能である。言語モデルを考慮する場合には、キューブ枝狩り (Chiang 2007) などの近似法により探索空間を抑えつつ、目的言語モデルを考慮した探索が可能である。

## 2.4 複数同期文脈自由文法

SCFG を複数の目的言語文の同時生成に対応できるように拡張した手法として、複数同期文脈自由文法 (MSCFG (Neubig, Arthur, and Duh 2015)) が提案されている。SCFG では導出規則中の目的言語記号列  $\bar{t}$  が単一であったが、MSCFG では以下のように  $N$  個の目的言語記号列を有する。

$$X \rightarrow \langle \bar{s}, \bar{t}_1, \dots, \bar{t}_N \rangle \quad (16)$$

通常の MSCFG 学習手法では、SCFG ルール抽出手法を一般化し、3 言語以上の言語間で意味の対応する文を集めた多言語コーパスから多言語導出規則が抽出され、複数の目的言語を考慮したスコアが付与される。本手法の利点として、原言語に対して主要な目的言語が 1 つ存在する場合に、他の  $N - 1$  言語のフレーズを補助的な言語情報として利用し、追加の目的言語モデルによって翻訳文の自然性評価を考慮した訳出を行うことで、結果的に主要な目的言語においても導出規則の選択が改善されて翻訳精度を向上可能なことが挙げられる。

SCFG は対応するフレーズ間で同一のインデックス付き非終端記号を同期して導出させることで、翻訳と単語並び替えを同時に行えるという単純な規則から成り立つために、多言語間の翻訳モデルへの拡張も容易であった。PBMT はフレーズの翻訳と単語並び替えを個別の問題としてモデル化しているため、MSCFG と同様の方法で 3 言語以上のフレーズ対応を学習して複数の目的言語へ同時に翻訳を行うには、並び替え候補をどのように翻訳スコアに反映させるかなどを新たに検討する必要がある。例えば日本語と朝鮮語のように語順が似通っており、ほとんど単語並び替えが発生しない言語の組み合わせでは、並び替えをまったく行わない場合でも高い並び替え精度となるため、並び替え距離に応じたペナルティを与える単純な手法でも高精度となるが、例えば日本語・朝鮮語とは語順の大きく異なる英語を加えた 3 言語間で並び替えをモデル化することは容易ではなく、第二の目的言語の存在が悪影響を与える可能性もある。そのため、本稿では PBMT の多言語拡張を行うことはせず、MSCFG に着目して議論を行う。

### 3 ピボット翻訳手法

2節では, SMT は対訳コーパスから自動的に翻訳規則を獲得し, 統計に基づいたモデルによって翻訳確率スコアが最大となるような翻訳を行うことを述べてきた. 統計モデルであるため, 言語モデルの学習に用いる目的言語コーパスと翻訳モデルの学習に用いられる対訳コーパスが大規模になるほど確率推定の信頼性が向上し, 精度の高い訳出が期待できる. 言語モデルについては, 目的言語の話者数やインターネット利用者数などの影響はあるものの, 比較的取得が容易であるため問題になることは少ない. 一方で対訳コーパスは SMT の要であり, 学習データにカバーされていない単語や表現の翻訳は不可能なため, 多くの対訳データ取得が望ましく, 実用的な SMT システムの構築には数百万文以上の対訳が必要とされている. ところが英語を含まない言語対, 例えば日本語とフランス語のような言語対を考えると, それぞれの言語では単言語コーパスが豊富に取得可能であるにも関わらず, 100 万文を超えるような大規模な対訳データを短時間で獲得することは困難である. このように, SMT の大前提である対訳コーパスは多くの言語対において十分な量を直ちに取得できず, 任意の言語対で翻訳を行うには課題がある.

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. また, 4節では, PBMT で有効性の確認されたピボット翻訳手法であるテーブル合成手法を SCFG で応用するための手法を提案し, 実験による比較評価と考察を述べる. 本節では原言語を *Src*, 目的言語を *Trg*, 中間言語を *Pvt* と表記し, これらの言語対を *Src-Pvt*, *Src-Trg*, *Pvt-Trg* のように表記して説明を行うこととする.

#### 3.1 逐次的ピボット翻訳手法

逐次的ピボット翻訳手法 (**Cascade**) (de Gispert and Mariño 2006) によって *Src* から *Trg* へと翻訳を行う様子を図 4 に示す. この方式では先ず, *Src-Pvt*, *Pvt-Trg* 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 *Src* の入力文を *Pvt* へ翻訳し, *Pvt* の訳文を *Trg* に翻訳することで, 結果的に *Src* から *Trg* への翻訳が可能となる. この手法は機械

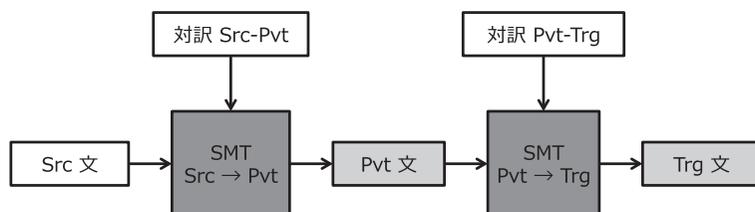


図 4 逐次的ピボット翻訳

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、加法的誤差によって精度が落ちることは欠点となる。Src-Pvt 翻訳システムで確率スコアの高い上位  $n$  文の訳出候補を出力し、Pvt-Trg 翻訳における探索の幅を広げるマルチセンテンス方式も提案されている (Utiyama and Isahara 2007) が、通常より  $n$  倍の探索時間が必要であり、大きな精度向上も報告されていない。

### 3.2 擬似対訳コーパス手法

擬似的に Src-Trg 対訳コーパスを作成することで SMT システムを構築する擬似対訳コーパス手法 (Synthetic) (de Gispert and Mariño 2006) によって、Src-Trg 翻訳を行う様子を図5に示す。この手法ではまず、Src-Pvt, Pvt-Trg のうちの片側、図の例では Pvt-Trg の対訳コーパスを用いて SMT システムを構築する。そして Src-Pvt 対訳コーパスの Pvt 側の全文を Pvt-Trg 翻訳にかけることで、Src-Trg 擬似対訳コーパスが得られる。これによって得られた Src-Trg 擬似対訳コーパスを用いて、SMT の翻訳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、一度擬似対訳コーパスを作ってしまうと、それ以降は通常の SMT と同じ学習手法を用いられることは利点となる。

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手法と擬似対訳コーパス手法によるピボット翻訳手法の比較実験 (de Gispert and Mariño 2006) 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。

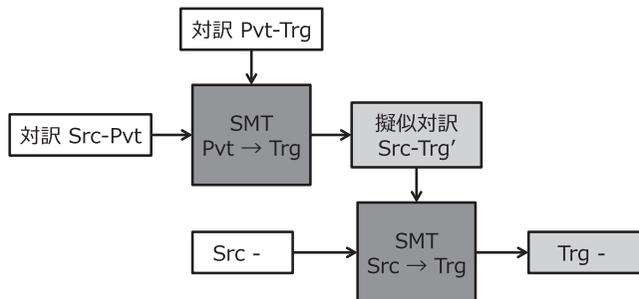


図5 擬似対訳コーパス手法

### 3.3 テーブル合成手法

PBMT, SCFG 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, それぞれフレーズテーブル, ルール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  $Src-Trg$  のピボット翻訳を行う様子を図 6 に示す. Cohn らによるテーブル合成手法 (**Triangulation**) (Cohn and Lapata 2007) では, 先ず  $Src-Pvt$  および  $Pvt-Trg$  の翻訳モデルを対訳コーパスによって学習し, それぞれをフレーズテーブル  $T_{SP}, T_{PT}$  として格納する. 得られた  $T_{SP}, T_{PT}$  から,  $Src-Trg$  の翻訳確率を推定してフレーズテーブル  $T_{ST}$  を合成する.  $T_{ST}$  を作成するには, フレーズ翻訳確率  $\phi(\cdot)$  と語彙翻訳確率  $\phi_{lex}(\cdot)$  を用い, 以下の数式に従って翻訳確率の推定を行う.

$$\phi(\bar{t}|\bar{s}) = \sum_{\bar{p} \in T_{SP} \cap T_{PT}} \phi(\bar{t}|\bar{p}) \phi(\bar{p}|\bar{s}) \tag{17}$$

$$\phi(\bar{s}|\bar{t}) = \sum_{\bar{p} \in T_{SP} \cap T_{PT}} \phi(\bar{s}|\bar{p}) \phi(\bar{p}|\bar{t}) \tag{18}$$

$$\phi_{lex}(\bar{t}|\bar{s}) = \sum_{\bar{p} \in T_{SP} \cap T_{PT}} \phi_{lex}(\bar{t}|\bar{p}) \phi_{lex}(\bar{p}|\bar{s}) \tag{19}$$

$$\phi_{lex}(\bar{s}|\bar{t}) = \sum_{\bar{p} \in T_{SP} \cap T_{PT}} \phi_{lex}(\bar{s}|\bar{p}) \phi_{lex}(\bar{p}|\bar{t}) \tag{20}$$

ここで,  $\bar{s}, \bar{p}, \bar{t}$  はそれぞれ  $Src, Pvt, Trg$  のフレーズであり,  $\bar{p} \in T_{SP} \cap T_{PT}$  はフレーズ  $\bar{p}$  が  $T_{SP}, T_{PT}$  の双方に含まれていることを示す. 式 (17)–(20) は, 以下のような条件を満たす無記憶通信路モデルに基づいている.

$$\phi(\bar{t}|\bar{p}, \bar{s}) = \phi(\bar{t}|\bar{p}) \tag{21}$$

$$\phi(\bar{s}|\bar{p}, \bar{t}) = \phi(\bar{s}|\bar{p}) \tag{22}$$

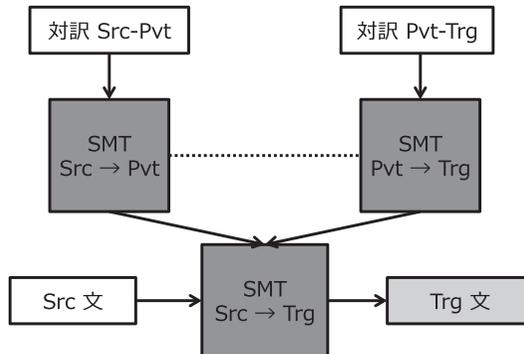


図 6 テーブル合成手法

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み合わせを求めて算出する必要があるため、大規模な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翻訳を期待できる。

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手法とテーブル合成手法によるピボット翻訳で比較実験を行った (Utiyama and Isahara 2007)。その結果、テーブル合成手法では、 $n = 1$  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 同期文脈自由文法におけるテーブル合成手法の応用

3 節で説明したピボット翻訳手法のうち、逐次的ピボット翻訳および擬似対訳コーパス手法は SMT の枠組にとらわれない手法であるため、SCFG を用いる SMT でもそのまま適用可能であるが、テーブル合成手法は本来、PBMT のフレーズテーブルを合成するために提案されたものである。SCFG を用いる翻訳方式では、式 (7) のように表現される同期導出規則をルールテーブルという形式で格納する。次節以降では、SCFG ルールテーブルを合成することで、PBMT におけるテーブル合成手法と同等のピボット翻訳を行うための手法について説明し、その後 PBMT および SCFG における複数のピボット翻訳手法による翻訳精度の差を実験によって比較評価し、考察を行う。

### 4.1 同期導出規則の合成

SCFG ルールテーブル合成手法では、まず  $Src-Pvt$ ,  $Pvt-Trg$  それぞれの言語対について、対訳コーパスを用いて同期導出規則を抽出し (2.3 節)、各規則の確率スコアなどの素性を算出してルールテーブルに格納する。その後、 $Src-Pvt$ ,  $Pvt-Trg$  ルールテーブル双方に共通の  $Pvt$  記号列を有する導出規則  $X \rightarrow \langle \bar{s}, \bar{p} \rangle$ ,  $X \rightarrow \langle \bar{p}, \bar{t} \rangle$  をすべて見つけ出し、新しい導出規則  $X \rightarrow \langle \bar{s}, \bar{t} \rangle$  の翻訳確率を、式 (17)–(20) に従って推定する。PBMT においては  $\bar{s}, \bar{p}, \bar{t}$  が各言語のフレーズ (単語列) を表しており、SCFG においては非終端記号を含む各言語の記号列を表す点で異なるが、計算式については同様である。また、 $X \rightarrow \langle \bar{s}, \bar{t} \rangle$  のワードペナルティおよびフレーズペナルティは  $X \rightarrow \langle \bar{p}, \bar{t} \rangle$  と同じ値に設定する。

本節で提案したルールテーブル合成手法によるピボット翻訳が、他の手法や他の翻訳枠組と比較して有効であるかどうかを調査するため、後述する手順によって比較実験を行った。

### 4.2 実験設定

3 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逐次的ピボット翻訳手法と、PBMT で高い実用性が示されたテーブル合成手法によるピボット翻訳を、PBMT

および SCFG において実施し、翻訳精度の比較評価を行った。

本実験では、学習および評価に用いる対訳コーパスとして、国連文書を元にして作成された国連多言語コーパス (Ziemski, Junczys-Dowmunt, and Pouliquen 2016) を用いて翻訳精度の比較評価を行った。本コーパスには、英語 (En), アラビア語 (Ar), スペイン語 (Es), フランス語 (Fr), ロシア語 (Ru), 中国語 (Zh) の 6 言語間で意味の対応する約 1,100 万文の対訳文が含まれている。これら 6 言語は複数の語族をカバーしているため、言語構造の違いにより複雑な単語並び替えが発生しやすく、SMT の枠組とピボット翻訳手法の組み合わせの影響を調査する目的に適している。現実的なピボット翻訳タスクを想定し、英語を中間言語として固定し、残りの 5 言語のすべての組み合わせでピボット翻訳を行った。ピボット翻訳では、*Src-Pvt*, *Pvt-Trg* のそれぞれの言語対の対訳を用いて *Src-Trg* の翻訳を行うが、ピボット翻訳が必要となる場面では直接的な対訳はほとんど存在しないものと想定し、それぞれの対訳の *Pvt* 側には共通の文が存在しない方が評価を行う上で望ましい。本コーパスのアーカイブには学習用データ (train) 約 1,100 万文、評価用データ (test) 4,000 文、パラメータ調整用データ (dev) 4,000 文が予め用意されているが、前処理として、それぞれのデータに対して重複して出現する英文を含む対訳文を取り除き、また、長い文は学習・評価時の計算効率上の妨げとなるため train に対して 60 単語、test, dev に対して 80 単語を超える文はすべて取り除いたところ、train は約 800 万文、test, dev はそれぞれ約 3,800 文が残った。しかし、評価対象となる組み合わせ数が膨大であるため、前処理後のデータサイズに比較すると小規模であるが、前処理後の train から *Src-Pvt* の学習用に train1, *Pvt-Trg* の学習用に train2 をそれぞれ 10 万文、英文の重複がないように取り出し、test, dev はそれぞれ 1,500 文ずつを実際的评价とパラメータ調整に用いた。

複数の言語の組み合わせで PBMT, SCFG のそれぞれについて以下のように SMT 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。

#### **Direct (直接翻訳) :**

直接的な対訳を得られる理想的な状況下における翻訳精度を得て比較を行うため、*Pvt* を用いず *Src-Trg* の直接対訳コーパス train1, train2 を個別に用いて翻訳モデルを学習し評価。train1, train2 による翻訳スコアをそれぞれ 「Direct 1」, 「Direct 2」とし、まとめて 「Direct 1 / 2」と表記

#### **Cascade (逐次的ピボット翻訳) :**

*Src-Pvt*, *Pvt-Trg* それぞれの対訳 train1, train2 で学習された翻訳モデルでパイプライン処理を行い、*Src-Trg* 翻訳を評価

#### **Triangulation (テーブル合成手法) :**

*Src-Pvt*, *Pvt-Trg* それぞれの対訳 train1, train2 で学習された翻訳モデルから、翻訳確率の推定により *Src-Trg* 翻訳モデルを合成し評価

コーパス中の中国語文は単語分割が行われていない状態であったため、KyTea (Neubig, Nakata,

and Mori 2011) の中国語モデルを用いて単語分割を行った。PBMT モデルの構築には Moses (Koehn, Hoang, Birch, Callison-Burch, Federico, Bertoldi, Cowan, Shen, Moran, Zens, Dyer, Bojar, Constantin, and Herbst 2007), SCFG 翻訳モデルの構築には Travatar (Neubig 2013) の Hiero 学習ツールを利用した。すべての翻訳システムでは KenLM (Heafield 2011) と train1+train2 の目的言語側 20 万文を用いて学習した 5-gram 言語モデルを訳出の自然性評価に用いている。また、翻訳結果の評価には、自動評価尺度 BLEU (Papineni et al. 2002) を用い、各 SMT システムについて MERT (Och 2003) により、開発用データセットに対して BLEU スコアが最大となるようにパラメータ調整を行った。

### 4.3 実験結果

様々な言語と機械翻訳方式の組み合わせについて Direct 1 / 2, Triangulation, Cascade の各ピボット翻訳手法で翻訳を行い評価した結果を表 1 に示す。太字は言語と翻訳枠組の各組み合わせで精度の高いピボット翻訳手法を示す。先行研究では、PBMT のピボット翻訳手法において Triangulation で Cascade よりも高い翻訳精度が示されており、このことは実験結果の表からも、すべての言語の組み合わせで確認できた。同様に、4.1 節で提案した SCFG ルールテーブルの Triangulation によっても、Cascade より高い翻訳精度が示された。このことから、SMT の枠組によらず、Triangulation 手法を用いることで Cascade 手法よりも安定して高いピボット翻訳精度が得られるものと考えられる。また、Triangulation の翻訳精度を Direct と比較した場合、例えばスペイン語・フランス語の Hiero 翻訳における Direct の平均 BLEU スコアが 35.34 であるのに対し、Triangulation の BLEU スコアが 32.62 と、2.72 ポイントの大きな差が開いており、Direct 翻訳で高い精度が出る言語対では Triangulation 手法でも依然として精度が大きく低下する傾向が見られた。逆に、Direct 翻訳の BLEU スコアが 15 を下回っていて翻訳が困難な言語対では、Triangulation 手法でも大きな差は見られず、フランス語・中国語の Hiero 翻訳のように、Triangulation のスコアが僅かながら Direct のスコアを上回る例も少数見られたが、誤差の範囲であろう。

一方、翻訳精度を PBMT と Hiero で比較した場合、言語対によって優劣が異なっているものの、傾向としては中国語を含む言語対において Hiero の翻訳精度が PBMT を大きく上回っており、ロシア語を含む言語対では僅かに下回り、それ以外の言語対では僅かに上回る例が多く見られた。20 組の言語対のうち 14 組で Hiero における Triangulation のスコアが PBMT の場合を上回っており、平均して 0.5 ポイント以上の BLEU スコアが向上しているため、Hiero をピボット翻訳に応用する手法は総じて有効であると考えられる。

表 1 ピボット翻訳手法毎の翻訳精度比較

Source	Target	MT Method	BLEU Score (%)		
			Direct 1 / 2	Triangulation	Cascade
Ar	Es	PBMT	28.57 / 28.69	<b>27.23</b>	26.16
		Hiero	29.12 / 29.41	<b>27.84</b>	26.78
	Fr	PBMT	23.31 / 22.00	<b>21.17</b>	20.03
		Hiero	23.68 / 22.19	<b>21.37</b>	20.76
	Ru	PBMT	16.16 / 16.22	<b>16.15</b>	14.45
		Hiero	17.28 / 16.82	<b>16.22</b>	14.71
	Zh	PBMT	12.68 / 12.88	<b>12.65</b>	10.41
		Hiero	14.51 / 14.54	<b>14.38</b>	13.85
Es	Ar	PBMT	14.27 / 14.12	<b>13.96</b>	13.18
		Hiero	14.46 / 14.17	<b>13.97</b>	13.87
	Fr	PBMT	36.03 / 34.79	<b>32.56</b>	29.66
		Hiero	35.87 / 34.81	<b>32.62</b>	29.72
	Ru	PBMT	22.17 / 21.87	<b>21.45</b>	19.89
		Hiero	21.58 / 22.18	<b>20.91</b>	19.55
	Zh	PBMT	15.03 / 14.92	<b>14.80</b>	14.04
		Hiero	17.56 / 18.10	<b>17.79</b>	17.54
Fr	Ar	PBMT	12.16 / 12.27	<b>12.18</b>	11.24
		Hiero	12.37 / 12.49	<b>12.35</b>	11.82
	Es	PBMT	38.71 / 38.85	<b>36.24</b>	32.89
		Hiero	38.34 / 38.82	<b>36.10</b>	33.16
	Ru	PBMT	20.54 / 20.53	<b>19.87</b>	18.58
		Hiero	20.10 / 20.86	<b>19.51</b>	18.33
	Zh	PBMT	13.92 / 13.96	<b>13.84</b>	13.03
		Hiero	15.99 / 15.90	<b>16.17</b>	15.73
Ru	Ar	PBMT	12.21 / 12.18	<b>11.99</b>	11.19
		Hiero	12.29 / 12.13	<b>11.75</b>	11.47
	Es	PBMT	30.99 / 30.41	<b>29.76</b>	29.10
		Hiero	31.01 / 31.39	<b>29.60</b>	28.67
	Fr	PBMT	25.32 / 25.37	<b>25.19</b>	24.10
		Hiero	25.74 / 25.71	<b>24.60</b>	23.64
	Zh	PBMT	14.66 / 15.09	<b>14.47</b>	13.40
		Hiero	15.85 / 15.75	<b>16.12</b>	15.99
Zh	Ar	PBMT	8.93 / 9.25	<b>8.96</b>	8.32
		Hiero	10.25 / 10.13	<b>9.90</b>	9.85
	Es	PBMT	21.83 / 22.49	<b>22.38</b>	20.90
		Hiero	23.55 / 23.76	<b>23.41</b>	23.10
	Fr	PBMT	17.69 / 17.77	<b>17.50</b>	16.26
		Hiero	18.83 / 18.72	<b>18.39</b>	18.15
	Ru	PBMT	13.78 / 14.12	<b>13.76</b>	12.33
		Hiero	14.96 / 15.44	<b>14.00</b>	13.92

#### 4.4 異なるデータを用いた場合の翻訳精度

4.3節までは、国連文書コーパスを用いた、複数の語族にまたがる言語間でのピボット翻訳を行った内容について説明し考察している。本研究の過程で、国連文書コーパスの他、欧州議会議事録を元にした Europarl コーパス (Koehn 2005) を用いて 4.2 節と同様の実験も実施している。このコーパスは、欧州の諸言語を広くカバーしており、多言語翻訳タスクに多用されるが、語族がほとんど共通しており比較的似通った言語間での翻訳となる。この実験では英語を中間言語として固定し、欧州でも話者数の多いドイツ語、スペイン語、フランス語、イタリア語の4言語の組み合わせでピボット翻訳を行った。Europarl から 10 万文の対訳で学習し、1,500 文ずつの評価とパラメータ調整を行ったが、*Src-Pvt* と *Pvt-Trg* それぞれの翻訳モデルの学習にはすべて中間言語データが一致しているものを用い、直接的な翻訳と比較してどの程度精度に影響があるかも調査した。この実験結果から、4.3 節の実験結果と同様に、PBMT と Hiero の双方において、すべての言語対において Triangulation が Cascade よりも高精度となることが確認された。また、2つの翻訳モデルの学習で中間言語側が共通のデータを用いているにも関わらず、Triangulation の精度は Direct と比較して大きく減少しており、この精度差が中間言語側の曖昧性の影響を強く受けて発生したものであると考えられる。一方、PBMT と Hiero の精度を比較した場合には、精度差が言語対に依存するのと同様であり、大きな単語並び替えが発生しないような言語の組み合わせが多いため、計算コストが低く、標準設定でより長いフレーズ対応を学習できる PBMT の方が有利と考えられる点も多かった。

#### 4.5 考察および関連研究

3節では、PBMT で提案されてきた代表的なピボット翻訳手法について説明し、本節ではテーブル合成手法を SCFG のルールテーブルに適用するための手法について述べ、また言語対・機械翻訳方式・ピボット翻訳手法の組み合わせによって翻訳精度の影響を比較評価した。その結果、SCFG においてもテーブル合成手法によって高い翻訳精度を得られることが示され、また言語対や用いるデータによっては PBMT の場合よりも高い精度が得られることも分かった。ピボット翻訳におけるその他の関連研究は、PBMT のテーブル合成手法をベースに、さらに精度を上げるための議論が中心である。テーブル合成手法はピボット翻訳手法の中でも高い翻訳精度が報告されているが (Utiyama and Isahara 2007)、1節で述べたような中間言語側の表現力に起因する曖昧性の問題や、異なる言語対やデータセット上で推定された単語アラインメントから抽出されるフレーズの不一致によって、得られる翻訳規則数が減少する問題などがあり、直接的な対訳が得られる理想的な状況と比較すると翻訳精度が大きく下回ってしまう。これらの問題に対処するための関連研究として、翻訳確率推定の前にフレーズの共起頻度を推定することでサイズが不均衡なテーブルの合成を改善する手法 (Zhu et al. 2014)、単語の分散表現を用いて単語レベルの翻訳確率を補正する手法 (Levinboim and Chiang 2015)、複数の中間言語を用い

る手法 (Dabre, Cromieres, Kurohashi, and Bhattacharyya 2015) などが挙げられ、曖昧性の解消には中間表現の工夫と信頼度の高い言語資源の有効利用が必要と言える。

## 5 中間言語情報を記憶するピボット翻訳手法の提案

3節ではSMTで用いられているピボット翻訳手法について紹介し、4節では従来手法の中で高い翻訳精度が報告されているテーブル合成手法をSCFGで応用するための手順について説明した。また、比較評価実験により、SCFGにおいてもPBMTと同様、テーブル合成手法によって逐次的ピボット翻訳手法よりも高い精度が得られた。しかし、直接の対訳を用いて学習した場合と比較すると、翻訳精度の差は未だ大きいため、精度が損なわれてしまう原因を特定し、解消することができれば、さらなる翻訳精度の向上が期待できる。テーブル合成手法で翻訳精度が損なわれる原因の一つとして、翻訳時に重要な手がかりとなるはずの中間言語の情報はテーブル合成後には失われてしまい、不正確に推定された*Src-Trg*のフレーズ対応と翻訳確率のみが残る点が挙げられる。本節では、従来では消失してしまう中間言語情報を記憶し、この追加の情報を翻訳時に用いることで精度向上に役立てる、新しいテーブル合成手法を提案する。

### 5.1 従来のテーブル合成手法の問題点

従来のテーブル合成手法の問題点について、1節中でも紹介したが、本節で改めて説明を行う。テーブル合成手法では、*Src-Pvt*、*Pvt-Trg*それぞれの言語対におけるフレーズの対応と翻訳確率のスコアが与えられており、この情報を元に、*Src-Trg*言語対におけるフレーズ対応と翻訳確率の推定を行う。ところが、語義曖昧性や言語間の用語法の差異により、*Src-Trg*のフレーズ対応を正確に推定することは困難である。

図7はテーブル合成手法によって対応を推定するフレーズの例を示しており、図中では日本語とイタリア語それぞれにおける3つの単語が、語義曖昧性を持つ英単語「approach」に結び付いている。このような場合、*Src-Trg*のフレーズ対応を求め、適切な翻訳確率推定を行うのは複雑な問題となってくる。その上、図8に示すように、従来のテーブル合成手法では、合成時に*Src*と*Trg*の橋渡しをしていた*Pvt*フレーズの情報が、合成後には保存されず失われてしまう。現実の人手翻訳の場合を考えても、現在着目しているフレーズに関する追加の言語情報が

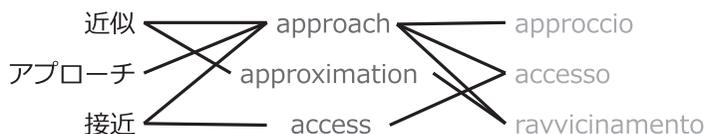


図7 モデル学習に用いるフレーズ対応 (日-英-伊)

与えられているなら、その言語を知る者にとって重要な手がかりとなって曖昧性解消などに用いることができる。そのため、*Src-Trg* を結び付ける *Pvt* フレーズは重要な言語情報であると考えられ、本研究では、この情報を保存することで機械翻訳にも役立てるための手法を提案する。

### 5.2 中間言語情報を記憶するテーブル合成手法

前節で述べた問題を克服するため、本研究では *Src* と *Trg* を結び付けていた *Pvt* フレーズの情報も翻訳モデル中に保存し、*Src* から *Trg* と *Pvt* への同時翻訳確率を推定することによって翻訳を行う新しいテーブル合成手法を提案する。図 9 に、本提案手法によって得られるフレーズ対応の例を示す。本手法の利点は、英語のように中間言語として選ばれる言語は豊富な単言語資源も得られる傾向が強いため、このような追加の言語情報を翻訳システムに組み込み、翻訳文の導出時に利用できることにある。

中間言語フレーズの情報を翻訳時に役立てるため、SCFG (2.3 節) を複数の目的言語文の同時生成に対応できるよう拡張した MSCFG (2.4 節) を用いて翻訳モデルの合成を行う。MSCFG による翻訳モデルを構築するためには、*Src-Pvt*, *Pvt-Trg* の SCFG 翻訳規則が格納されたルールテーブルを元に、SCFG ルールテーブルとしてではなく、*Src-Trg-Pvt* の MSCFG ルールテーブルとして合成し、これによって *Pvt* フレーズを記憶する。訳出候補の探索時には、生成文の



図 8 従来手法によって得られるフレーズ対応



図 9 提案手法によって得られるフレーズ対応

自然性を評価し、適切な語彙選択を促すために言語モデルを用いるが、目的言語モデルのみでなく、中間言語モデルも同時に用いた探索を行う。次節から、SCFG 翻訳モデルの同期導出規則から MSCFG 翻訳モデルの複数同期導出規則を合成するための手順について説明する。

### 5.3 同期導出規則から複数同期導出規則への合成

4.1 節では、SCFG の同期規則を合成するために、*Src-Pvt*, *Pvt-Trg* ルールテーブル双方に共通の *Pvt* 記号列を有する導出規則  $X \rightarrow \langle \bar{s}, \bar{p} \rangle$ ,  $X \rightarrow \langle \bar{p}, \bar{t} \rangle$  を見つけ出し、新しい導出規則  $X \rightarrow \langle \bar{s}, \bar{t} \rangle$  の翻訳確率を、式 (17)–(20) に従って確率周辺化を行い推定することを述べた。一方、中間言語情報を記憶するテーブル合成手法では  $X \rightarrow \langle \bar{s}, \bar{p} \rangle$ ,  $X \rightarrow \langle \bar{p}, \bar{t} \rangle$  を元に、以下のよう複数同期規則を合成する。

$$X \longrightarrow \langle \bar{s}, \bar{t}, \bar{p} \rangle \quad (23)$$

このような規則を用いて翻訳を行うことによって、同時生成される中間言語文を通じて中間言語モデルなどのような追加の素性を取り入れることが可能となる。式 (17)–(20) に加えて、*Trg* と *Pvt* を同時に考慮した翻訳確率  $\phi(\bar{t}, \bar{p}|\bar{s})$ ,  $\phi(\bar{s}|\bar{p}, \bar{t})$  を以下のように推定する。

$$\phi(\bar{t}, \bar{p}|\bar{s}) = \phi(\bar{t}|\bar{p}) \phi(\bar{p}|\bar{s}) \quad (24)$$

$$\phi(\bar{s}|\bar{p}, \bar{t}) = \phi(\bar{s}|\bar{p}) \quad (25)$$

*Src-Pvt* の翻訳確率  $\phi(\bar{p}|\bar{s})$ ,  $\phi(\bar{s}|\bar{p})$ ,  $\phi_{lex}(\bar{p}|\bar{s})$ ,  $\phi_{lex}(\bar{s}|\bar{p})$  はルールテーブル  $T_{SP}$  のスコアをそのまま用いることが可能である。これら 10 個の翻訳確率  $\phi(\bar{t}|\bar{s})$ ,  $\phi(\bar{s}|\bar{t})$ ,  $\phi(\bar{p}|\bar{s})$ ,  $\phi(\bar{s}|\bar{p})$ ,  $\phi_{lex}(\bar{t}|\bar{s})$ ,  $\phi_{lex}(\bar{s}|\bar{t})$ ,  $\phi_{lex}(\bar{p}|\bar{s})$ ,  $\phi_{lex}(\bar{s}|\bar{p})$ ,  $\phi(\bar{t}, \bar{p}|\bar{s})$ ,  $\phi(\bar{s}|\bar{p}, \bar{t})$  に加えて、 $\bar{t}$  と  $\bar{p}$  に含まれる非終端記号数を 2 つのワードペナルティとし、定数 1 のフレーズペナルティの、合わせて 13 個のスコアが MSCFG ルールにおける素性となる。

### 5.4 同期規則のフィルタリング

前節で説明した、中間言語情報を記憶するテーブル合成手法は、このままでは  $\langle \bar{s}, \bar{t} \rangle$  ではなく、 $\langle \bar{s}, \bar{t}, \bar{p} \rangle$  の全組み合わせを記録するため、従来より大きなルールテーブルが合成されてしまう。計算資源を節約するためには、幾つかのフィルタリング手法が考えられる。Neubig らによると、主要な目的言語  $T_1$  と補助的な目的言語  $T_2$  で翻訳を行う際には、 $T_1$ -フィルタリング手法 (Neubig et al. 2015) が効果的である。このフィルタリング手法を、提案するテーブル合成手法に当てはめると、 $T_1 = Trg$ ,  $T_2 = Pvt$  であり、原言語フレーズ  $\bar{s}$  に対して、まず *Trg* において  $\phi(\bar{t}|\bar{s})$  が上位  $L$  個までの  $\bar{t}$  を残し、それぞれの  $\bar{t}$  に対して  $\phi(\bar{t}, \bar{p}|\bar{s})$  が最大となるような  $\bar{p}$  を残す。

## 6 実験的評価

前節で提案した中間言語情報を記憶するテーブル合成手法の有効性を検証するため、多言語コーパスを用いたピボット翻訳の比較評価実験を実施した。

### 6.1 実験設定

用いたデータやツールは、4.2節の実験と大部分が共通しているため、差分を明らかにしつつ説明を行う。本実験では、4.2節の実験で得られた同じ対訳データを用いて各翻訳モデルの学習に用いた。すなわち、国連多言語コーパスを用いて、英語 (En) を中間言語とするアラビア語 (Ar)、スペイン語 (Es)、フランス語 (Fr)、ロシア語 (Ru)、中国語 (Zh) の5言語の組み合わせでピボット翻訳の翻訳精度を比較し、それぞれの言語対について *Src-Pvt* 翻訳モデルの学習用 (train1) に10万文、*Pvt-Trg* 翻訳モデルの学習用 (train2) に10万文、評価 (test) とパラメータ調整用 (dev) にそれぞれ1,500文ずつを用いた。目的言語モデルの学習にも、4.2節と同様に train1+train2 の目的言語文20万文を用いている。また、多くの場合、英語においては大規模な単言語資源が取得可能であるため、最大500万文までのデータを用いて段階的に学習を行った複数の中間言語モデルを用意した。SCFG および MSCFG を用いるデコーダとして Travatar (Neubig 2013) を使い、付属の Hiero ルール抽出プログラムを用いて SCFG 翻訳モデルの学習を行った。翻訳結果の比較には、自動評価尺度 BLEU (Papineni et al. 2002) を使い、各翻訳モデルは MERT (Och 2003) により、開発用データに対して BLEU スコアが最大となるようにパラメータ調整を行った。提案手法のテーブル合成手法によって得られた MSCFG ルールテーブルは、 $L = 20$  の  $T_1$ -フィルタリング手法によって枝刈りを行った。本実験ではまず、以下の6つの翻訳手法を比較評価する。

**Cascade** (逐次的ピボット翻訳) :

*Src-Pvt* および *Pvt-Trg* の SCFG 翻訳モデルで逐次的ピボット翻訳 (3.1節)。「w/ PvtLM 200k/5M」は *Src-Pvt* 翻訳時にそれぞれ20万文、500万文で学習した中間言語モデルを用いることを示す

**Tri. SCFG** (SCFG ルールテーブル合成手法) :

*Src-Pvt* および *Pvt-Trg* の SCFG モデルを合成し、*Src-Trg* の SCFG モデルによって合成 (3.3節, ベースライン)

**Tri. MSCFG** (MSCFG ルールテーブル合成手法) :

*Src-Pvt* および *Pvt-Trg* の SCFG モデルを合成し、*Src-Trg-Pvt* の MSCFG モデルによって翻訳 (5節)。「w/o PvtLM」は中間言語モデルを用いないことを示し、「w/ PvtLM 200k/5M」はそれぞれ20万文、500万文で学習した中間言語モデルを用いることを示す

## 6.2 翻訳精度の比較

表2に、英語を介したすべての言語対におけるピボット翻訳の結果を示す。実験で得られた結果は、ブートストラップ・リサンプリング法 (Koehn 2004) により統計的有意差を検証した。それぞれの言語対において、太字は従来手法 Tri. SCFG よりも BLEU スコアが高いことを示し、下線は最高スコアを示す。短剣符は各ピボット翻訳手法の翻訳精度が Tri. SCFG よりも統計的有意に高いことを示す ( $\dagger: p < 0.05, \ddagger: p < 0.01$ )。評価値から、提案したテーブル合成手法で中間言語モデルを考慮した翻訳を行った場合、すべての言語対において従来のテーブル合成手法よりも BLEU スコアの向上が確認できる。すべての組み合わせにおいて、テーブル合成手法で中間言語情報を記憶し、500 万文の言語モデルを考慮して翻訳を行った場合に最も高いスコアを達成しており、従来法に比べ最大で 1.8、平均で 0.75 ほどの BLEU 値の向上が見られる。このことから、中間言語情報を記憶し、これを翻訳に利用することが曖昧性の解消に繋がり、安定して翻訳精度を改善できたと言えよう。

また、異なる要因による影響を切り分けて調査するため、MSCFG へ合成するが、中間言語モデルを用いずに翻訳を行った場合の比較も行った (Tri. MSCFG w/o PvtLM)。この場合、保

表 2 ピボット翻訳手法と中間言語モデル規模の組み合わせによる翻訳精度比較

Src	Trg	BLEU Score (%)					
		Cascade w/ PvtLM 200k	Cascade w/ PvtLM 5M	Tri. SCFG (baseline)	Tri. MSCFG w/o PvtLM	Tri. MSCFG w/ PvtLM 200k	Tri. MSCFG w/ PvtLM 5M
Ar	Es	26.78	$\ddagger$ <b>28.52</b>	27.84	$\ddagger$ <b>28.44</b>	$\ddagger$ <b>28.48</b>	$\ddagger$ <b>29.13</b>
	Fr	20.76	$\ddagger$ <b>22.02</b>	21.37	$\ddagger$ <b>21.94</b>	$\ddagger$ <b>22.01</b>	$\ddagger$ <b>22.52</b>
	Ru	14.71	15.91	16.22	<b>16.38</b>	$\dagger$ <b>16.61</b>	$\ddagger$ <b>16.76</b>
	Zh	13.85	$\ddagger$ <b>15.01</b>	14.38	<b>14.43</b>	$\ddagger$ <b>14.93</b>	$\ddagger$ <b>15.50</b>
Es	Ar	13.87	$\dagger$ <b>14.35</b>	13.97	<b>14.19</b>	$\dagger$ <b>14.34</b>	$\ddagger$ <b>14.42</b>
	Fr	29.72	30.20	32.62	<b>32.70</b>	<b>32.81</b>	$\dagger$ <b>32.94</b>
	Ru	19.55	20.52	20.91	<b>20.92</b>	<b>20.95</b>	$\ddagger$ <b>21.52</b>
	Zh	17.54	$\ddagger$ <b>18.37</b>	17.79	17.56	$\dagger$ <b>18.13</b>	$\ddagger$ <b>18.70</b>
Fr	Ar	11.82	11.96	12.35	12.00	12.35	$\ddagger$ <b>12.71</b>
	Es	33.16	33.90	36.10	36.00	<b>36.33</b>	$\ddagger$ <b>36.93</b>
	Ru	18.33	18.95	19.51	$\dagger$ <b>19.91</b>	$\ddagger$ <b>20.23</b>	$\ddagger$ <b>20.44</b>
	Zh	15.73	<b>16.35</b>	16.17	<b>16.20</b>	<b>16.40</b>	$\dagger$ <b>16.58</b>
Ru	Ar	11.47	11.72	11.75	<b>11.84</b>	<b>11.91</b>	$\ddagger$ <b>12.18</b>
	Es	28.67	$\ddagger$ <b>30.23</b>	29.60	$\dagger$ <b>29.90</b>	$\ddagger$ <b>30.59</b>	$\ddagger$ <b>31.40</b>
	Fr	23.64	$\dagger$ <b>25.09</b>	24.60	<b>24.63</b>	$\ddagger$ <b>25.20</b>	$\ddagger$ <b>25.32</b>
	Zh	15.99	$\dagger$ <b>16.68</b>	16.12	15.75	$\ddagger$ <b>16.66</b>	$\ddagger$ <b>16.81</b>
Zh	Ar	9.85	9.79	9.90	<b>9.98</b>	<b>10.04</b>	<b>10.07</b>
	Es	23.10	$\dagger$ <b>23.98</b>	23.41	<b>23.62</b>	$\dagger$ <b>23.83</b>	$\ddagger$ <b>24.45</b>
	Fr	18.15	<b>18.45</b>	18.39	18.30	<b>18.51</b>	$\ddagger$ <b>19.16</b>
	Ru	13.92	13.65	14.00	<b>14.24</b>	$\ddagger$ <b>14.41</b>	$\ddagger$ <b>14.67</b>

存された中間言語情報が語彙選択に活用されないため、本手法の優位性は特に現れないものと予想できたが、実際には、SCFGに合成する場合よりも多くの言語対で僅かに高い翻訳精度が見られた。これは、追加の翻訳確率などのスコアが有効な素性として働き、パラメータ調整を行った上で、適切な語彙選択に繋がったことなどが原因として考えられる。

大規模な中間言語モデルを用いる手法は、本稿で提案するテーブル合成後のモデルのみならず、従来の逐次のピボット翻訳手法でも可能であるため、500万文の大規模な中間言語モデルを用いた場合の精度評価も行った (Cascade w/ PvtLM 5M)。Cascade w/ PvtLM 5Mは20万文の中間言語モデルしか用いない逐次のピボット翻訳手法 (Cascade w/ PvtLM 200k)と比較した場合にZh-Ar, Zh-Ruを除いたすべての言語対で高精度であり、単純に大規模な言語モデルを用いることで精度向上に繋がることは確認された。しかし、従来のテーブル合成手法であるTri. SCFGと比較した場合の精度差は言語対依存であり、安定した精度向上とはならなかった。また、本稿の提案手法 Tri. MSCFG w/ PvtLM 5Mと、Cascade w/ PvtLM 5Mを比較すると、同じ規模の中間言語モデルを用いてもすべての言語対で提案手法の方が高精度であり、このことからテーブル合成手法と大規模な中間言語モデルを用いることの有効性が高いと言えるだろう。

### 6.3 中間言語モデルの規模が翻訳精度に与える影響

中間言語モデルの規模がピボット翻訳精度に与える影響の大きさは言語対によって異なっているが、中間言語モデルの学習データサイズが大きくなるほど精度が向上することも確認できる。図10は、中国語・スペイン語 (左) およびアラビア語・ロシア語 (右) のピボット翻訳において異なるデータサイズで学習した中間言語モデルが翻訳精度に与える影響を示す。図からも中間言語モデルが曖昧性を解消して翻訳精度向上に寄与している様子が確認できる。中間

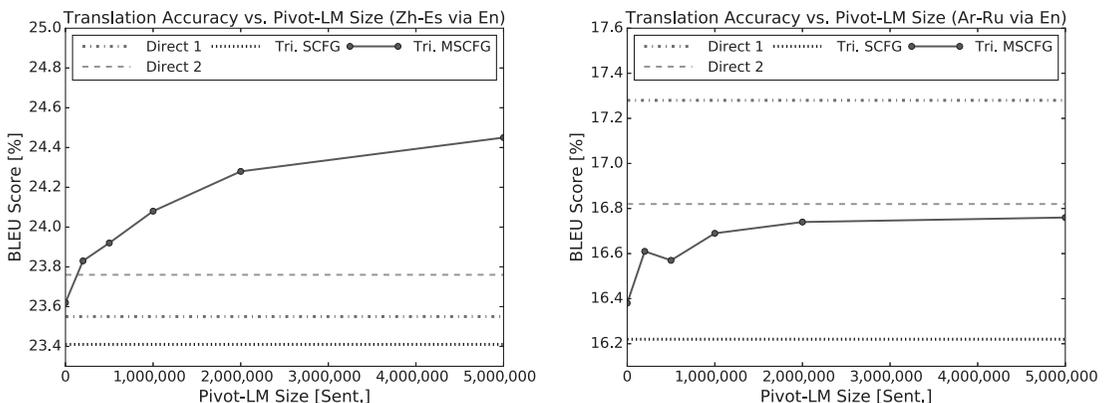


図 10 中間言語モデル規模がピボット翻訳精度に与える影響

言語モデルの学習データサイズを増加させることによる翻訳精度への影響は対数的であることも見てとられるが、これは目的言語モデルサイズが翻訳精度へ与える影響と同様の傾向である (Brants, Popat, Xu, Och, and Dean 2007). 中国語・スペイン語の翻訳ではグラフの傾向から、さらに大規模な中間言語モデルを用いることで精度向上の見込みがあるが、一方でアラビア語・ロシア語の場合には学習データサイズが200万文から500万文に増加しても精度にほとんど影響が見られないため、これ以上の精度向上には限界があると考えられる。

#### 6.4 曖昧性が解消された例と未解決の問題

本提案手法によって中間言語側で曖昧性が解消されて翻訳精度向上に繋がったと考えられる訳出の例を示す。

入力文 (フランス語) :

Le nom du candidat **proposé** est indiqué dans l'annexa à la présente note .

参照訳 (スペイン語) :

El nombre del candidato **propuesto** se presenta en el anexo de la presente nota .

対応する英文 :

The name of the candidate **thus nominated** is set out in the annex to the present note .

**Tri. SCFG:**

El nombre del **proyecto** de un candidato se indica en el anexo a la presente nota .

(BLEU+1: 34.99)

**Tri. MSCFG w/ PvtLM 5M:**

El nombre del candidato **propuesto** se indica en el anexo a la presente nota . (BLEU+1: 61.13)

The name of the candidate **proposed** indicated in the annex to the present note .

(同時生成された英文)

上記の例では、入力文中のフランス語の分詞「proposé (指名された)」には参照役中のスペイン語の分詞「propuesto」が対応しているが、従来のテーブル合成手法では、誤った対応である名詞「proyecto (計画, 立案)」が結び付き翻訳に用いられた結果、不正確な訳出となっている。一方で提案手法においては、入力文の「proposé」に対してスペイン語の「propuesto」と英語の「proposed」が同時に結び付いており、生成される英文中の単語の前後関係から適切な語彙選択を促し、訳出の改善に繋がったものと考えられる。

逆に、提案手法では語彙選択がうまくいかず、直接対訳で学習した場合よりも精度が落ちた訳出の例を示す。

入力文 (フランス語) :

J . Risques d'**aspiration** : **citère** de viscosité pour la classification des **mélanges** ;

参照訳 (スペイン語) :

J . Peligros por **aspiración** : **criterio** de viscosidad para la clasificación de **mezclas** ;

対応する英文 :

J . **Aspiration** hazards : viscosity **criterion** for classification of **mixtures** ;

**Direct 1:**

J. Riesgos d'aspiration : **criterio** de viscosité para la clasificación de **los mélanges** ; (BLEU+1: 34.20)

**Direct 2:**

J. Riesgos d'aspiration : **criterio** de viscosité para la clasificación de **mezclas** ; (BLEU+1: 49.16)

**Tri. MSCFG w/ PvtLM 2M:**

J. Riesgos d'aspiration : viscosité **criterios** para la clasificación de **mélanges** ; (BLEU+1: 27.61)

J. Risk d'aspiration : viscosité **criteria** for the categorization of **mélanges** ; (同時生成された英文)

この例では、フランス語の「d'aspiration (吸引)」や「mélanges (混合物)」といった専門用語は、コーパス中の出現頻度が少なく、「d'aspiration」は train1 や train2 にも一度も出現しないため、Direct 1 / 2 の双方で翻訳不可能で未知語扱いとなっており、「mélanges」は train2 でのみ出現しており、Direct 1 では未知語扱いである。テーブル合成手法では、2つの翻訳モデルで共通して出現する中間言語フレーズのみしか学習できないため、他方のみには含まれない専門用語は未知語となってしまう。この問題は、その他のピボット翻訳手法である逐次的ピボット翻訳手法や擬似コーパス手法でも当然解決不可能なため、複数の対訳データでカバーできない表現は外部辞書などを用いて補う必要があるだろう。

また、この例では未知語の問題以外にもスペイン語の単数形の名詞「criterio (基準)」がテーブル合成手法では複数形の「criterios」となっていたり語順が誤ったりしている問題も見られる。こういった問題は、本提案手法である程度は改善されているものの、活用形や語順の問題により正確に対処するためには、統語的情報を明示的に扱う手法の導入が必要であると考えられる。

## 6.5 Europarl を用いた評価および品詞毎の翻訳精度

提案手法の有効性を調査するための比較評価実験も、国連文書コーパスのみでなく、研究の過程で 4.4 節と同様に Europarl を用いた欧州の言語間でのピボット翻訳においても実施した。この実験においても 10 万文の対訳データを用いて学習した翻訳モデルを合成するが、中間言語モデルの学習には最大 200 万文までの英文を利用した。この実験からも、中間言語情報を記憶するテーブル合成手法と 200 万文で学習した中間言語モデルを用いた場合に、すべての言語対において従来のピボット翻訳手法である Tri. SCFG や Cascade を上回る精度が得られた。このことから、本提案手法は言語構造の類似度に関わらず有効に機能するものと考えられる。

また、英語は他の欧州諸言語と比較して、性・数・格に応じた活用などが簡略化された言語として有名であり、語形から統語情報が失われることで発生する曖昧性の問題もある。本節では、英語を介したドイツ語・フランス語の両方向の翻訳において、誤りの発生しやすい品詞について調査する。まず、独仏・仏独翻訳における評価データの参照訳および各ピボット翻訳手法の翻訳結果に対して Stanford POS Tagger (Toutanova and Manning 2000; Toutanova, Klein, Manning, and Singer 2003) を用いて品詞付与を行い、参照訳と翻訳結果を比較して、語順は考慮せずに適合率と再現率の調和平均である F 値を算出した。

表3および表4は、各翻訳における高頻出品詞の正解率を表している。出現頻度は、参照訳中の各品詞の出現回数を意味し、丸括弧内に示された数値は、提案手法とベースライン手法のF値の差分である。結果は言語対依存であるが、特に目的言語に強く依存していることが明らかである。

表3の独仏翻訳の例では、提案手法によって、ベースライン手法よりも特に前置詞、定冠詞、動詞でF値が大きく向上している。一般名詞、形容詞、動詞、副詞は重要な内容語であり、語彙選択の幅も広いため、どの手法でも全体的にF値が低くなっている。一般名詞に関しては、通常のテーブル合成手法でもDirectと大きな差は出しておらず、そのため提案手法でもほとんど改善されなかった。一方で、動詞のF値は提案手法で大きく向上しており、中間言語モデルによって語の並びを考慮して語彙選択を行うことで翻訳精度向上に繋がったと考えられる。しかし、それでもDirectには大きく及ばず、頻度が比較的低い内容語の語彙選択を適切に行うことは困難と言えるだろう。一方で、機能語においては提案手法においてDirectと近いF値となった。

表3 独仏翻訳における品詞毎の翻訳精度

品詞	出現頻度	F-Measure (%)		
		Direct	Tri. MSCFG w/ PvtLM 2M	Tri.SCFG
NC (一般名詞)	8,360	78.94	78.19 (+0.01)	78.18
P (前置詞)	6,553	92.58	92.46 (+1.02)	91.44
DET (定冠詞)	5,363	91.04	90.63 (+1.50)	89.13
PUNC (区切り記号)	4,258	93.89	93.77 (+0.44)	93.33
ADJ (形容詞)	3,725	71.36	69.85 (+0.67)	69.18
V (動詞)	3,324	69.52	63.86 (+1.06)	62.80
ADV (副詞)	2,238	81.98	77.81 (+0.34)	77.47

表4 仏独翻訳における品詞毎の翻訳精度

品詞	出現頻度	F-Measure (%)		
		Direct	Tri. MSCFG w/ PvtLM 2M	Tri.SCFG
NN (一般名詞)	10,813	78.04	75.42 (+0.53)	74.89
ART (冠詞)	4,502	90.38	87.21 (+1.24)	85.97
CARD (数詞)	3,849	94.56	94.38 (-0.44)	94.82
APPR (前置詞)	2,577	89.66	84.73 (+1.50)	83.23
ADJA (形容詞)	2,311	65.96	65.29 (+0.47)	64.82
NE (固有名詞)	2,005	77.52	75.34 (+0.47)	74.87
ADV (副詞)	1,848	74.32	70.13 (+0.18)	69.95
PPEF (再帰代名詞)	1,665	92.64	87.46 (+1.67)	85.79

表4の仏独翻訳の例では、表3の場合と少し変わっており、冠詞、前置詞、再帰代名詞のような機能語で、提案手法によってF値が大きく改善されているものの、Directと比べると大きな差があり、ピボット翻訳で精度が大きく落ちる原因と考えられる。冠詞や前置詞は機能語であるものの、ドイツ語では男性系・女性系に加えて、英語にもフランス語にもない中性系の活用を持っているため、フランス語よりも活用の種類が多く、また冠詞や前置詞も格に応じた活用をすることが知られている。原言語を英語に翻訳した際には統語的情報が失われることが多いため、機能語にも活用幅があるような目的言語に対してはピボット翻訳が特に困難になることが多い。

本節では、品詞毎の単語正解率の分析を行ったが、言語対毎に特定品詞の翻訳精度が落ちる現象が見られた。英語には同じ語形で多品詞の語が多いため、単語の対応だけで統語的な役割を判断するのは不可能な場合もある。統語的な情報を汲み取るには複数の単語からなるフレーズを考慮する必要があるが、文頭と文末のような離れた位置での依存関係も存在するため、単語列としてではなく構文構造を考慮することも重要と考えられる。そこで、本研究の今後の課題として、構文構造を中間言語の表現として用いるピボット翻訳手法を検討している。

## 7 まとめ

本研究の目的は、多言語機械翻訳における翻訳精度の向上を目指し、従来のピボット翻訳手法を調査、問題点を改善して翻訳精度を高めることであった。そのために、PBMTで既に有効性が示されている、テーブル合成手法によるピボット翻訳をSCFGに適用し、どのような処理が有効であるかに着目した。さらに、従来のテーブル合成手法では中間言語情報が失われ、曖昧性により翻訳精度が減少する問題に対処するため、中間言語情報を記憶し、中間言語モデルを利用して自然な語順の語彙選択を促すことで精度を向上させるテーブル合成手法についても提案した。本論文で提案した手法を用いて、国連文書を元にした多言語コーパスの6言語のデータを用いてピボット翻訳の比較評価を行ったところ、英語を中間言語としたすべての言語の組み合わせで従来のテーブル合成手法よりも高い翻訳精度が示された。また、特に大規模な中間言語モデルを用いることで、より適切な語彙選択が促されて翻訳の質を高められることが分かった。しかし、提案手法でも、直接の対訳コーパスを用いて学習を行った理想的な状況と比較すると、精度の開きが大きく、中間言語モデルをさらに大きくするだけでは解決できないであろう点も示唆された。本提案手法で解決できなかった曖昧性の問題を調査すべく、特定の言語対で品詞毎の単語正解率を求めたところ、語順を考慮するだけでは解消されない曖昧性の問題もあり、構文情報を用いてこの点を改善することが今後の課題として考えられる。

ピボット翻訳の曖昧性の問題は、主として中間言語の表現力に起因しており、中間言語の単語列だけでは原言語の情報が失われてしまい、目的言語側で確率的に正しく再現することは困

難である。そのため、今後の課題として、中間言語を特定の言語の単語列としてではなく、より高い表現力を持った構文構造を中間表現とすることで、中間言語側の多品詞語の問題に対処したり、原言語側の情報を保存して、より正確に目的言語側で情報を再現するための手法を検討する。

1つ目は、中間表現に統語情報を用いた翻訳規則テーブルの軽量化・高精度化である。本研究で提案した手法では、SCFGの翻訳モデルを学習するために、階層的フレーズベース翻訳という枠組の翻訳手法で翻訳規則を獲得している。これは、翻訳において重要な、単語並び替え問題を高精度に対処できる点で優れているが、統語情報を用いず、総当り的な手段で非終端記号を含んだ翻訳規則を学習するため、テーブルサイズが肥大化する傾向がある。その上、テーブル合成時には、中間表現の一致する組み合わせによってテーブルサイズはさらに増加する。これは、曖昧性により多くの誤ったフレーズ対応も保存されることを意味するため、不要な規則は除去してサイズを削減するべきである。このような、中間表現が一致する組み合わせの中には、文法上の役割は異なるが表記上同じようなものも含まれるため、意味の対応しない組み合わせに対して高い翻訳確率が推定されてしまう場合もある。こういった問題は、中間言語の表現に統語情報を組み込むことで、品詞や句構造が異なればフレーズの対応も結び付かないという制約が働き、誤った句対応を容易に除外できるため、テーブルサイズは減少し、曖昧性が解消されて翻訳精度の向上が期待できる。

2つ目は、中間表現に原言語の統語情報を保存するピボット翻訳手法の提案である。ピボット翻訳においては、中間言語の表現力が悪影響を及ぼして、原言語の情報が失われてしまうことを述べてきたが、これは機械翻訳に限らず、人手による翻訳でも度々起こる問題である。例えば、英語には人称接尾辞のような活用体系がないため、英語に訳した際に性・数・格などの統語的情報が失われ、結果的に英語を元にした翻訳では原意と大きく異なってしまう現象などがある。本枠組では、前述の中間言語側の統語情報と組み合わせることで、より原意を汲んだ翻訳の実現を目指す。

## 謝 辞

本研究の一部はJSPS 科研費 16H05873, 24240032 および ATR-Trek 社共同研究の助成を受けて実施されました。

## 参考文献

Brants, T., Popat, A. C., Xu, P., Och, F. J., and Dean, J. (2007). “Large Language Models in Machine Translation.” In *Proceedings EMNLP*, pp. 858–867.

- Brown, P. F., Pietra, V. J., Pietra, S. A. D., and Mercer, R. L. (1993). “The Mathematics of Statistical Machine Translation: Parameter Estimation.” *Computational Linguistics*, **19**, pp. 263–312.
- Chappelier, J.-C. and Rajman, M. (1998). “A Generalized CYK Algorithm for Parsing Stochastic CFG.” In *Proceedings TAPD*, Vol. 98, No. 5, pp. 133–137.
- Chiang, D. (2007). “Hierarchical Phrase-based Translation.” *Computational Linguistics*, **33** (2), pp. 201–228.
- Cohn, T. and Lapata, M. (2007). “Machine Translation by Triangulation: Making Effective Use of Multi-Parallel Corpora.” In *Proceedings ACL*, pp. 728–735.
- Dabre, R., Cromieres, F., Kurohashi, S., and Bhattacharyya, P. (2015). “Leveraging Small Multilingual Corpora for SMT Using Many Pivot Languages.” In *Proceedings NAACL*, pp. 1192–1202.
- de Gispert, A. and Mariño, J. B. (2006). “Catalan-English Statistical Machine Translation without Parallel Corpus: Bridging through Spanish.” In *Proceedings of LREC 5th Workshop on Strategies for Developing Machine Translation for Minority Languages*, pp. 65–68.
- Dyer, C., Cordova, A., Mont, A., and Lin, J. (2008). “Fast, Easy, and Cheap: Construction of Statistical Machine Translation Models with MapReduce.” In *Proceedings WMT*, pp. 199–207.
- Goto, I., Utiyama, M., Sumita, E., Tamura, A., and Kurohashi, S. (2013). “Distortion Model Considering Rich Context for Statistical Machine Translation.” In *Proceedings ACL*, pp. 155–165.
- Heafield, K. (2011). “KenLM: Faster and Smaller Language Model Queries.” In *Proceedings, WMT*, pp. 187–197.
- Koehn, P. (2004). “Statistical Significance Tests for Machine Translation Evaluation.” In Lin, D. and Wu, D. (Eds.), *Proceedings EMNLP*, pp. 388–395.
- Koehn, P. (2005). “Europarl: A Parallel Corpus for Statistical Machine Translation.” In *MT Summit*, Vol. 5, pp. 79–86.
- Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). “Moses: Open Source Toolkit for Statistical Machine Translation.” In *Proceedings ACL*, pp. 177–180.
- Koehn, P., Och, F. J., and Marcu, D. (2003). “Statistical Phrase-Based Translation.” In *Proceedings NAACL*, pp. 48–54.
- Levinboim, T. and Chiang, D. (2015). “Supervised Phrase Table Triangulation with Neural Word Embeddings for Low-Resource Languages.” In *Proceedings EMNLP*, pp. 1079–1083.
- 三浦明波, Neubig Graham, Sakti Sakriani, 戸田智基, 中村哲 (2014). 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. 情報処理学会第 219 回自然言語処理研究会 (SIG-NL),

20号, pp. 1–7.

- 三浦明波, Neubig Graham, Sakti Sakriani, 戸田智基, 中村哲 (2015). 中間言語モデルを用いたピボット翻訳の精度向上. 情報処理学会第 222 回自然言語処理研究会 (SIG-NL), 2号, pp. 1–5.
- Miura, A., Neubig, G., Sakti, S., Toda, T., and Nakamura, S. (2015). “Improving Pivot Translation by Remembering the Pivot.” In *Proceedings ACL*, pp. 573–577.
- Neubig, G. (2013). “Travatar: A Forest-to-String Machine Translation Engine based on Tree Transducers.” In *Proceedings ACL Demo Track*, pp. 91–96.
- Neubig, G., Arthur, P., and Duh, K. (2015). “Multi-Target Machine Translation with Multi-Synchronous Context-free Grammars.” In *Proceedings NAACL*, pp. 484–491.
- Neubig, G., Nakata, Y., and Mori, S. (2011). “Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis.” In *Proceedings ACL*, pp. 529–533.
- Nirenburg, S. (1989). “Knowledge-Based Machine Translation.” *Machine Translation*, 4 (1), pp. 5–24.
- Och, F. J. (2003). “Minimum Error Rate Training in Statistical Machine Translation.” In *Proceedings ACL*, pp. 160–167.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). “BLEU: A Method for Automatic Evaluation of Machine Translation.” In *Proceedings ACL*, pp. 311–318.
- Shannon, C. E. (1948). “A Mathematical Theory of Communication.” *Bell System Technical Journal*, 27 (3), pp. 379–423.
- Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). “Feature-rich Part-of-speech Tagging with a Cyclic Dependency Network.” In *Proceedings NAACL*, pp. 173–180.
- Toutanova, K. and Manning, C. D. (2000). “Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger.” In *Proceedings EMNLP*, pp. 63–70.
- Utiyama, M. and Isahara, H. (2007). “A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation.” In *Proceedings NAACL*, pp. 484–491.
- Zhu, X., He, Z., Wu, H., Zhu, C., Wang, H., and Zhao, T. (2014). “Improving Pivot-Based Statistical Machine Translation by Pivoting the Co-occurrence Count of Phrase Pairs.” In *Proceedings EMNLP*, pp. 1665–1675.
- Ziemski, M., Junczys-Dowmunt, M., and Pouliquen, B. (2016). “The United Nations Parallel Corpus v1.0.” In *Proceedings LREC*, pp. 3530–3534.

## 略歴

**三浦 明波**：2013年イスラエル国テクニオン・イスラエル工科大学コンピュータ・サイエンス専攻卒業。2016年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。現在、同大学院博士後期課程在学。機械翻訳，自然言語処理に関する研究に従事。情報処理学会，言語処理学会，ACL 各会員。

**Graham Neubig**：2005年米国イリノイ大学アーバナ・シャンペーン校工学部コンピュータ・サイエンス専攻卒業。2010年京都大学大学院情報学研究科修士課程修了。2012年同大学院博士後期課程修了。2012～2016年奈良先端科学技術大学院大学助教。現在，カーネギーメロン大学言語技術研究所助教，奈良先端科学技術大学院大学客員准教授。機械翻訳，自然言語処理に関する研究に従事。

**Sakriani Sakti**：1999年インドネシア・バンドン工科大学情報卒業。2002年ドイツ・ウルム大学修士，2008年博士課程修了。2003～2011年ATR音声言語コミュニケーション研究所研究員，情報通信研究機構主任研究員。現在，奈良先端科学技術大学院大学情報科学研究科助教。2015～2016年フランスINRIA滞在研究員。統計的パターン認識，音声認識，音声翻訳，認知コミュニケーション，グラフィカルモデルの研究に従事。JNS, SFN, ASJ, ISCA, IEICE, IEEE 各会員。

**戸田 智基**：1999年名古屋大学工学部電気電子・情報工学科卒業。2003年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。同年日本学術振興会特別研究員-PD。2005年奈良先端科学技術大学院大学情報科学研究科助手。2007年同助教。2011年同准教授。2015年より名古屋大学情報基盤センター教授。工学博士。音声情報処理の研究に従事。IEEE，電子情報通信学会，情報処理学会，日本音響学会 各会員。

**中村 哲**：1981年京都工芸繊維大学工芸学部電子工学科卒業。京都大学工学博士。シャープ株式会社。奈良先端科学技術大学院大学助教，2000年ATR音声言語コミュニケーション研究所室長，所長，2006年（独）情報通信研究機構研究センター長，けいはんな研究所長などを経て，現在，奈良先端科学技術大学院大学教授。ATRフェロー。カールスルーエ大学客員教授。音声翻訳，音声対話，自然言語処理の研究に従事。情報処理学会喜安記念業績賞，総務大臣表彰，文部科学大臣表彰，Antonio Zampoli 賞受賞。ISCA 理事，IEEE SLTC 委員，IEEE フェロー。

(2016年5月18日 受付)

(2016年8月31日 再受付)

(2016年10月12日 採録)