

# 語義曖昧性解消の機械翻訳への利用可能性を探る

藤田 早苗

NTT コミュニケーション科学基礎研究所 奈良先端科学技術大学院大学

fujita.sanae@lab.ntt.co.jp

Graham NEUBIG

neubig@is.naist.jp

## 1 はじめに

語義曖昧性解消（以下、WSD）は古くから研究されてきている課題だが、実用化には至っていない。実用化に近づくためには、精度の向上をはかるとともに、対象とする語や語義にも着目する必要があるだろう。

現在 WSD では、対象語ごとにモデルを構築する方法が主流である。しかし WSD の結果を何にどう利用するかによって、対象語や語義粒度、必要とされる精度は異なると考えられる。そのため、具体的に利用先を想定し、適切な対象語や語義粒度を検討する必要がある。

日本語の WSD タスクのコンペティションは、Senseval-2, SemEval-2010 での 2 回行われている。両方とも、対象語<sup>1</sup>が決められている lexical sample タスクであり、岩波国語辞典 [9] の語義を推定するタスク設定となっている。岩波国語辞典では、語義は階層構造を持っている。複合語の一部かどうかを示す階層を含めると、階層構造の最大の深さは 4 である。本稿では、最も深い階層を細分類、一つ上位の階層を中分類と呼ぶ。Senseval-2 は細分類までの一致の評価や、階層構造にしたがって部分的なスコアを与える評価方法を取り入れたタスク設定となっている。SemEval-2010 は中分類の語義を推定するタスク設定であり、Project Next NLP<sup>2</sup> の WSD 班では SemEval-2010 と同じ設定でエラー解析を行っている。

本稿では機械翻訳（以下、MT）への利用を前提とし、Project Next NLP の MT 班で報告されたエラー例を基に適切な WSD の対象語や語義粒度を検証する。

なお、WSD を日英の MT に利用することを意図したタスクとして Senseval-2 の翻訳タスク [6] がある。翻訳タスクでは、トランスレーション・メモリーと呼ばれる対訳用例集合のうちから、テスト文の対象語の翻訳に利用できる用例を選ぶか、翻訳結果そのものを答えるというタスク設定だった。本稿では、語義を推

定するという従来型の WSD が MT のエラー解消に役立てるかどうかという点に着目したい。

以降、2 章では、MT 班で報告されたエラー例を分析し、WSD でエラーが解消できる可能性があるかどうかという観点で分析を行う。WSD により、エラーが解消出来る可能性があると考えた対象語に対しては、語義を推定する実験を行う。まず 3 章では、評価・訓練を行うためのデータについて述べる。4 章では、本稿で実験に利用するシステムを紹介する。5 章では実験結果の紹介と分析を行う。さらに、6 章ではそれまでの実験結果を受け、従来の辞書の語義を推定する方法ではなく、訳語を語義として語義曖昧性解消をすることを提案し、実験を行う。7 章では、語義曖昧性解消結果を統計的機械翻訳に利用し、翻訳に与える影響について考察する。最後に 8 章でまとめと今後の課題について述べる。

## 2 WSD 関連 MT エラー例の分析と対象語・語義粒度の決定

Project Next NLP の中間報告会において MT 班により MT 結果のエラー原因が語彙選択誤りだと報告された語を対象語の候補とする。

本章ではまず、これらの語を WSD の対象語として扱うことができるか、また、岩波国語辞典の語義分類に分けることにより、正しい英訳が選択できる可能性があるかどうかという観点で分類した。結果を表 1 に示す。

表 1 で、[A][B] は、岩波国語辞典の語義分類に分けることができれば正しい英訳の選択ができそうな例、[C] は辞書の語義をさらに細分化することで英訳の選択ができそうな例である。また、[D][E] は辞書の分類では英訳の選択ができないと考えられる例である。

[E] は“内容語以外”であり、訳語選択は動詞などの用言との関係解析に依存すると考えられる。動詞自身の訳語選択も、「絵を かく」「色を ぬる」「マリオを

<sup>1</sup>Senseval-2 は名詞と動詞各 50 語ずつ計 100 語、SemEval-2010 は 50 語が対象語

<sup>2</sup><https://sites.google.com/site/projectnextnlp/>

表 1: MT エラー例の分類

[A] 中分類で可能 (5, 35.7 %)	
(a-1) 3絵をかいたり、色をぬったりして、きれいにしあげる。	
(a-2) 水難での救援するためのベストの名前は、なんというのですか？	
(a-3) ...おじさんって独特の匂いがある方居ますよね？	【細分類なし】
(a-4) ※バリクラブ、米 ブルッキングズ研究所などによる	【細分類なし。品詞推定か NE が
(a-5) ドビルパン 仏 外相は十八日、仏 記者団との...仏 政府にも、...	正しければ不要?】
[B] 細分類で可能 (2, 14.3 %)	
(b-1) 単位が取れなくても「落第」はしないが、全員に受講を義務付ける。	
(b-2) どうやらまさやんの曲がNHKドラマ...	【品詞推定か NE が正しければ中分類で可】
[C] 更に細分化が必要 (3, 21.4 %)	
(c-1) 別れた夫が嫉妬して、という線もあるわ。	【「方針」と「水準」の意味を分ける】
(c-2) 3絵をかいたり、色をぬったりして、きれいにしあげる。	【「かく」「ぬる」は辞書では単義語】
[D] 辞書以外の分類が必要 (2, 14.3 %)	
(d-1) ファミコンでマリオをしたことしかありません。	【「マリオ」の WSD が必要】
(d-2) そんなところか？	【いろいろな解釈が可能】
[E] 内容語以外 (2, 14.3 %)	
(e-1) 露天風呂の石のところに隙間があって、男性社員に裸を見られた...	
(e-2) ふたつとして同じものはないそうですよ.....	

ただし、下線部は MT での誤り箇所。[A]-[E]、(a-1)-(e-2) は参照用に付与した記号。“...” は省略箇所、【】内はコメントを示す。また、NE は固有表現抽出を示す。

する」など、項構造解析 (以下、SRL) と同時に解くべきだと考える。ただし、「マリオをする」の場合は、「マリオ」がゲームであることがわかる必要がある。その部分は WSD の範疇だと捉えることもできるが、「マリオ」は岩波国語辞典には記載がない。[D] の「ところ」は形式名詞<sup>3</sup>であり、形式名詞が実際にあらわしている内容は辞書定義文から選べるような性質のものではない。

このように表 1 から、少なくともこれらの例では、MT における誤り箇所の品詞によって、WSD で対応が可能かどうか分けることができるといえる。つまり、形式名詞を除く名詞 (「ベスト」「仏」「米」「曲」「おじ」「線」「単位」)、および、形状詞 (「きれい」) では、WSD による改善の可能性がある。そのため、本稿ではこれらの語を対象として WSD に取り組む。

これらのうち、SemEval-2010 の設定である辞書の中分類でよいと分類したのは、5 例だけである。また、中分類といっても、必ずしも細分類が存在するわけではなく、[A] に含まれる対象語の中で、細分類が存在する語は「きれい」<sup>4</sup>と「ベスト」<sup>5</sup>のみである。そこで、本稿では細分類での語義曖昧性解消を基本とする。ただし、例 (c-1) は細分類の語義をさらに分割し、語義を増やすこととする。

### 3 評価・訓練用データ

本章では、評価、および、訓練に用いるデータと、語義付与の方針について述べる。

#### 3.1 語義付与方針

現代日本語書き言葉均衡コーパス (BCCWJ)<sup>6</sup> の Core データに岩波国語辞典の語義を付与した Sensebank [7] をラベルありデータのベースとして利用した。この Sensebank は SemEval-2010 で利用されたものを拡張したものであり、多義語の内容語にはすべて語義を付与するという方針で作成されている。

しかしながら、今回の対象語に対する語義付与率は、Core データ中の対象語の 22.9 %のみだった。原因の詳細は不明だが、品詞や読みの不一致によって作業対象から漏れたことなどが原因として考えられる。

例えば、「米」には「コメ」「ベイ」など複数の読みが存在する。‘亜米利加’の語義があるのは、「ベイ」だけなので、読みの一致を語義付与の条件に含めている場合、読み推定が誤っていると正しい語義が候補に挙がらないことになる。

また、品詞についても、例えば「曲」‘音楽上の作品を数える語。’の場合、“接尾辞-名詞的-助数詞”になると考えられるが、実際は、“普通名詞”と解析されることも多く、品詞による語義付与の制限を行うと正しい語義を付与できない場合がある。

<sup>3</sup>UniDic では“名詞-非自立”である

<sup>4</sup>細分類を用いると 3 クラス増える

<sup>5</sup>細分類を用いると 1 クラス増える

<sup>6</sup><http://www.ninjal.ac.jp/kotonoha/>

つまり、こうした問題を WSD の問題ではなく、形態素解析、あるいは、固有表現抽出の問題だと捉えることも考えられるが、本稿では、読みや品詞に限らず、形態素の出現形の字面が一致する語はすべて対象語として語義を付与する。ただし、「米」に対して「米露」のように、形態素区切りがエントリと一致しない場合は対象外とした。

### 3.2 ラベルありデータ

BCCWJ の Core データのうち、CoreA を評価用データ、その他のコアデータ (以下、CoreO) を訓練用ラベルありデータとする。また、Core データに出現しない語義のラベルありデータを獲得するため、辞書の例文も利用する。また、京大コーパスに語義付与したデータ (「檜」センスバンク [1]) も利用する。ただし、「檜」センスバンクには、Lexeed[5] の語義が付与されているため、Lexeed の語義から岩波国語辞典の語義が一意に決まるものだけを利用した。

また、辞書では読みが異なるとエントリも異なるが、字面が同じならまとめて対象語義として扱う必要がある。実用的には、どのエントリに対応する語なのかを推定すること自体が必要だと考えられるからである<sup>7</sup>。

最終的に語義を付与したデータのサイズを表 2 に示す。表 2 には、対象語の横に【低/中/高難易度】の分類を記載している。難易度の分類は、語義の頻度分布のエントロピー (式 (1), [10]) に基づくものであり、高難易度 ( $E(w) \geq 1$ )、中難易度 ( $0.5 \leq E(w) < 1$ )、低難易度 ( $E(w) < 0.5$ ) と分けられている。なお、式 (1) で、 $p(s_i|w)$  は、単語  $w$  の語義が  $s_i$  となる確率を示している。エントロピーの計算には、CoreO における語義の頻度分布を利用した。

$$E(w) = - \sum_i p(s_i|w) \log p(s_i|w) \quad (1)$$

### 3.3 ラベルなしデータ

後述のように、本稿では半教師あり学習を行う。半教師あり学習に用いるラベルなしデータには、対訳コーパスの日本語側テキストを利用する。対訳コーパスは、MT 班の解析で利用された Project Next MT 班が利用した対訳データで利用されているもの [11]<sup>8</sup> を利用した。表 3 に各対象語毎に得られたラベルなしデータのサイズを示す。

<sup>7</sup>SemEval-2010 では、エントリ自体は指定されている

<sup>8</sup><http://phontron.com/japanese-translation-data.php>

## 4 システム

### 4.1 学習器

ベースラインシステムの学習器には Support Vector Machine (SVM, [2]) の線形 SVM を利用する。これは、WSD 班の分析用のベースラインシステムとして利用されており、SemEval-2010 のベースラインシステムと基本的に同じ設定である [13]。

また、我々のシステムとしては半教師あり学習 (Maximum Hybrid Log-likelihood Expectation; MHLE, [3]) を利用する。MHLE はラベルありデータとテストデータが非常に異なる場合などにもロバストだという特徴がある。[4] は、MHLE の WSD への利用を報告しており、ラベルありデータが非常に少数でも、高い分類性能を持つことを示している。

### 4.2 素性

WSD 班の分析用システムで利用されている素性セットを BL とする。BL に含まれる素性は、対象語の出現形、品詞大分類、品詞細分類、前後 2 単語の出現形、品詞大分類、品詞細分類、分類語彙表の値 (5 桁、複数可)、および、対象語と係り受け関係にある語の出現形である。

BL に、対象語の出現する文中の全内容語の基本形 (BOW) を素性として加え、BL+BOW とする。さらに、分類語彙表の代わりに日本語語彙大系 [12] の意味カテゴリを利用したものを BLGT+BOW とする。なお、[4] では Gibbs サンプルングを用いたトピック分類<sup>9</sup> による結果を利用しているが、本稿では利用していない。対訳データには短い文が多く含まれるため、トピック分類による効果は少ないと考えたためである。

## 5 実験 (1)

### 5.1 結果

表 4 に結果を示す。表 4 で、MFS は最頻語義を正解とした場合の精度である。SVM は BL、BL+BOW、BLGT+BOW のいずれの素性を利用した場合でも、学習データに京大コーパスと岩波の例文を追加した場合でも、全く結果に変化がなく、いずれも最頻語義が選ばれている。

<sup>9</sup><http://gibbslda.sourceforge.net/>

表 2: ラベルありデータのサイズ

Class	Core		例文	京大	計	Sense ID	定義文 (一部略)
	A	O					
対象語: ベスト【中難易度語】							
1	0	46	0	11	57	46550-0-0-1-1	最上級。最良。もっともすぐれたもの。
2	0	1	1	0	2	46550-0-0-1-2	全力。「ベストを尽くす」▽best
3	1	6	0	0	7	46550-0-0-2-0	チョッキ。▽vest
4	0	0	0	0	0	46550-0-0-3-0	写真判の一つ。四センチに六・五センチの大きさ。▽vest
5	0	3	0	0	3	46550-1-0-0-0	(ある期間に) 最もよく売れた本。▽best seller
計	1	56	1	11	69		
対象語: 仏/フツ, ホトケ/【中難易度語】							
1	0	6	3	1	10	45292-0-0-1-0 48017-0-0-1-0	梵語の音訳「仏陀」の略。ほとけ。「仏教・仏道・仏法・仏説・仏者」
2	8	21	1	54	84	45292-0-0-2-0	「仏蘭西」の略。「フツ」と読む。「英仏・仏語」
3	0	0	0	0	0	48017-0-0-2-0	故人。死者。
4	0	0	0	0	0	48017-0-0-3-0	仏(1)のように温厚な、または情け深い人。
計	8	27	4	55	94		
対象語: きれい(綺麗・綺麗)【中難易度語】							
1	5	42	3	3	53	12645-0-0-1-0	美しい、特に見た目がきらびやか・はでで美しい様子。「花が綺麗だ」
2	0	14	0	0	14	12645-0-0-2-0	よごれ・余計なものがない様子。
3	0	0	3	2	5	12645-0-0-2-1	(利益や色恋のからまない) さっぱりした様子。「綺麗なつきあい」
4	2	0	0	2	4	12645-0-0-2-2	清潔。「綺麗・綺麗」
5	0	0	4	0	4	12645-0-0-2-3	《「綺麗に」の形で副詞的に》あとかたなく、さっぱり。「借金を綺麗に返す」
6	0	1	1	0	2	12645-1-0-1-0	見かけ・口先だけは、いちおう非難されないように体裁よくしてある物事。
7	0	0	0	0	0	12645-1-0-2-0	よごれないで済む仕事。
8	0	0	0	0	0	12645-2-0-0-0	美しく装った女。▽多く芸者を指す。
計	7	57	11	7	82		
対象語: 米/コメ, ベイ/【中難易度語】							
1	0	98	4	27	129	46313-0-0-1-0 18132-0-0-0-0	稲の実。こめ。よね。「米作・米飯」。以下「マイ」と読む。「白米・玄米」
2	0	0	0	0	0	46313-0-0-2-0	八十八歳。「米寿」▽「米」を分解すると「八十八」となることから。
3	12	250	0	312	574	46313-0-0-3-0	「亜米利加」、また「亜米利加合衆国」の略。「米国・英米」
4	0	0	0	0	0	46313-0-0-4-0	長さの単位「メートル」に当てて用いる。▽(4)は「平米・立米」など。
計	12	348	4	339	703		
対象語: 曲【中難易度語】							
1	0	0	0	0	0	12261-0-0-1-0	まっすぐでない。よこしま。まげる。↔直。「曲折・曲直・曲線」
2	0	0	0	0	0	12261-0-0-2-0	こまかい一つ一つの部分。すみ。くわしい。「委曲」
3	0	0	1	0	1	12261-0-0-3-0	音楽のふし。「何の曲もない」「曲技・曲芸・曲馬・曲節・歌曲」
4	1	87	2	17	107	12261-0-0-4-1	「曲を作る」「楽曲・名曲・秘曲・序曲・編曲・小夜曲・交響曲」
5	0	21	0	3	24	12261-0-0-4-2	音楽上の作品を数える語。
6	0	0	1	0	1	12261-0-0-5-0	詩の一体。「春風馬埭曲」
7	0	1	1	0	2	12261-0-0-6-0	屏風の仕立て方を言う語。「六曲一双」
計	1	109	5	20	135		
対象語: おじ【低難易度語】							
1	0	0	0	0	0	5527-0-0-1-0	▽父や母より年上なら「伯父」、年下なら「叔父」と書く。
2	2	10	1	0	13	5527-0-0-2-0	よその大人の男。▽普通は「おじさん」の形で使う。
計	2	10	1	0	13		
対象語: 線【高難易度語】							
1	2	27	3	2	34	28495-0-0-1-0	せん。糸すじ。糸のように細長く作られたもの。「線を引く」「線香・線路」
2	1	0	0	0	1	28495-0-0-2-0	面の上に描かれるすじ。
3	0	0	0	0	0	28495-0-0-2-1	数学で、点の移動によって生ずる図形。「線条・線画・実線・白線」
4	2	21	5	0	28	28495-0-0-2-2	相接する二つの面のさかい目。「水平線・前線・不連続線・境界線」
5	1	53	1	0	55	28495-0-0-3-0	まっすぐに進み、または一つらなりになっていると、考えたもの。「光線」
6	2	100	6	0	108	28495-0-0-3-1	定期交通機関の経路。「路線・本線」。鉄道の線路。「単線・二番線ホーム」
7	2	0	2	1	5	28495-0-0-3-2a	方針。「この線で行こう」「国策の線に沿う」
8	0	4	1	1	6	28495-0-0-3-2b	水準。「いい線に達した」
9	0	0	1	0	1	28495-0-0-4-0	「身体の線」「脚線」
10	0	1	1	0	2	28495-0-0-5-0	「線の細い人」
計	10	206	20	4	240		
対象語: 単位【高難易度語】							
1	1	24	0	3	28	32181-0-0-1-1	数量を測る基準として定めた量。例、長さを測るメートル・尺等。
2	1	10	1	4	16	32181-0-0-1-2	高等学校以上で、学習の基準量。「卒業するには単位が足りない」
3	0	28	1	6	35	32181-0-0-2-0	ある組織の要素としての一まとまり。「防火単位」
計	2	62	2	13	79		
合計	43	875	48	449	1,415		

表 3: ラベルなしデータのサイズ

	ベスト	仏	きれい	米	曲	おじ	線	単位	Total
ラベルなしデータ	641	1,523	2,071	2,240	1,942	350	7,457	954	17,178

表 4: 実験結果 (1)

Target	ベスト (1)		仏 (8)		きれい (7)		米 (12)		曲 (1)		おじ (2)		線 (10)		単位 (2)		Total (43)	
	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)
MFS	0	0	8	100	5	71.4	12	100	1	100	2	100	2	20	0	0	30	69.8
SVM BL	0	0	8	100	5	71.4	12	100	1	100	2	100	2	20	0	0	30	69.8
MHLE BL	1	100	4	50.0	0	0.0	10	83.3	1	100	1	50.0	3	30.0	2	100	22	51.2
MHLE BLGT+BOW	1	100	8	100	1	14.3	10	83.3	1	100	2	100	7	70.0	2	100	32	74.4

一方、MHLE を利用した場合、BL より、BLGT+BOW の素性を利用する方が精度が高い。SVM や MFS の結果に比べ、「きれい」と「米」以外ではすべて精度が向上している。「きれい」も「米」も、テストデータに最頻語義しか出現していないため、他の語義だと推定した箇所が全て誤りとなってしまったという問題がある。逆に、テストデータに出現する語義が最頻語義以外だった「ベスト」「単位」では、正しい語義の推定が可能となった。全体として、74.4%の精度が得られた。

表 5 上部に、「米」「線」の MHLE(BL+BOW) で誤ったテスト文を示す。表 5 の左端には、[13] の藤田の分析と同様、エラー原因、あるいは、正解するために必要だと考えられる知識を記載した。

例えば、「米」については、現システムでは利用していない素性だが、SRL や並列解析の結果を利用することができれば、正解出来る可能性があると考えた。一方、「線」の場合、ラベルありデータが全くない未知語義や、人間によるラベル付与でも非常に迷うと考えられる例が多い。また、「線」は複合名詞の一部となる場合が多く、MT への利用を考えた場合、複合名詞としてどのように翻訳されるべきかを考慮した訓練データや語義クラスの設定が必要だと感じた。

表 5 には、本章の実験での「きれい」の誤り例は載せていないが、所感として、「きれい」の場合、辞書で定義された語義に分けるのは人間にも非常に難しいと感じた。つまり、このように人間にとって語義付与が難しいものは、機械学習でも難しいと思われる。そこで次章では、システムは同設定のまま、正解語義として、辞書で定義された語義以外を利用することを考える。

表 6: 「きれい」の英訳とクラス番号

Class	English
1	Clean,Cleanly,clear,clearly
2	Beautiful,Beautifully,beautiful,beautifully
3	Complete,Completely,complete,completely
4	Whol,Wholly,whol,wholly
5	Entire,Entirely,entire,entirely
6	Neat,Neatly,neat,neatly

## 6 実験 (2)

前章までは、一般的な WSD のタスク設定と同様、辞書の語義を推定する問題として実験を行った。ただし、本稿の目的は、WSD によって、MT のエラーを改善できるかどうかの調査である。そもそも MT の改善が目的なので、より直接的に訳語を語義とした場合の実験を行った。本章では、前章で最も推定精度が低かった「きれい」を対象語とした実験を行う。

### 6.1 ラベルありデータ (2)

訳語を語義としたラベルありデータは次のように獲得した。

1. 対訳辞書から、対象語 ( $j$ ) の訳語 ( $e$ ) を抽出
2. 各訳語に異なるクラス番号 ( $i$ ) を付与
3. 対訳文で対象語  $j$  を含む文の対訳文に  $e_i$  が含まれる場合、クラス  $i$  のラベルありデータとする

具体的には、まず日英の単語辞書 (ALT-J/E) から、「きれい/綺麗/奇麗」に対する英訳を抽出、派生語を含めて展開し、クラス番号を付与した。その結果得られたクラス番号と英訳の組み合わせは、表 6 の通りである。

次に日英対訳コーパス (3.3 節) から、日本語側が「きれい/綺麗/奇麗」で、英語側が英訳候補にあるも

表 5: エラー例と分析

COR	SYS	文 (一部略)	分析
実験 (1)			
3	1	◆*米*が新規原発を建設の方針	SRL
3	1	中でも日本、ロシア、仏、独、*米*が大口債権国だ。	SRL/並列解析
1	7	今季加わった貴重な戦力は、いま「個の輝き」レベルから*線*としてつながり、「塊」として...	Difficult
2	4	... 特有の体臭成分であるノネナールが、特に体の中心*線*に沿って発生します。	Unseen, BothAreOK
5	4	... 我が国の生命*線*ともいうべき製造業の発展を担い、ものづくりの基盤技術を支える...	Difficult
実験 (2)			
2	6	それを俺が*綺麗*にデザインしてやって、形をつくってブロンズを流す。	SRL

ただし、COR は正解クラス、SYS はシステムが選択したクラスを示す。

表 7: 英訳を語義とした場合のデータサイズ

Class	英訳	CoreA		対訳データから	
		No.	%	No.	%
1	clean	2	28.6	314	39.4
2	beautiful	2	28.6	442	55.5
3	complete	0	0	3	0.4
6	neat	3	42.9	37	4.6
Total		7		796	

表 8: 評価結果 (2)

	きれい (7)
	No. (%)
MFS	2 28.6
SVM BL	4 57.1
SVM BLGT+BOW	2 28.6
MHLE BL	6 85.7
MHLE BLGT+BOW	5 71.4

のを自動的に抽出した。例えば、例 (1) の対訳文<sup>10</sup>の場合、対訳文の日本語側に「きれい」が、英語側に‘clean’が出現するため、日本語側の文をクラス 1 のラベルありデータとして利用する。

- (1) 「きれいなお茶わんがほしーぜ」  
‘I want a clean cup.’

なお、評価データとしては、CoreA に出現する 7 文 (表 2) に、正解のクラスを付与した<sup>11</sup>。表 7 に英訳を語義とした場合のデータサイズを示す。

## 6.2 結果と分析 (2)

表 8 に実験結果を示す。実験 1 の結果 (表 4) と異なり、SVM、MHLE の両方で、BL の方が精度が高くなった。ラベルありデータが 10 倍以上に増え、かつ、語義分類が変わったため、BL のみで高い精度が得られた可能性がある。

MHLE BL では、「きれい」のエラーはひとつだけになった<sup>12</sup>。表 5 下部にエラーとなったテスト文を示す。表 5 のエラー例の場合、「デザイン」が「綺麗」であることが分かれば、正しい解析が可能かもしれない。また、「デザイン」の意味クラスが <1048:絵画,1557:創作 (形状) > であるという情報と組み合わせることで、より正解となる可能性が高くなるかもしれない。

<sup>10</sup> 青空文庫より抜粋

<sup>11</sup> 所感だが、正解タグの付与においても、辞書語義より英訳の方が付与しやすかった

<sup>12</sup> ただし、辞書語義を利用する場合に比べ、クラス数は 8 から 6 に減少、ラベルありデータに出てくるクラス数も 6 から 4 に減少している。

## 6.3 英訳を語義とする場合の問題点

MT への利用を考えた場合、すべての対象語に対し、実験 2 の方法が適用できればよいかもしれない。適切な訳語と対訳コーパスが得られれば、実験 2 の手法で一定の精度が得られると期待できる。ただし、適切な訳語が辞書やコーパスから得られない場合の対処方法を考える必要がある。例えば、「線」の場合、辞書に載っていた英訳は、‘line, route, track, wire’のみであり、MT 班で報告されたエラー例 (表 1) に対し、適切な訳語は得られない。こうした語義に対しどのように対応するかは今後の課題としたい。

## 7 機械翻訳への適用実験

WSD での語義付与結果を MT システムの学習データに利用する実験を行った。具体的には、学習用の対訳コーパス (3.3 節) の対象単語に自動的に語義を付与し、Travatar[8] に基づく forest-to-string 翻訳システム統計的機械翻訳システムを学習した。さらに、評価データに自動的に語義を付与し、単語を語義ごとに区別する場合としない場合の翻訳結果の差分を調査した。なお、語義推定以外の要因の影響を最低限に押さえるために、パラメータ調整を 1 回語義なしのモデルで学習し、語義ありのモデルでも同じパラメータを利用した。語義付与に利用したモデルは、「きれい」以外の語は実験 1 (5 章)、「きれい」は実験 2 (6 章) で最も精度の高かったモデルを利用した。

調査には次の 2 種類の評価データを用いた。

**samp:** 学習データから無作為に抜粋した1,276文。このデータには対訳が付与されているため、翻訳精度の自動評価に用いる。

**core:** 上述の BCCWJ CoreA 評価データ (3,024 文)。このデータには対訳が存在しないため、定性評価のみに用いる。

表 9: 自動評価結果 (samp, BLEU)

	全文
語義なし	33.82
語義あり	33.82

まず、表 9 に samp に対する評価結果を、語義なしと語義ありに分けて示す。この結果から、語義を考慮することによる精度の変化がなかったことが分かる。コーパス全体で変化がなかった理由として、1,276 文中、対象語彙が 5 文にしか現れなかったことが挙げられる。

次に、core に対して定性的な評価と分析を行う。表 10 に、語義のありなしにより翻訳結果に差分があった例をできるだけ示した。

表 10 から、全体的に翻訳結果にあまり大きな変化は見られない。理由として、統計翻訳に利用する場合、対訳コーパスに対する語義付与も高い精度が要求されるにも関わらず、その部分の語義推定があまりうまくいかず、学習時に語義毎に十分なバイヤスをかけられなかったということが考えられる。

例えば、表 10 「仏」の場合、3 例目は翻訳結果が良くなったと言えるが、それ以外の例は、訳語選択には変化がない。また、「米」の 1,2 例目は訳語選択が悪くなっている。これらの例は、WSD で間違った例であり、WSD のエラーによる影響である。「線」の例では、いずれも訳語は変わらないが、構文等に変化が出ており、特に 3 例目は比較的良くなっている。

実験 2 の結果を利用した「きれい」についても傾向は同様である。2 例目は特徴的であり、訳語選択はより良くなったと思われるが、文としては語義なしの方が良い。これは、語義に分けることによって各語義毎の対訳コーパスが減り、文として良い翻訳が得られなかったのではないかと考えられる。

まとめると、WSD の結果を MT に利用して精度向上に貢献するためには、対訳コーパスを含めてほぼ完璧な語義推定が出来る必要があること、非常に高い精度で推定できた場合でも、語義毎の用例数が少なくなるために翻訳精度は必ずしも向上しないことがわかった。

今後は、WSD の精度をより向上させるとともに、語義として分けるだけでなく、同時に同じような用法の語や語義をまとめることで翻訳精度が改善できるか、といったことを調査したい。

## 8 まとめと今後の課題

WSD の実用化を目指すには、精度の向上を目指すと共に、具体的な下流タスクを想定し、必要とされる対象語や語義を検討する必要があると考えている。本稿では、WSD の下流タスクとして MT を想定し、MT 班で報告されたエラー例について、一般的な WSD のタスク設定で解消出来る可能性があるかどうかを検討した (2 章)。その結果、形式名詞を除く名詞、形状詞は辞書の語義を推定する WSD により解消出来る可能性があると考えた。そこで、それらの語を対象語とした場合に、どの程度の精度で語義推定が可能かどうか実験により確かめた。そのため、まず 3 章では、評価・訓練を行うためのデータについて、4 章では、本稿で実験に利用するシステムについて述べた。5 章では、我々のシステムにより、全体で 74.4 % の精度で語義推定ができることを示した。エラー例に対しては、[13] と同様に分析を行った。ただし、一部の対象語については、辞書の語義を推定することは非常に難しかった。そこで、6 章では語義として訳語を利用する方法を提案し、5 章の実験で最も精度の低かった対象語「きれい」を対象に実験を行った。その結果、「きれい」でも 85.7 % の精度が得られた。両方の実験から、語義を選べば高い精度で語義推定が可能であることが示せた。

さらに、7 章では、WSD の結果を統計的機械翻訳に利用する実験を行った。全体として翻訳結果にあまり大きな変化はなかったが、翻訳結果が良くなった例、悪くなった例も存在した。悪くなった理由としては、対訳コーパスに対する語義付与の精度が十分ではなかった、語義に分けることによって各語義ごとの対訳コーパスが減った、などが考えられる。

今後は、WSD の精度をより向上するために、本稿では利用していない特徴である SRL や並列解析などの導入を検討したい。さらに、SRL と同時に解くことができれば、本稿では対象外とした機能語や動詞も今後、対象語に含めることが可能かもしれない。また、語義としての英訳の利用方法をより精練させること、語義として分けるだけでなく、同時に同じような用法の語や語義をまとめることで翻訳精度が改善できるか、といったことを調査したい。

表 10: 翻訳結果の比較

仏	
	とくに、イラク戦争に反対してきた <u>仏</u> 独露から前向きな回答を引き出した意味は大きい。
wo	In particular, it drew a positive response from Russia, <u>France</u> and Germany, which had been opposed to the war in Iraq.
w	There is a big drew a positive response from Russia, <u>France</u> and Germany, in particular, was opposed to the war in Iraq.
(a-5) ドビルバン 仏 外相は十八日、仏 記者団との内輪の懇談で、米欧対立の解消に向け、「フセイン拘束が大いに役立つ」と述べ、独露両国と“反戦連合”を組んだ <u>仏</u> 政府にも、イラク再建に積極関与していく環境が整ってきたことを強調した。	
wo	<u>French</u> Foreign Minister Villepin emphasised 18 in both Japan and Germany and was "anti-war coalition government of <u>Buddha</u> , was actively involved in the reconstruction of Iraq environment conflict resolution at a meeting of the private <u>Buddha</u> reporters," benefit greatly from the capture of Saddam Hussein," he said.
w	<u>French</u> Foreign Minister, Villepin reconciliation, and "benefit greatly from the capture of Saddam Hussein," he said at a meeting of the private <u>Buddha</u> reporters, 18, and emphasized that both Russia and was "anti-war coalition government of <u>Buddha</u> was actively involved in the reconstruction of Iraq.
ベーカー特使の訪仏を歓迎した シラク仏大統領 が、債務問題を協議する場として、欧米、日本の十九か国で構成する主要債権国会議（パリクラブ＝）の多国間の枠組みを改めて念押ししたのも、そのためだった。	
wo	It was because of the <u>President Chirac Baker</u> was welcomed as an opportunity to discuss the debt issue, and multilateral framework of the major Claims Conference (Paris Club-) composed in 19 countries in Europe and America, Japan and was.
w	It was because of the <u>French President Chirac</u> , welcomed Baker, as a place to discuss the debt issue, and was multilateral framework of the major Claims Conference (Paris Club-) composed in 19 countries in Europe and Japan.
米	
(表 5) 中でも日本、ロシア、仏、独、 <u>米</u> が大口債権国だ。	
wo	In the <u>U.S.</u> , Japan, Russia, France, Germany, is a large country of claims.
w	It is a large country of Japan, Russia, France, Germany, <u>rice</u> in particular.
(表 5) ◆ <u>米</u> が新規原発を建設の方針	
wo	◆ <u>U.S.</u> policy on the construction of new nuclear power stations.
w	◆ <u>rice</u> policy for the construction of new nuclear power stations.
米 CNNテレビのプロデューサー、ウェンディ・ウィットワースさんの五十歳の誕生日、夫のラルフさんが“プレゼント”したのは何と、元ビートルズのポール・マッカートニーのプライベートコンサートだった。	
wo	<u>U.S.</u> cable TV producer, Wendy, and Whitworth's fiftieth birthday, her husband Ralph was "present" was a private concert of former Beatles member, Paul McCartney.
w	It was her husband Ralph Whitworth Wendy's fiftieth birthday of <u>U.S.</u> cable TV producer, and "present" was a private concert of former Beatles member, Paul McCartney.
線	
(表 5) この脂肪酸が酸化や皮膚常在菌によって分解されると、特有の体臭成分であるノネナールが、特に体の <u>中心線</u> に沿って発生します。	
wo	The fatty acid decomposed by the oxidation and skin bacteria nonenal, which is a unique characteristic component is particularly along the <u>center line</u> of the body.
w	The fatty acid decomposed by the oxidation and skin bacteria nonenal, which is a component of its body is particularly along the <u>center line</u> of the body.
(表 5) それとともに、我が国の <u>生命線</u> ともいべき製造業の発展を担い、ものづくりの基盤技術を支える実践的な人材の養成も推進していく必要がある。	
wo	Together with the practical training of personnel who supported the foundation of manufacturing technology should be promoted in the development of the manufacturing industry, and can be described as a <u>lifeline</u> of our country.
w	Together, or it is necessary for the development of the manufacturing industry, and promoted practical training of personnel who supported the foundation of manufacturing technology can be described as a <u>lifeline</u> of our country.
慰謝料を搾り取られた恨みという <u>線</u> も。	
wo	<u>Line</u> was sapped the compensation of resentment.
w	He was sapped the compensation <u>line</u> of resentment.
きれい	
そうすると、腸が <u>綺麗</u> になって便秘が改善されて肌が良くなると思います。	
wo	Constipation and the intestines and so do you think is better for my skin.
w	I think so intestines are <u>clean</u> , constipation improved skin will change for the better.
ニューカレドニアのウベア島の海とモルディブの海ではどちらが <u>綺麗</u> ですか??	
wo	In the sea of New Caledonia, Wallis and the Maldives, which is <u>pretty</u> ?
w	In the sea of New Caledonia, Wallis and the sea in the Maldives, <u>beautiful</u> ?
綺麗ですね o ( ? ) o	
wo	<u>Beautiful</u> , isn't it-o ( ? ) o
w	' <u>Anything</u> o ( ? ) o
(a-1/c-2) 3絵をかいたり、色をぬったりして、 <u>きれい</u> にしあげる。	
wo	Do a 3-e, and is not the color and finish.
w	Do a 3-e, and is not the color and finish it.

ただし、wo は「語義なし」w は「語義あり」の場合の結果を示す

## 参考文献

- [1] Francis Bond, Sanae Fujita, and Takaaki Tanaka. The Hinoki Syntactic and Semantic Treebank of Japanese. *Language Resources and Evaluation*, Vol. 40, No. 3-4, pp. 253-261, 2006. (Special issue on Asian language technology).
- [2] Chih-Chung Chang and Chih-Jen Lin. *LIB-SVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [3] Akinori Fujino, Naonori Ueda, and Masaaki Nagata. A robust semi-supervised classification method for transfer learning. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*, pp. 379-388, 2010.
- [4] Sanae Fujita and Akinori Fujino. Word Sense Disambiguation by Combining Labeled Data Expansion and Semi-Supervised Learning Method. *Transactions on Asian Language Information Processing, Association for Computing Machinery (ACM)*, Vol. 12, No. 2, pp. 1-26, 2013.
- [5] 笠原要, 佐藤浩史, Francis Bond, 田中貴秋, 藤田早苗, 金杉友子, 天野昭成. 「基本語意味データベース:lexeed」の構築. 情報処理学会 自然言語処理研究会 (2004-NLC-159), pp. 75-82, 2004.
- [6] Sadao Kurohashi and Kiyotaka Uchimoto. Senseval-2 japanese translation task (自然言語処理特集号「senseval-2 日本語タスク」). 自然言語処理, Vol. 10, No. 3, pp. 25-37, apr 2003.
- [7] Okumura Manabu, Shirai Kiyooki, Kanako Komiya, and Yokono Hikaru. On Semeval-2010 Japanese WSD Task. *Journal of Natural Language Processing*, Vol. 18, No. 3, pp. 293-307, 2011.
- [8] Graham Neubig. Travatar: A forest-to-string machine translation engine based on tree transducers. In *Proceedings of the ACL Demonstration Track*, Sofia, Bulgaria, August 2013.
- [9] 西尾実, 岩淵悦太郎, 水谷静夫. 岩波国語辞典. 岩波書店, 1994.
- [10] 白井清昭. SENSEVAL-2 日本語辞書タスク. 自然言語処理, Vol. 10, No. 3, pp. 3-24, 2003.
- [11] 赤部晃一, Graham Neubig, 工藤拓, John Richardson, 中澤敏明, 星野翔. Project Next における機械翻訳の誤り分析. 言語処理学会第 21 回年次大会 (NLP-2015), 自然言語処理におけるエラー分析ワークショップ, 2015.
- [12] 池原悟, 宮崎雅弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系. 岩波書店, 1997.
- [13] 新納浩幸, 白井清昭, 村田真樹, 福本文代, 藤田早苗, 佐々木稔, 古宮嘉那子, 乾孝司. 語義曖昧性解消の誤り分析. 言語処理学会第 21 回年次大会 (NLP-2015), 自然言語処理におけるエラー分析ワークショップ, 2015.