階層的機能表現辞書の意味的等価クラス および対訳用例を用いた機能表現の日英翻訳

阿部佑亮　(筑波大学大学院システム情報工学研究科)
鈴木敬文　(筑波大学大学院システム情報工学研究科)
宇津呂武仁　(筑波大学システム情報系)*
山本幹雄　(筑波大学システム情報系)
松吉俊　(山梨大学大学院医学工学総合研究部)
河田容英　(株式会社ナビックス)

Japanese to English Machine Translation of Functional Expressions based on Semantic Equivalence Classes of Hierarchical Lexicon and Translation Examples

Yusuke Abe　(University of Tsukuba)
Takafumi Suzuki　(University of Tsukuba)
Takehito Utsuro　(University of Tsukuba)
Mikio Yamamoto　(University of Tsukuba)
Suguru Matsuyoshi　(University of Yamanashi)
Yasuhide Kawada　(Navix Co., Ltd.)

1. はじめに

日本語には 16,000 種類以上の機能表現 (助詞・助動詞・接続詞相当語句) の異形が存在する．日本語機能表現には非常に多様な異形が存在するが，それらの異形を網羅的に正しく翻訳することは難しい．この問題に対応する手法として，先行研究では，日本語機能表現を網羅的に列挙した大規模日本語機能表現階層辞書における機能表現の意味的等価クラスを利用して，日英対訳特許文中に出現する日本語機能表現の日英翻訳を対象として，日本語機能表現の集約的な日英機械翻訳を行う手法を提案している．この手法を 53 の意味的等価クラスに適用した結果，20 クラスについては，意味的等価クラスに属する日本語機能表現の翻訳規則を 1 規則ないし 2 規則に集約出来ることが分かった．しかし，一方で，他の 33 クラスについては，意味的等価クラスに属する日本語機能表現の翻訳規則を集約することが出来なかった．これは，日本語機能表現を英訳する際の曖昧性のためであった．より正確な翻訳を行うためには，これら機能表現表記のもつ曖昧性を考慮した翻訳の仕組みが必要不可欠である．

以上を踏まえて，本論文では，NTCIR-7 の特許翻訳タスクで配布された 1,798,571 件の日英対訳特許文対から得た対訳用例を用いて，日本語機能表現を英訳する方式を提案する．この方式においては，機能表現の意味的等価クラスごとに，様々な対訳用例からデータベースを構築し，英訳対象となる機能表現表記の用例と最も類似した対訳用例の訳語を適用することで，上記の曖昧性に対応する．評価実験として，句に基づく統計的機械翻訳モデル Moses [Koehn07] を，日英対訳特許文を用いて訓練したものとの翻訳精度比較を行った．両手法の作成時に参照するテキストと同ジャンルである特

*utsuro @ iit.tsukuba.ac.jp

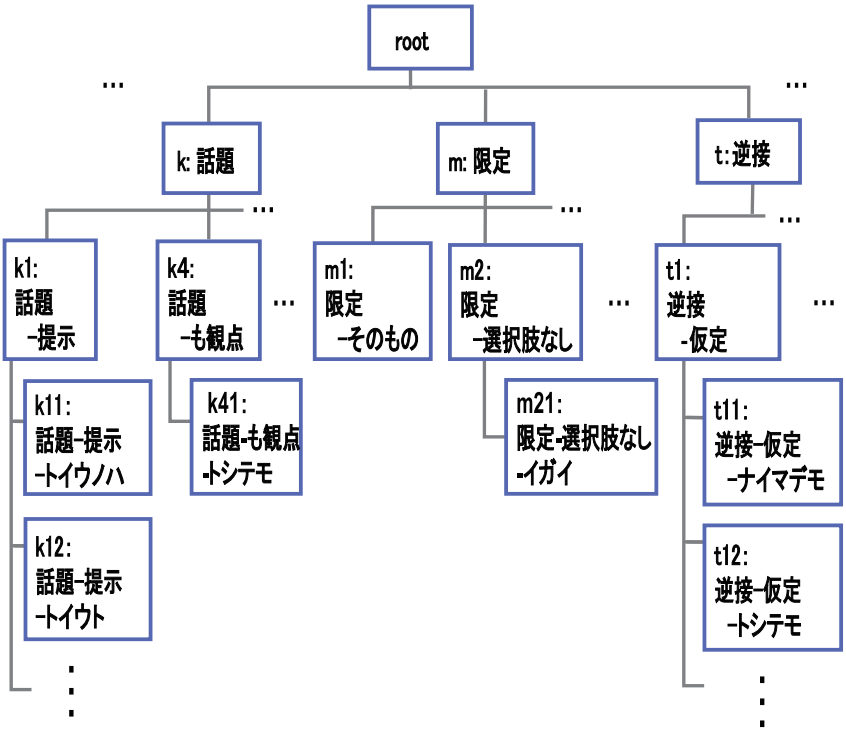

図 1: 意味的等価クラスに基づく階層構造

許文における翻訳精度は，多くの意味的等価クラスにおいて Moses の方が優れていたが,「日本語書き言葉均衡コーパス」および「日本語学習者用用例集」における翻訳精度は，多くの意味的等価クラスにおいて提案手法の方が優れていた．このことから，対訳用例を選定したテキストとは異なるジャンルのテキストにおける英訳においても，提案手法は比較的安定した翻訳性能を示すことを実証できた．

2. 階層的日本語機能表現辞書

[松吉 07, 松吉 08] は，日本語の機能表現の異型を，機能表現の構成要素の組み合わせとして階層的に収録した辞書を編纂した (日本語機能表現一覧「つつじ」[1])．日本語機能表現一覧「つつじ」には，16,801 の機能表現が収録されており，この辞書によって，日本語機能表現の網羅的取り扱いが可能になった．

また，日本語機能表現一覧「つつじ」では，図 1 に示すように，辞書に収録されている見出し語について，意味的等価クラスという形での階層的分類も行っている．この最下層に位置する全 199 個の意味的等価クラスについて，同一クラス内の機能表現は，日本語文中で言い換え可能であるとされている [松吉 08]．この意味的等価クラスを用いることにより，日本語機能表現の言い換え候補を網羅的に取り扱うことが可能となった．

3. 機能表現表記の曖昧性

日本語機能表現表記の適切な英訳を行うためには，日本語機能表現表記の持つ曖昧性に対応する必要がある．日本語機能表現表記を英訳するにあたって，対応すべき曖昧性は，大きく分けて 3 種類

[1] http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/

表 1: 機能表現表記の曖昧性の例

(a) 機能的用法/内容的用法の曖昧性

	表記	例文	用法
(1)	ものの	乾燥に供した加熱空気は蒸発した水蒸気を含み，多くの熱エネルギーを持っている ものの ，回収して循環利用するには限界があり，多くの場合廃棄されている．	機能的用法 t24(逆接-確定-モノノ) (～ ものの= *although* ～)
(2)	ものの	ここで，ブロックが存在しない場合は，探索対象段の位置を，保持されたアベイラブルエリアで最後の ものの 左上隅点とし（ステップ１１０６），その後，後述する図１２に示される処理を実行する。	内容的用法 (～ ものの = *of*)

(b) 複数の機能的用法間の曖昧性

	表記	例文 (英訳文)	用法
(3)	としても	このため、誤って装置に物等を落下した としても ，その衝撃は反射ミラー８ｆに伝わり難くなっている．	機能的用法 t12(逆接-仮定-トシテモ) (としても= *even when*)
(4)	としても	さらに，ブレード４５は接触ローラ３７の外周面３７ａの汚れを除去するクリーニング手段 としても 作用する．	機能的用法 k41(話題-も観点-トシテモ) (としても= *as*)

(c) 対訳英語の曖昧性

	表記	例文 (英訳文)	用法
(5)	による	原稿台１１側からの光のミラー１４ による 反射光路上には結像レンズ１６とプラテン２０がこの順に配置されている。	機能的用法 c11(仲介-原因-ニヨッテ) (による = *by*)
(6)	による	本発明 による 可変差動制限装置２の制御は、以下の (1), (2), (3) の３種の制御の組合せから構成される。	機能的用法 c11(仲介-原因-ニヨッテ) (による= *according to*)
(7)	による	つまり、放電開始 による 電圧の低下が、極間異常状態と判定されてしまうことがある。	機能的用法 c11(仲介-原因-ニヨッテ) (による= *due to*)

ある．1 つは，文中の表現が機能表現の意味として用いられているもの (機能的用法) と，その表現を構成する語本来の意味で用いられているもの (内容的用法) との間の曖昧性である (表 1 (a))．もう 1 つは，機能表現の意味が文脈によって異なるという機能的用法の曖昧性である (表 1 (b))．そして最後の 1 つは，対訳英語の曖昧性である (表 1 (c))．

4. 対訳用例データベースの構築

本論文では，NTCIR-8 の特許翻訳タスク [Fujii10] で配布された日英対訳特許文の文対応データのうち，1,798,571 件をフレーズテーブルの訓練用データとして使用した．この文対応データに対して，句に基づく統計的機械翻訳モデルのツールキットである Moses [Koehn07] を適用し，日英の句の組および日英の句の組が対応する確率を示したフレーズテーブルを作成する．

このフレーズテーブルを用いて，先の約 180 万件の日英対訳特許文対から，対訳用例データベースを構築した．その構築手順を，図 2 に示す．具体的には，対訳用例データベースの構築対象としている意味的等価クラスに属する各日本語機能表現表記について，以下の条件を満たす「日本語機能表現表記- 英訳語」組をフレーズテーブルから抽出する．

- 日英対訳特許文対における日本語機能表現表記の出現頻度が 20 以上
- Moses によって，日英対訳特許文対における「日本語機能表現表記 - 英訳語」組が句対応していると自動判定された箇所の頻度が 10 以上
- フレーズテーブルにおける日英翻訳確率が 0.05 以上

そして，抽出した各「日本語機能表現表記 - 英訳語」組について，この表記および英訳語が対応関係であると人手で判断された対訳文対を，約 180 万件の日英対訳特許文対から収集し，対訳用例デー

表 2: 対訳用例データベースを構築した意味的等価クラス

意味的等価クラス			表現数	表現の例
日本語機能表現表記の用法の曖昧性	大きい	M11(不必要 - 不必要 - ナクテヨイ)	299	なくてもよい，までもない，ずともよく
		P11(例示 - 程度 - クライ)	6	くらい，ばかり，ほど
		c11(仲介 - 原因 - ニヨッテ)	15	により，をもって，によります
		m12(限定 - そのもの - ノミ)	5	きり，だけ，のみ
		n12(添加 - 非限定 - ダケデナク)	12	のみならず，だけじゃなく，上に
		s11(理由 - 因状況 - イジョウハ)	9	からには，うえは，以上
		t12(逆接 - 仮定 - トシテモ)	21	にしても，としましても，たところで
	小さい	D11(判断 - 当為 - ナケレバナラナイ)	213	ないといけない，ねばならない，べき
		b11(対象 - 関連 - ニツイテ)	26	に関する，について，につきまして
		u12(対比 - 般 - カワリニ)	14	代わりに，代りに，かと思うと

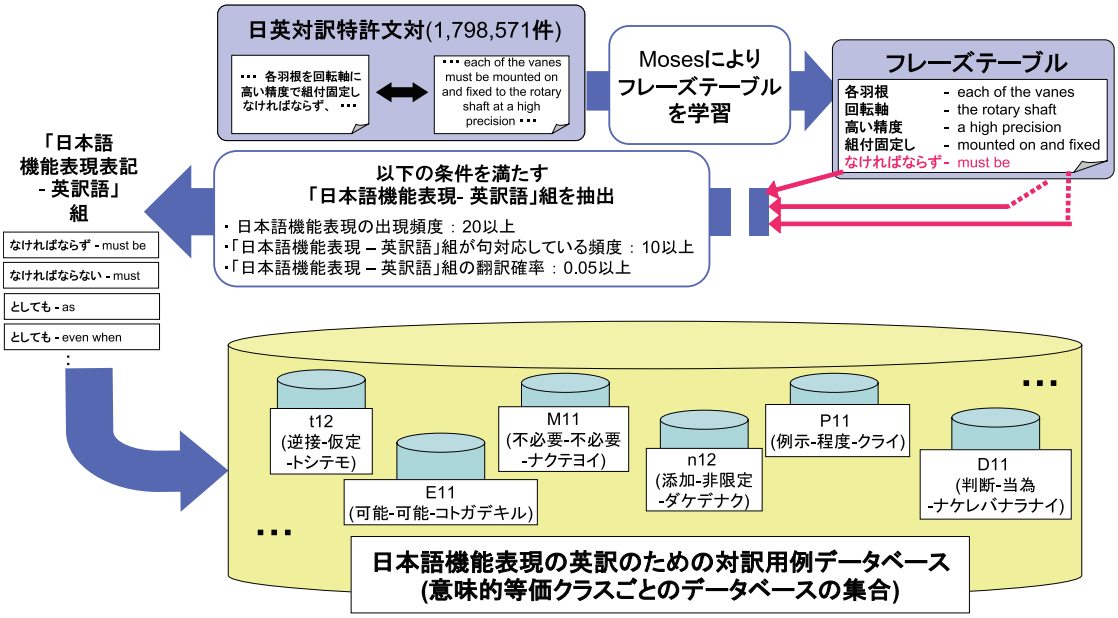

図 2: 意味的等価クラスごとの対訳用例データベースの構築手順

タベースへ登録する．ただし，本研究では，意味的等価クラスごとのデータベースを用意し，当該日本語機能表現が属する意味的等価クラスのデータベースにのみ，対訳用例を追加する．この構築手順に従って，表 2 に示した 10 の意味的等価クラスの対訳用例データベース構築を行った．その結果，表 3 に示すように，10 クラス合計で 5,253 用例の対訳用例データベースを構築することができた．

5. 対訳用例を用いた機能表現の日英翻訳

対訳用例を用いた機能表現の日英翻訳においては，翻訳対象である日本語用例が与えられると，まず，用例中の機能表現が属する意味的等価クラスの対訳用例データベースが参照される．次に，用例間の類似度に基づいて，与えられた用例中の日本語機能表現表記と用法の最も類似した対訳用例を選択する．そして，その対訳用例における英訳語を，翻訳対象の日本語機能表現表記の英訳語として採用する．

以下，まず，入力された日本語用例を $e_j = \langle m_{pre}, M_c, m_{suf} \rangle$，その日本語用例 e_j 中の日本語機能表現表記を $f_j(e_j)$ とする．ただし，m_{pre}，m_{suf} はそれぞれ，日本語機能表現表記に前接する，あるいは，後接する形態素を表し，M_c は，日本語機能表現表記を構成する形態素列を表す．また，データベース中のある用例を $e_{je}^{db} = \langle e_j^{db}, t_e^{db} \rangle$ とする．ただし，e_j^{db} は，データベース中の用例の日本語部分を，t_e^{db} は e_j^{db} 中の機能表現部分の英訳を，それぞれ表す．ここで，日本語用例 e_j 中の日本語機能

表 3: 対訳用例データベース中の日本語機能表現表記数・用例数

意味的等価クラス			日本語表記数	「日本語機能表現表記 - 英訳語」組数	用例数
日本語機能表現表記の用法の曖昧性	大きい	M11(不必要 - 不必要 - ナクテヨイ)	5	16	96
		P11(例示 - 程度 - クライ)	3	5	113
		c11(仲介 - 原因 - ニヨッテ)	5	15	1489
		m12(限定 - そのもの - ノミ)	2	4	773
		n12(添加 - 非限定 - ダケデナク)	6	12	797
		s11(理由 - 因状況 - イジョウハ)	2	5	494
		t12(逆接 - 仮定 - トシテモ)	6	16	778
		上記 7 クラスの合計	29	73	4540
	小さい	D11(判断 - 当為 - ナケレバナラナイ)	6	9	114
		b11(対象 - 関連 - ニツイテ)	6	19	455
		u12(対比 - 般 - カワリニ)	4	7	144
		上記 3 クラスの合計	16	35	713
合計			45	108	5253

表現表記 $f_j(e_j)$ の属する意味的等価クラスの集合を $SS(f_j(e_j))$ とし[2]，$SS(f_j(e_j))$ 中の一つの意味的等価クラスを S，S の対訳用例データベースを EDB_S とすると，参照すべき対訳用例全体の集合は，以下の $EDB(e_j)$ で表される．

$$EDB(e_j) \quad = \bigcup_{S \in SS(f_j(e_j))} EDB_S$$

また，日本語用例 e_j と対訳用例データベース中の用例 e_{je}^{db} との類似度 $Sim\big(e_j, e_{je}^{db} = \langle e_j^{db}, t_e^{db} \rangle\big)$ は，以下で定義される．

$$\begin{aligned} Sim\Big(e_j, e_{je}^{db} = \Big\langle e_j^{db}, t_e^{db} \Big\rangle\Big) \quad = \quad & Sim_{pre}\Big(m_{pre}(e_j), m_{pre}(e_j^{db})\Big) + \ Sim_c\Big(M_c(e_j), M_c(e_j^{db})\Big) \\ & + \ Sim_{suf}\Big(m_{suf}(e_j), m_{suf}(e_j^{db})\Big) \end{aligned}$$

ここで，Sim_{pre} および Sim_{suf} はそれぞれ，e_j，e_j^{db} の前接形態素，および，後接形態素の類似度であり，Sim_c は，e_j の構成形態素列と e_j^{db} の構成形態素列の類似度である．厳密には，Sim_{pre} および Sim_{suf} はそれぞれ，前接形態素 $m_{pre}(e_j)$，$m_{pre}(e_j^{db})$，および後接形態素 $m_{suf}(e_j)$，$m_{suf}(e_j^{db})$ の品詞・活用形の一致数を正規化した値を表し，Sim_c は，構成形態素列 $M_c(e_j)$，$M_c(e_j^{db})$ の品詞・活用形の一致数 [3]を正規化した値を表している．図 3 は，用例間の類似度 $Sim\Big(e_j, e_{je}^{db} = \Big\langle e_j^{db}, t_e^{db} \Big\rangle\Big)$ の定義を，具体例で図示したものである．この図中の 2 つの用例の場合，前接・後接形態素および構成形態素列の品詞・活用形が，いずれも完全に一致しているため，Sim_{pre}，Sim_{suf} および Sim_c はいずれも 1 となり，この 2 つの用例間の類似度 $Sim\Big(e_j, e_{je}^{db}\Big)$ は，最大値である 3 となる．

以上に基づき，参照すべき対訳用例全体の集合中で最も類似する対訳用例に基づいて機能表現の翻訳を行う関数 $tran_{fe}\big(f_j(e_j)\big)$ は，以下で定義される．

$$tran_{fe}\Big(f_j(e_j)\Big) \quad = \quad t_e^{db} \quad s.t. \ \Big\langle e_j^{db}, t_e^{db} \Big\rangle = \underset{e_{je}^{db} \in EDB(e_j)}{\operatorname{argmax}} \ Sim\Big(e_j, e_{je}^{db} = \Big\langle e_j^{db}, t_e^{db} \Big\rangle\Big)$$

提案手法による訳語選択の例を，図 4 に示す．

[2]ここでは，一つの機能表現表記が複数の意味的等価クラスに属する場合があることを想定した定式化となっている．

[3]2 つの構成形態素列 $M_c(e_j)$，$M_c(e_j^{db})$ の形態素数が異なる場合，形態素列の前側から順に形態素を対応付け，品詞・活用形の一致数を計上していき，形態素列が長い方について，対応付ける形態素が無い場合，その形態素との品詞・活用形の一致数は 0 とする．

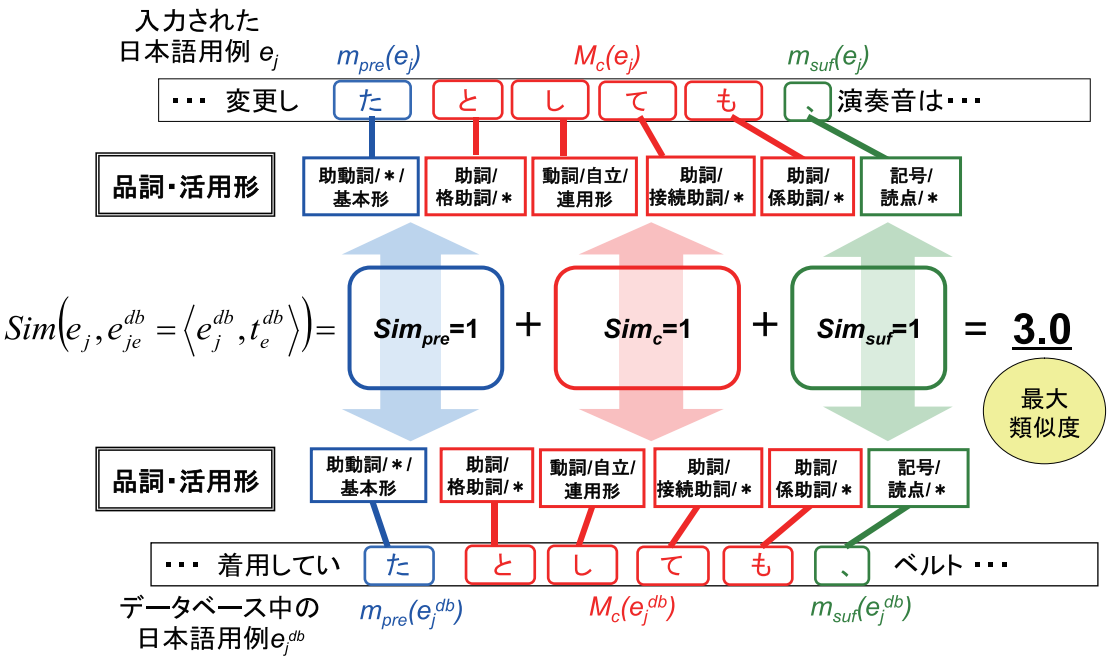

図 3: 機能表現表記の用例間の類似度

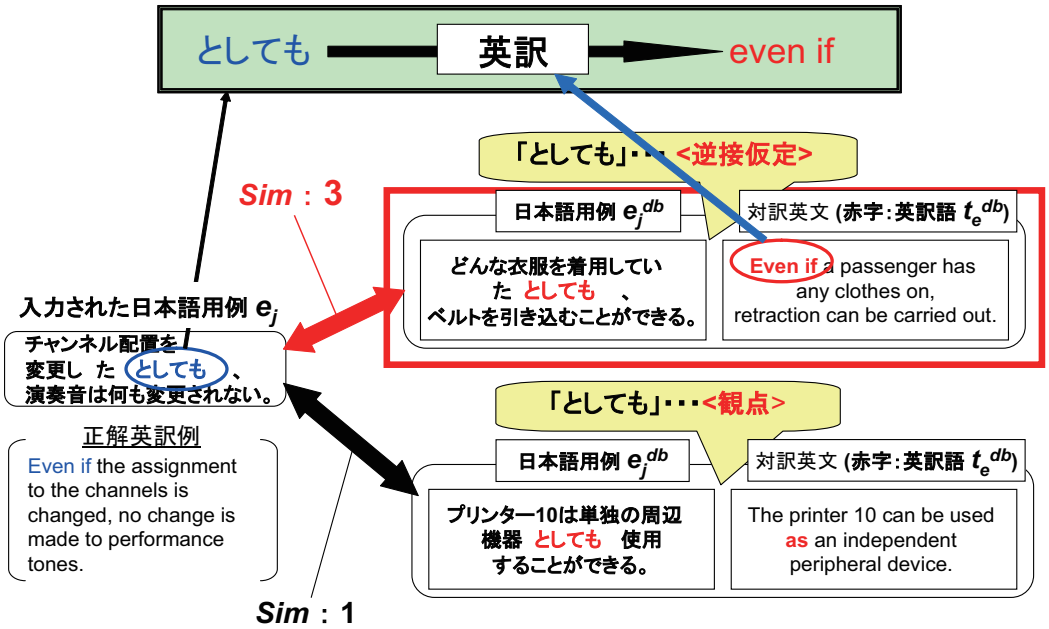

図 4: 日本語機能表現表記の用例間の類似度を用いた訳語選択

6. 評価

提案手法の評価として，提案手法および句に基づく統計的機械翻訳モデルである Moses [Koehn07] を日英対訳特許文を用いて訓練したものについて，翻訳精度の比較を行った．評価は，4 節で述べた手順に従って対訳用例データベースを構築した 10 の意味的等価クラスを対象として行った．

評価文は，NTCIR-8 の特許翻訳タスク [Fujii10] で配布された日英対訳特許文のうちの約 140 万件 (以下,「特許文」),「現代日本語書き言葉均衡コーパス」 [BCCWJ 総括班 11](以下,「書き言葉コー

表 4: 評価文の収集対象となるテキストの内訳

テキストの種類	テキストの特徴	文数
日英対訳特許文 (特許文)	2001-2007 年発行の日本公開特許および，それと対応する米国公開特許の文書について，互いに対訳関係にある度合いの高い部分として「背景」および「実施例」の部分を抽出し，日英間で文対応を付けられている．	1,387,713
現代日本語書き言葉均衡コーパス (書き言葉コーパス)	新聞や小説，雑誌，ウェブページ，国会議事録等，さまざまなジャンルの文書から収集された文を収録している．	4,212,638
日本語学習者用用例集	日本語学習者向けに，日常会話文を中心とした文を収録している．	9,125

表 5: 評価結果

評価対象		評価文数	翻訳精度 (%)			
			ベースライン	Moses	提案手法	提案手法 (上限値)
テキストジャンル別	特許文	140	53	**66**	65	70
	書き言葉コーパス	292	53	34	**77**	78
	日本語学習者用用例集	168	51	42	**56**	62
評価文集合別	評価文集合 (1) (提案手法：英訳が容易，Moses：英訳が容易)	210	54	**66**	**66**	73
	評価文集合 (2) (提案手法：英訳が容易でない，Moses：英訳が容易)	180	48	47	**54**	58
	評価文集合 (3) (提案手法：英訳が容易，Moses：英訳が容易でない)	90	58	31	**74**	82
	評価文集合 (4) (提案手法：英訳が容易でない，Moses：英訳が容易でない)	120	55	20	**63**	65
合計		600	53	46	**63**	69

パス」)，および,「日本語学習者用用例集」 [グループ・ジャマシイ 98]，の 3 種類のテキストから収集した．各テキストの特徴および文数を，表 4 に示す．また，評価文は,「対訳用例データベース中の各用例との類似度の最大値 $Sim_{max}(e_j)$ [4]の範囲」および「評価文 e_j 中の日本語機能表現表記 $f_j(e_j)$ が Moses の学習データである約 180 万件の対訳特許文中に出現するか」の 2 つの尺度で分類した．前者に関しては，$2.33 \leq Sim_{max}(e_j) \leq 3$ を満たす評価文 e_j を「提案手法での英訳が容易」，$0 \leq Sim_{max}(e_j) < 2.33$ を満たす評価文 e_j を「提案手法での英訳が容易でない」と，それぞれ判断した．同様に後者に関しては，日本語機能表現表記 $f_j(e_j)$ が Moses の学習データである約 180 万件の対訳特許文中に出現する評価文 e_j を「Moses での英訳が容易」，出現しない評価文 e_j を「提案手法での英訳が容易でない」と，それぞれ判断した．この 2 つの尺度に従って，各意味的等価クラスについて 4 種類の評価文集合を作成した．また，評価文は，各評価文集合から選定された評価文の数が均等になるように選定した．

評価結果を表 5 に示す.「提案手法」,「Moses」は，それぞれの手法での翻訳精度を示しており,「提案手法 (上限値)」は，提案手法において複数の英訳語が出力され，その中に評価文中の機能表現表記

[4] $Sim_{max}(e_j) = \max_{e_{je}^{db}} Sim\left(e_j, e_{je}^{db} = \left\langle e_j^{db}, t_e^{db} \right\rangle\right)$ と定義する．

に対して適切な英訳語と不適切な英訳語が混在する場合，その中から適切な英訳のみを選択し，出力した場合の翻訳精度を示している.「ベースライン」は，各意味的等価クラスの対訳用例データベースにおける対訳用例の英訳語 t_e^{db} のうち，頻度最大のものを英訳語として出力した場合の精度である．評価の結果，評価文書集合全体では，提案手法の方が Moses よりも優れた翻訳精度となった．また，ベースラインと比較して，どの評価文集合においても，提案手法の方が翻訳精度が高いことから，提案手法によって機能表現表記の曖昧性に対応した翻訳が行われていることが分かる．

テキストのジャンル別の翻訳精度を見てみると，両手法の作成時に参照するテキストと同ジャンルである「特許文」では，Moses の方が精度が高いが，それ以外のジャンルのテキストでは，提案手法の方が精度が高いことが分かる．このことから，提案手法は，特定のジャンルのテキストから対訳用例を集めても，それとは異なるジャンルのテキストへ適応できる可能性が高いことが分かる．通常，一般的なジャンルにおいては，二言語対訳コーパスの作成は高コストであり，使用できるものが限られるため，この特性は重要であると考えられる．

評価文中の日本語機能表現表記が訓練データ中に出現しない評価文集合 (3)，(4) における翻訳精度の比較結果をふまえると，Moses は訓練データから作成したフレーズテーブルを用いて翻訳を行うため，訓練データ中に出現しない表記の翻訳を行うことが困難であることが分かる．一方，提案手法は用例間の類似度として品詞・活用形の情報のみを利用しているため，訓練データ中に出現しない表記の翻訳にも対応することができている．

7. 関連研究

本論文の関連研究としては，大きく分けて「日本語機能表現の検出および解析に関する研究」と「機械翻訳に関する研究」の 2 種類がある．以下，それぞれについて述べる．

7.1 日本語機能表現の検出および解析に関する研究

機能表現の検出および解析を目的とした研究としては [土屋 07,注連 07,小早川 09,鈴木 12b] が，機能表現を考慮した係り受け解析に関する研究としては [注連 07] が，それぞれ挙げられる．このうち，機能表現の検出を目的とした研究は，機能表現表記の持つ曖昧性のうち，機能的用法/内容的用法の曖昧性の解消を行う研究と言える．特に， [鈴木 12a] においては，階層的日本語機能表現辞書 [松吉 07] の機能表現の間の階層的関係を利用して，上層に位置する代表的な機能表現の用例を利用して，その派生表現にあたる機能表現の検出を行う手法を提案している．

7.2 機械翻訳に関する研究

本論文で採用した**用例ベース翻訳** [Nagao84,佐藤 92,Sommers03] の枠組みは，大きく分けて，**対訳用例データベース**と**用例を参照するための用例間の類似度**の 2 つの要素から構成される．用例ベース翻訳の代表的研究の一つとして，[佐藤 91] では，翻訳対象を動詞とその必須格の名詞からなる動詞フレームに限定して，用例ベースの英日翻訳を行っている．具体的には，対訳用例を動詞フレームという枠で扱い，注目する動詞の必須格である名詞の出現頻度の傾向が類似する動詞フレームを参照している． [村田 01] では，日英翻訳における動詞の時制や進行形，完了形や助動詞の選択を，用例に基づいて行っている．参照する用例を決定するための類似度として，文末からの一致文字列の長さという，簡便なものを用いているが，分類語彙表 [国立国語研究所 64] の分類番号や語の変化形といった情報も利用している．

その他には，代表的表現への言い換えを介した機械翻訳の研究として，内容語と口語的な機能表現を扱った [山本 01,山本 02] が知られている．同様に，階層的日本語機能表現辞書 [松吉 07] の機能表

現を対象として，代表的表現への言い換えを介して機械翻訳を行う手法の研究として,「日本語学習者用用例集」 [グループ・ジャマシイ 98] 中の例文を対象とした集約的英訳についての研究事例 [坂本 09]，特許文を対象とした集約的英訳についての研究事例 [島内 11]，および，集約的中国語訳についての研究事例 [劉 10] がある．

8. おわりに

本論文では，対訳用例に基づく日本語機能表現表記の英訳手法を提案した．対訳用例データベースは，約 180 万件の日英対訳特許文より収集した．意味的等価クラス別の対訳用例データベースとすることで，参照する用例の探索範囲を，用法の類似する用例に限定した．さらに，機能表現表記の用法の区別をせずに対訳用例を収集し，品詞・活用形の情報に基づく用例間の類似度を利用して参照する用例を選択することで，機能表現表記の持つ曖昧性への対応も可能とした．評価実験として，句に基づく統計的機械翻訳モデル Moses [Koehn07] を日英対訳特許文を用いて訓練したものとの翻訳精度比較を行った．両手法の作成時に参照するテキストと同ジャンルである特許文における翻訳精度は，多くの意味的等価クラスにおいて Moses の方が優れていたが,「日本語書き言葉均衡コーパス」および「日本語学習者用用例集」における翻訳精度は，多くの意味的等価クラスにおいて提案手法の方が優れていた．このことから，対訳用例を選定したテキストとは異なるジャンルのテキストにおける英訳においても，提案手法は比較的安定した翻訳性能を示すことを実証できた．

今後の課題としては，対訳用例データベース構築の手順において，対訳用例収集の対象とする「日本語機能表現表記 - 英訳語」組の条件を修正するか，あるいは,「日本語機能表現表記 - 英訳語」組の人手による選定を行い，より多くの英訳語を含むようデータベースを拡張することが挙げられる．また，他の意味的等価クラスについても同様に対訳用例データベースを構築し，その翻訳性能を検証することも挙げられる．そして，Moses のような統計的機械翻訳の手法との併用によって，翻訳精度向上が可能であると考えられるので，機械学習の導入により，それを実現することを目指す．

参考文献

[Fujii10] Fujii, A., Utiyama, M., Yamamoto, M., Utsuro, T., Echizen-ya, H., Ehara, T. and Shimohata, S.: Overview of the Patent Translation Task at the NTCIR-8 Workshop, *Proc. 8th NTCIR Workshop Meeting*, pp. 371–376 (2010).

[グループ・ジャマシイ 98] グループ・ジャマシイ（編）：教師と学習者のための日本語文型辞典, くろしお出版 (1998).

[小早川 09] 小早川健, 関場治朗, 木下明徳, 熊野正, 加藤直人, 田中英輝：単語格子とマルコフモデルによる日本語機能表現の解析 — 日本語機能表現辞書「つつじ」を用いて —, 電子情報通信学会技術研究報告, NLC2009-1, pp. 15–20 (2009).

[Koehn07] 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.: Moses: Open Source Toolkit for Statistical Machine Translation, *Proc. 45th ACL, Companion Volume*, pp. 177–180 (2007).

[BCCWJ 総括班 11] 文部科学省科学研究費特定領域研究「日本語コーパス」総括班: 特定領域研究「日本語コーパス」研究成果報告 (2011).

[劉 10] 劉颯, 長坂泰治, 宇津呂武仁, 松吉俊：意味的等価クラスを用いた日本語機能表現の集約的日中翻訳規則の作成と分析, 言語処理学会第 16 回年次大会論文集, pp. 194–197 (2010).

[松吉 07] 松吉俊, 佐藤理史, 宇津呂武仁：日本語機能表現辞書の編纂, 自然言語処理, Vol. 14, No. 5, pp. 123–146 (2007).

[松吉 08] 松吉俊, 佐藤理史：文体と難易度を制御可能な日本語機能表現の言い換え, 自然言語処理, Vol. 15, No. 2, pp. 75–99 (2008).

[村田 01] 村田真樹, 馬青, 内元清貴, 井佐原均：用例ベースによるテンス・アスペクト・モダリティの日英翻訳, 人工知能学会論文誌, Vol. 16, No. 1, pp. 20–28 (2001).

[Nagao84] Nagao, M.: A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, Elithorn, A. and Banerji, R. (eds.), Artificial and Human Intelligence, Elsevier Science Publishers, B.V (1984).

[国立国語研究所 64] 国立国語研究所：分類語彙表, 秀英出版 (1964).

[坂本 09] 坂本明子, 宇津呂武仁, 松吉俊：日本語機能表現の集約的英訳, 言語処理学会第 15 回年次大会論文集, pp. 654–657 (2009).

[佐藤 91] 佐藤理史：MBT1: 実例に基づく訳語選択, 人工知能学会誌, Vol. 6, No. 4, pp. 592–600 (1991).

[佐藤 92] 佐藤理史：実例に基づく翻訳, 情報処理, Vol. 33, No. 6, pp. 673–681 (1992).

[島内 11] 島内蘭, 阿部佑亮, 鈴木敬文, 宇津呂武仁, 松吉俊：特許文における日本語機能表現の集約的英訳規則の作成と評価, 言語処理学会第 17 回年次大会論文集, pp. 396–399 (2011).

[注連 07] 注連隆夫, 土屋雅稔, 松吉俊, 宇津呂武仁, 佐藤理史：日本語機能表現の自動検出と統計的係り受け解析への応用, 自然言語処理, Vol. 14, No. 5, pp. 167–197 (2007).

[Sommers03] Sommers, H.: An Overview of EBMT, Carl, M. and Way, A. (eds.), Recent Advances in Example-Based Machine Translation, pp. 3–57, Kluwer Academic (2003).

[鈴木 12a] 鈴木敬文, 阿部佑亮, 宇津呂武仁, 松吉俊, 土屋雅稔：『現代日本語書き言葉均衡コーパス』における複合辞の検出と評価, 『コーパス日本語学ワークショップ』予稿集 (2012).

[鈴木 12b] 鈴木敬文, 阿部佑亮, 宇津呂武仁, 松吉俊, 土屋雅稔：代表・派生関係を利用した日本語機能表現の解析方式の評価, 言語処理学会第 18 回年次大会論文集 (2012).

[土屋 07] 土屋雅稔, 注連隆夫, 高木俊宏, 内元清貴, 松吉俊, 宇津呂武仁, 佐藤理史, 中川聖一：機械学習を用いた日本語機能表現のチャンキング, 自然言語処理, Vol. 14, No. 1, pp. 111–138 (2007).

[山本 01] 山本和英, 白井諭, 坂本仁, 張玉潔：SANDGLASS: 両言語換言機構を基軸とする音声翻訳, 言語処理学会第 7 回年次大会発表論文集, pp. 221–224 (2001).

[山本 02] 山本和英：換言と言語変換の協調による機械翻訳モデル, 言語処理学会第 8 回年次大会発表論文集, pp. 307–310 (2002).