

# 統計的機械翻訳における WordNet を用いた フレーズ意味曖昧性解消手法の提案

大山 鉄郎<sup>1,a)</sup> 関 洋平<sup>2,b)</sup>

**概要:** 文脈の特徴から単語の意味を推定する語義曖昧性解消 (Word Sense Disambiguation: WSD) は、機械翻訳の分野において、多義語の翻訳に有用な要素技術として認識されている。しかし、フレーズベース統計的機械翻訳 (Phrase-Based Statistical Machine Translation: PBSMT) では、フレーズ単位による翻訳を行うため、単語を単位とした WSD の効果が必ずしも期待できない。本研究では、翻訳単位であるフレーズの意味の曖昧性解消手法を提案する。提案手法は、概念辞書 WordNet を用いて、入力文の周辺語と対象フレーズを含む対訳文の周辺語との意味的な類似度からフレーズの意味を推定する。提案手法を組み込んだ機械翻訳システムの翻訳精度は、組み込まない場合と比較して、自動評価尺度 BLEU において有効性を示した。

**キーワード:** フレーズ意味曖昧性解消, フレーズベース統計的機械翻訳, 語義曖昧性解消, WordNet

## Phrase Sense Disambiguation Method Using WordNet in Statistical Machine Translation

OHYAMA TETSURO<sup>1,a)</sup> SEKI YOHEI<sup>2,b)</sup>

**Abstract:** Word Sense Disambiguation to estimate the meaning of a word from context features is recognized as useful element technology for translation of polysemous words in Machine Translation (MT). On the other hand, WSD is not effective in Phrase-Based Statistical Machine Translation because it assumed the word as a translation unit, not phrase unit. In this study, we propose a disambiguation method of the meaning of a phrase unit. Our method estimates the meaning of the phrase by semantic similarity from neighboring words between the input sentence and parallel sentences using WordNet. The proposed MT system showed the effectiveness in BLEU scores against a baseline system.

**Keywords:** phrase sense disambiguation, PBSMT, WSD, WordNet

### 1. はじめに

統計的機械翻訳 (Statistical Machine Translation: SMT) では、翻訳する語が多義性を持つことが原因で適切な翻訳ができないことがある。文脈の特徴から単語の意味を推定する語義曖昧性解消 (Word Sense Disambiguation: WSD) は、多義語の翻訳において適切な訳語を選択し、

SMT の精度を向上させるための要素技術として認識されている [3], [10]。

SMT における WSD では、WSD の対象語とその語義を対訳コーパスから取得することが多い [1], [3], [10]。アライメントによって対訳文それぞれの単語対応が取得され、得られた単語対応の原言語側が WSD の対象語に、目的言語側がその語義として扱われる。語義の決定には教師あり学習がよく用いられ、対象単語と対応する翻訳候補から適切な訳語を選択する単語翻訳 [10] や、n-best の翻訳仮説を WSD を用いてリランキングする手法 [1] など、

<sup>1</sup> 筑波大学大学院 図書館情報メディア研究科

<sup>2</sup> 筑波大学 図書館情報メディア系

<sup>a)</sup> ohyama@slis.tsukuba.ac.jp

<sup>b)</sup> yohei@slis.tsukuba.ac.jp

さまざまな研究がされている．フレーズベース統計的機械翻訳 (Phrase-Based Statistical Machine Translation: PBSMT) においては, Carpuat ら [3] が, 曖昧性解消の対象を単語からフレーズとした, フレーズ意味曖昧性解消 (Phrase Sense Disambiguation: PSD) の提案を行い, 曖昧性解消と翻訳の単位を同一にすることで翻訳精度が向上することを示した．この際のフレーズとは, 対訳コーパスから自動的に抽出された複数の単語のまとまりのことを指し, 一般に用いられるフレーズとは定義に相違がある．PSD は, WSD を機械翻訳に適用させるための有効な手法といえる．PSD では, 曖昧性解消の対象であるフレーズとそのフレーズ意味は対訳コーパスから自動的に決定される．そのため, 対象フレーズは辞書に登録されていない場合が多く, 対象語とその周辺語を WordNet<sup>\*1</sup>に問い合わせて語義を推定する手法 [9] などの, 対象語で辞書に問い合わせるといった処理が適用できないことがある．Carpuat らの手法では言語資源として対訳コーパスのみを使用して PSD を行っているが, 対訳コーパスの量が十分でない場合, フレーズ意味を推定することが難しい．

本研究では, PSD を行うには対訳コーパスの量が不十分な場合を想定し, 外部資源として WordNet を用いた PSD 手法を提案する．提案手法は, 対象フレーズとその仮説を用いて対訳コーパスから PSD に有用な対訳文を選択し, 入力文と対訳文集合それぞれのフレーズ周辺語の意味的な類似度から, フレーズ意味を推定する．

本稿の構成を以下に示す．まず, 2 節で関連研究について紹介し, 3 節で提案手法である WordNet を用いたフレーズ意味曖昧性解消手法について説明する．4 節で翻訳による実験を行う．5 節で考察, 6 節でまとめと今後の課題を示す．

## 2. 関連研究

Carpuat ら [2] の PSD は, Kernel PCA に基づいたモデルに加えて Naive Bayes, 対数線形モデル, ブースティング法を組み合わせたアンサンブル手法を用いている．この手法では, 対訳コーパスに含まれている語のみを考慮して評価値を決定する．そのため, 対訳コーパスが十分に用意できない場合, 適切な評価値が与えられない場合がある．本稿で提案する手法では, WordNet を利用することで, 対訳コーパスに含まれていない語についても概念間の類似度を用いて評価値を与えることができる．

Pedersen ら [9] は, WSD の対象語を含む文脈に対して, WordNet を用いて語義間に定義される任意の関連性尺度を定義することで語義を推定する手法を提案している．Pedersen らの語義推定は, 以下の式で行われる．

$$\arg \max_{i=1}^{m_i} \sum_{j=1, j \neq t}^n \max_{k=1}^{m_j} \text{rel}(s_{ti}, s_{jk}) \quad (1)$$

$w_1, w_2, \dots, w_n$  は, 文脈内の語を,  $w_t$  は語義を付与する対象語を表す． $s_{i1}, s_{i2}, \dots, s_{im_i}$  は,  $w_i$  が持つ,  $m_i$  個の語義である． $\text{rel}$  は語義のペアに対してその関連性スコアを与える．Pedersen らの手法は, 任意の語義ペアに評価値を与える手法であるが, その語義が WordNet に登録されていないと適用できない．本研究の PSD の対象であるフレーズは, 対訳コーパスから選択される複数の形態素のため, 直接問い合わせることができない．

提案手法では, フレーズの周辺語と対訳文集合に含まれる語義との類似度を WordNet を用いて計算することでフレーズ意味を決定する．対訳文集合とは, フレーズ意味を推定するために有用な対訳文の集合であり, 対訳コーパスから選択される．フレーズの周辺語の意味的な類似度を用いることで, フレーズや周辺語が対訳コーパスでほとんど出現しない場合でも, WordNet に登録されていれば類似度を算出することができる．

## 3. WordNet を用いたフレーズ意味曖昧性解消

本節では, 提案手法である, WordNet を用いた PSD 手法について述べる．本研究では, PSD 手法を素性関数の形で実装することで, PBSMT に PSD の評価値を与える．これにより PSD の処理単位は, PBSMT の翻訳単位と同一となる．一般的な PBSMT [4] では, 以下の対数線形モデルに基づいて出力文  $\hat{e}$  が決定される．

$$\hat{e} = \arg \max_e \sum_{i=0} \lambda_i h_i(e, f) \quad (2)$$

$e$  は出力である目的言語文の仮説,  $f$  は入力である原言語文を表す． $h(e, f)$  は素性関数であり,  $\lambda$  は各素性関数に対する重みである．

提案手法の PSD は素性関数として, 以下の式で与えられる．

$$\phi_{\text{PSD}}(e, f) = \sum_i \text{eval}(\bar{e}_i, \bar{f}_i) \quad (3)$$

$$\text{eval}(\bar{e}, \bar{f}) = \sum_{i=1}^n \max_{j,k,l} \text{rel}(t_{kl}, s_{ij}) \quad (4)$$

$\bar{e}, \bar{f}$  は,  $e, f$  に含まれるフレーズである． $s_{i1}, s_{i2}, \dots, s_{im_i}$  は,  $f$  が含む語  $w_i$  が持つ  $m$  個の語義である．ただし, 対象のフレーズそのものは含まないため,  $w_i \notin \bar{f}$  である． $\text{rel}(c_1, c_2)$  は語義のペアに対してその関連性スコアを与える． $t_{kl}$  は, 対訳コーパスから生成される対訳文集合が含む語  $x_k$  が持つ  $l$  番目の語義である． $x_k$  は  $w_i$  と同様に, 対象のフレーズは含まないため,  $x_k \notin \bar{e}$  である．

提案手法の流れを図 1 に示す．提案手法は, (1) 対訳文

\*1 <http://wordnet.princeton.edu/>

集合の構築, (2) 語義集合の構築, (3) 類似度の計算の3つの処理から構成されている. 次節よりそれぞれについて説明する.

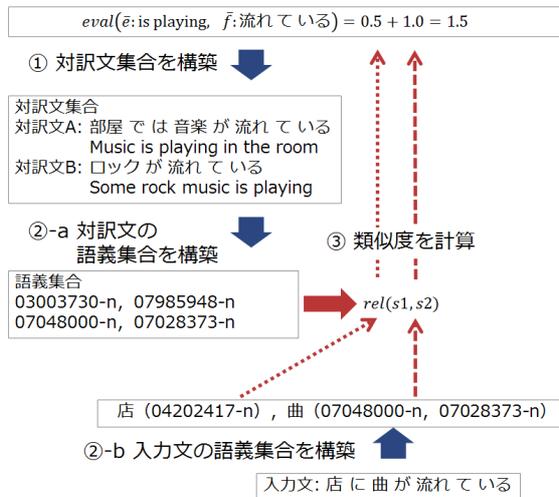


図 1 提案手法の流れ

提案手法は, フレーズとその仮説に対して, フレーズを除く入力文の語 (フレーズの周辺語) の語義それぞれについて, 対訳文集合から構築された語義集合から最も類似する語義の類似度を計算し, その総和を評価値として与えることで, フレーズの曖昧性解消を行う.

### 3.1 フレーズと仮説に関連する対訳文の選択

対訳文集合は, フレーズと仮説を用いて対訳コーパスから選択される. あらかじめ, 対訳コーパスに含まれる対訳文ごとにフレーズを抽出しておき, 素性関数の入力であるフレーズと仮説とをそれぞれ含む対訳文を, 対訳文集合とする.

### 3.2 対訳文を用いた語義集合の構築

フレーズと仮説から選択された対訳文集合から, 入力文の周辺語と比較するための語義集合を構築する. 語義集合は, 対訳文が含む名詞, 動詞, 形容詞, 副詞に対応する WordNet の synset (語義) の集合である.

本研究では, 以下の2種類の方法で語義集合を構築した.

- (1) union: 対訳文ごとに原言語側, 目的言語側の文が含む語が属する語義を抽出し, 2つの集合の和を対訳文の語義集合とする. 最後に, それぞれの対訳文から得られた語義集合をまとめて, フレーズと仮説とから得られた対訳文集合に対する語義集合とする.
- (2) intersection: 対訳文ごとに原言語側, 目的言語側の文が含む語が属する語義を抽出し, 2つの集合の積を対訳文の語義集合とする. 最後に, それぞれの対訳文から得られた語義集合をまとめて, フレーズと仮説とから得られた対訳文集合に対する語義集合とする.

単語から語義を選択する際は, 単語の原形と, その品詞の一致を条件とした. 構築した2種類の語義集合のうち, 積を取った intersection のほうが, その対訳文において不適切な語義を削除できていると考えられる. しかし, 提案手法は, 入力文におけるフレーズの周辺語それぞれに対して, 最も意味的な類似度の高い語義を選択するものであるため, 不適切な語義が含まれていることによる悪影響はそれほど大きくないと考えられる. そのため, 本研究では2種類の語義集合を構築することとした.

提案手法は, フレーズと仮説が得られるたびに対訳文集合と語義集合の構築が必要となる. それらは時間的コストの大きい処理となるため, あらかじめ語義集合を構築しておく. このとき入力文は未知であるため, すべての可能性を考慮して, フレーズテーブルのフレーズ対それぞれについて語義集合を構築する. この処理によって, フレーズと仮説が得られたとき, それに対応する語義集合は読み込み処理のみで獲得できる. 翻訳時における提案手法における主な処理は, 入力文から獲得するフレーズ周辺語の語義集合の構築と, 次節の類似度計算のみとなる.

### 3.3 WordNet を用いた類似度

語義ペアに対する関連性スコアは, WordNet を用いて求める. WordNet は synset と呼ばれる同義語集合が階層構造となって表現されている英語の概念辞書である. 日本語版として日本語 WordNet [5] がある. 日本語 WordNet では, 現在, 57,238 概念 (synset 数), 93,834 語, 158,058 語義 (synset と単語のペア) が登録されている. WordNet を用いることにより, 語と語の意味的な類似度を計算することができる.

式 (1) の  $rel(c_1, c_2)$  で与えられる類似度 [9] には, 以下を使用する.

$$rel_{path}(c_1, c_2) = \frac{1}{length(c_1, c_2) + 1} \quad (5)$$

$$rel_{lin}(c_1, c_2) = \frac{2 * IC(LCS)}{IC(c_1) + IC(c_2)} \quad (6)$$

$$IC(c) = -\log P(c) \quad (7)$$

$length(c_1, c_2)$  は  $c_1$  と  $c_2$  の最短パス長を,  $LCS(c_1, c_2)$  は  $c_1$  と  $c_2$  の最小共通概念を,  $IC(c)$  は  $c$  の選択情報量を表す関数である.

式 (5) は, 語義間のパス長を, 式 (6) は語義の選択情報量を考慮した尺度となっている. 本研究では, 選択情報量の計算には対訳コーパスを用いている.

## 4. 実験: 翻訳精度の評価

提案手法による翻訳システムと, ベースラインシステムを用いて翻訳を行い, 翻訳精度を BLEU [8] と RIBES [11] により評価する.

表 1 翻訳精度

評価尺度	BLEU [8]			ベースライン との有意差	RIBES [11]			ベースライン との有意差
	CSTAR03	IWSLT04	IWSLT05		CSTAR03	IWSLT04	IWSLT05	
ベースライン	34.20	33.35	26.41		72.05	72.50	68.06	
(1) path, union	<b>35.27</b>	<b>35.22</b>	<b>28.30</b>	あり	<b>72.19</b>	<b>74.50</b>	<b>70.31</b>	なし
(2) path, intersection	35.18	<b>35.22</b>	27.72	あり	72.17	<b>74.50</b>	69.70	なし
(3) lin, union	35.19	35.15	26.84	なし	71.98	74.30	69.48	なし
(4) lin, intersection	35.22	<b>35.22</b>	27.40	あり	72.01	<b>74.50</b>	69.26	なし

#### 4.1 実験データ

実験データとして、高度言語情報融合フォーラム ALA-GIN \*2で公開されている日英翻訳エンジン学習・評価用対訳コーパスを用いた。コーパスから、訓練データとして19,972文、テストデータとしてCSTAR03(506文), IWSLT04(500文), IWSLT05(506文)の3セットを使用した。開発データとしては、CSTAR03を用いた。

#### 4.2 実験方法

翻訳のデコーダには Moses [4] を、アライメントには GIZA++ [7] を、チューニングには MERT [6] を用いた。

ベースラインは、素性関数に提案手法によるスコアを追加していない場合の翻訳精度とした。提案手法は、union と intersection の2種類の語義集合と  $rel_{path}$  と  $rel_{lin}$  の2種類の類似度の組み合わせからなる計4種類の設定で実験を行い、翻訳精度を測定した。以降、提案手法は、採用する戦略に基づき、union, intersection, path, lin, などと表記する。

#### 4.3 実験結果

実験結果を表1に示す。提案手法は、いずれの設定でもベースラインを上回る結果を得た。また、path は lin より高い精度を示した。提案手法で最も良い結果となった設定は、語義集合に union, 類似度に path を用いたものであった。最も良い翻訳精度を示した提案手法は、ベースラインと比較して、BLEU で、平均 1.61 ポイント、RIBES で、平均 1.46 ポイントの向上が見られた。

lin と intersection の組み合わせを除く3つの場合については、BLEU において、ベースラインとの両側 t 検定 ( $p < 0.05$ ) において有意差があった。RIBES ではすべての設定において有意差はなかった。

### 5. 考察

出力結果と翻訳精度より、以下の4点について考察を行う。

- (1) 類似度の計算方法の翻訳精度への影響
- (2) 対訳文集合や語義集合の違いによる影響
- (3) 改善された文の傾向

#### (4) 改悪された文の原因

#### 5.1 類似度の計算方法の翻訳精度への影響

path と lin による翻訳では、わずかに path の方が高い翻訳精度を示した。path は語義の距離を、lin は語義の選択情報量を考慮した類似度である。

lin では、適切な選択情報量が計算できないことが原因で誤った翻訳をする例が見られた。本研究では、選択情報量は、翻訳の訓練データに用いた対訳コーパス(19,972文)から計算されている。このコーパスのサイズでは、一部の語義に関して選択情報量を計算するのに十分でないことがあった。本来、高頻度/低頻度であるはずの語義が対訳コーパス中では低頻度/高頻度になることで選択情報量の計算を誤ることがある。しかし、lin は選択情報量を適切に計算することができれば、用いるコーパスを考慮した評価値を与えられる有用な尺度であるといえる。本研究では、対訳コーパスを用いているが、選択情報量は他の単言語コーパスなどからでも計算することができる。したがって、より規模の大きいコーパスで選択情報量を計算することにより、翻訳精度が改善する可能性がある。

また、提案手法では、フレーズの周辺語の語義と、対訳文から構築された語義集合を比較するが、対訳文から構築された語義集合には、不適切な語義が含まれている可能性がある。この不適切な語義は類似度を計算する際のノイズとなる。path では、“フレーズの周辺語の語義と距離的に近い語義”を選択するため、lin と比較すると比較的このノイズに対して頑健であるといえる。

#### 5.2 対訳文集合や語義集合の違いによる影響

提案手法では、フレーズと仮説より対訳文集合が構築され、そこから語義集合が構築される。そのため、フレーズと仮説によっては、対訳文集合が非常に大きくなり、他の仮説よりも有利に働きやすい。しかし、本コーパスにおいては、対訳文集合のサイズ差による影響はあまり見られなかった。これは、対訳文集合のサイズと素性関数として用いている翻訳確率との間にある程度の相関があるため、問題として表面化しなかった(もしくは翻訳確率に隠れて観測できなかった)か、語義間の類似度計算は、入力文の周辺語それぞれに対して最大値を取る1つの語義を選択す

\*2 <http://alaginrc.nict.go.jp/>

るという処理のため、サイズの影響がわずかであったことが原因だと考えられる。

語義集合による精度の差は、2つの類似度で反対の傾向を示した。path では union がより高い精度を示し、lin では intersection がより高い精度を示した。

まず、union がより高い精度を示した path について考察する。path の計算では、intersection の語義集合が類似度を計算するのに不十分だったため、より多い語義を含む union がより高い精度を示したと考えられる。union は、intersection より不適切な語義が含まれている可能性が高いといえるが、どちらの語義集合も、元となる対訳文集合は、“抽出されたフレーズ対集合が、フレーズと仮説を含む場合”という条件を満たす対訳文で構築されている。前節の通り、path は、語義集合に不適切な語義が含まれるというノイズに対して頑健であるといえる。そのため、path の計算に用いる語義集合は、対訳文集合構築の際の条件による選別である程度十分な効果を示し、その結果、intersection より多くの語義を含む union で、より高い精度が出たと考えられる。

lin では、intersection がより高い精度を示した。これは、union では不適切な語義を含んでしまっており、その語義を最大値として採用してしまっただと考えられる。lin では、語義の選択情報量を用いて概念間の類似度を計算する。そのため、path の“フレーズの周辺語の語義と距離的に近い語義”という条件が発生せず、不適切な語義に対して最大値を与えてしまうことがある。

2種類の語義集合に対する2つの類似度による結果より、より多くの語義を含むが不適切な語義も多くなる union、不適切な語義は少ないが、適切な語義も少なくなってしまう intersection と、3.2節の語義集合の構築で立てた仮説とほぼ同じ知見が得られた。語義集合としてどちらが優位かは、用いる類似度によって変化し、本研究で試した範囲では、概念間の距離から算出する path では union、選択情報量を用いる lin では intersection が有効であった。

### 5.3 改善された文の傾向

本研究で使用したコーパスでは、提案手法の利点は、動詞や助動詞で構成されるフレーズに多く表れた。“ありますか”や“ですか”といったフレーズは、単体ではどう翻訳されるのか決定できないことが多くある。提案手法では、フレーズの周辺語を考慮して翻訳を行うため、表現の似ている対訳文が対訳コーパスに含まれていたとき、その対訳文の語義が類似度で高い値となり、目的言語側の表現が採用されやすくなる。

逆に、名詞や形容詞といったフレーズでは、有効性が必ずしも確認できなかった。ただし、これに関しては、本研究で使用したコーパスが旅行英会話文を収録したもので、名詞や形容詞に多義的なものがあまり含まれていないことが有効性が見られなかった原因の一つだと考えられる。

翻訳結果が明確に改善された例としては、対訳コーパスに入力文と同じ、またはほぼ同じ文が含まれていた場合が挙げられる。提案手法によって翻訳が改善された例を以下に示す。

#### 例 1

入力文: もう これ 以上 我慢 できません。  
ベースライン: I can't stand more than I  
提案手法: I can't stand any more of this.

#### 例 2

入力文: これは 何 の 料金 ですか。  
ベースライン: This is What  
提案手法: What's this charge for? Do

#### 例 3

入力文: 休憩 は あります か。  
ベースライン: Do you have any including  
提案手法: Is there an intermission?

提案手法では、入力文と表現が似ている対訳文が、対訳コーパスに含まれるとき、その対訳文から生成される語義が、類似度の計算において最大値を取りやすい。このときの挙動は、入力文と類似する例文をもとに翻訳を行う、用例翻訳の翻訳処理と近いものになる。例1から例3は、入力文と同一の文が訓練に用いた対訳コーパスに含まれているものである。オーソドックスな統計的機械翻訳では、訓練に用いる対訳文はフレーズに分解して保持するため、テストとして訓練に用いた文と同様の文が入力されても翻訳結果が適切になることは保証されていない。提案手法でも、フレーズごとに翻訳が行われるため結果は保証されないが、類似する文が対訳コーパスに含まれることで、その語義が類似度計算において有利に働くため、適切な翻訳がされやすくなる。

今回提案手法によって翻訳が正確に改善された文は、対訳コーパス中に同様の文が出現する文でしか確認できなかったが、部分的な改善は各所に見られた。ただし、パス長や選択情報量といった尺度を用いるには、実験に用いた対訳コーパスでは不十分と考えられ、よりサイズの大きい対訳コーパスを使って実験を行う必要がある。

### 5.4 改悪された文の原因

出力文が悪くなってしまった文の原因として、以下の2点が挙げられる。

- (1) 対訳文集合が適切に構築されない。
- (2) フレーズの周辺語が手がかりにならない。

対訳文集合は、素性関数の入力であるフレーズと仮説をフレーズ対としてを持つ対訳文から構築される。このときアライメントの誤りが原因で、不適切なフレーズと仮説が

入力されることがある。誤ったアライメントのフレーズと仮説からは、適切な対訳文集合を構築できず、語義集合も不適切なものとなる。しかし、提案手法は、フレーズの周辺語の語義と語義集合の類似度から評価値を決定するため、アライメントが誤ったものに対しても高い評価値を与えてしまうことがある。

入力文には、フレーズの周辺語がフレーズを翻訳する手がかりにならないものもある。現在は、すべてのフレーズと仮説に対して、PSDの評価値を与えている。しかし、フレーズの周辺語が手がかりにならない場合においては、提案手法そのものがノイズとなってしまうため、結果として出力文を改悪する。提案手法は現在、フレーズの周辺語がフレーズの翻訳に有用であるという仮定のもとに成り立っている。この原因の改善のためには、すべてのフレーズと仮説について評価するのではなく、提案手法が有効なフレーズと仮説の条件を追加することが有用であると考えられる。

## 6. おわりに

本研究では、統計的機械翻訳における WordNet を用いたフレーズ意味曖昧性解消手法を提案した。提案手法は、ベースラインの機械翻訳システムと比較して、BLEU 平均 1.61 ポイント、RIBES 平均 1.46 ポイントの翻訳精度の向上が見られた。

フレーズの周辺語の語義と対訳コーパスから選択された文の語義集合との類似度を用いることで、“ありますか”や“ですか”といった動詞や助動詞で構成されたフレーズなど、一定数のフレーズの翻訳を改善できることを示した。

今後の課題として、提案手法が有効に働く条件の設定や、類似度を計算する際に比較する語義の制限が挙げられる。

提案手法は、一定のフレーズでは有効に働いたが、高頻度であったり、周辺語が手がかりにならないフレーズの翻訳においては、不適切なことがある。提案手法の改良または、提案手法が有効でない条件を定式化することで、より翻訳精度を向上できると考えられる。

類似度の計算は、フレーズの周辺語それぞれについて、対訳文から構築された語義集合のすべての語義を用いて行われる。現在の方法では、不適切な語義間の類似度を計算してしまう場合もあることから、フレーズに対する周辺語の制限、文の構造情報等から比較する語義を選択するなど、語義集合の中から比較に適切な語義を選択する必要があると考えられる。

謝辞 本研究の一部は、科学研究費補助金基盤研究 C (課題番号 24500291)、基盤研究 B (課題番号 25280110) ならびに萌芽研究 (課題番号 25540159) の助成を受けて遂行された。

## 参考文献

- [1] Marianna Apidianaki, Guillaume Wisniewski, Artem Sokolov, Aurelien Max, and Francois Yvon. WSD for n-best reranking and local language modeling in SMT. In *Proceedings of Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-6)*, pp. 1–9, Jeju, Korea, 2012.
- [2] Marine Carpuat, Weifeng Su, and Dekai Wu. Augmenting Ensemble Classification for Word Sense Disambiguation with a Kernel PCA Model. In *Proceedings of Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pp. 88–92, Barcelona, Spain, 2004.
- [3] Marine Carpuat and Dekai Wu. Improving Statistical Machine Translation using Word Sense Disambiguation. In *Proceedings of Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL2007)*, pp. 61–72, Prague, Czech Republic, 2007.
- [4] 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 *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL2007) demonstration session*, pp. 177–180, Prague, Czech Republic, 2007.
- [5] Kow Kuroda, Francis Bond, and Torisawa Kentaro. Why Wikipedia Needs to Make Friends with WordNet. In *Proceedings of the 5th International Conference of the Global WordNet Association (GWC-2010)*, pp. 9–16, Mumbai, India, 2010.
- [6] Franz Josef Och. Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of the 41th Annual Meeting of the Association for Computational Linguistics (ACL2003)*, pp. 160–167, Sapporo, Japan, 2003.
- [7] Franz Josef Och and Hermann Ney. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, Vol. 29, No. 1, pp. 19–51, 2003.
- [8] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL2002)*, pp. 311–318, Philadelphia, USA, 2002.
- [9] Ted Pedersen, Satyanjeev Banerjee, and Siddharth Patwardhan. Maximizing Semantic Relatedness to Perform Word Sense Disambiguation. In *Research Report UMSI*, 2005.
- [10] David Vickrey, Luke Biewald, Marc Teyssier, and Daphne Koller. Word-Sense Disambiguation for Machine Translation. In *Proceedings of the 2005 Conference on Empirical Methods in Natural Language (EMNLP2005)*, pp. 771–778, Vancouver, Canada, 2005.
- [11] 平尾努, 磯崎秀樹, Kevin Duh, 須藤克仁, 塚田元. RIBES: 順位相関に基づく翻訳の自動評価法. 言語処理学会第 17 回年次大会 (NLP2011), pp. 1115–1118, 2011.