2007－NL－177 (3)
2007／1／26

Web 上のテキスト情報と翻訳モデルを利用した翻訳品質評価法の検討

宮下広平†*，山本誠一††，安田圭志‡††，柳田益造†

†同志社大学
‡ATR 音声言語コミュニケーション研究所
††NICT

日英方向の機械翻訳による訳文の正解訳と誤訳を分類する手法として，先に筆者らは Web 上に存在する膨大な英文の用例に着目し，Web 上で翻訳文が検索ヒットするか否かにより正解訳と誤訳を分類する手法を提案している．提案手法では，訳文の流暢さ（Fluency）と適切さ（Adequacy）により，翻訳文を正解訳と 3 つの誤訳クラスに分類し，正解訳と各誤訳クラスの検索ヒット率が，文中に含まれる単語数に応じてどのように異なるか等について検証している．その結果，対象とする分野の会話文に適合した機械翻訳を用いれば，提案手法を用いることでかなりの割合で正解訳と誤訳に分類可能であることを示している．本稿では，先に提案した手法の幾つかの課題を指摘すると共に，解決策として，単語クラスの導入などの課題の対処法及び，翻訳モデルを用いた手法を組み合わせる手法を提案している．更にこの改良された手法により，正解訳と誤訳の判別性能が向上することを示している．

Quality Evaluation Method of Machine Translated Sentences by Comparing Text Retrieved from Web and Using Translation Model

Kohei Miyashita†*, Seiichi Yamamoto††, Keiji Yasuda‡††, and Masuzo Yanagida†

†Doshisha University
‡ATR Spoken Language Translation Research Labs.
††NICT

This paper proposes a method for evaluating quality of machine translated sentences by comparing them with enormous amount of texts retrieved from Web. Conducted in the paper is an experiment on investigating how many sentences exactly matched to correct translations and incorrect translations are retrieved from Web using translations from several translation systems. Authors' approach to classify correct and incorrect translations is proved to be effective for medium length sentences consisting of five to nine words. Some problems of the proposed method are pointed and methods for resolving these problems are proposed: One is to use word classes, and the other is to introduce a translation model. Additionally pointed is that the distinction performance of correct translations and incorrect translations is improved using the proposed method.

1 はじめに

経済のグローバル化の進展に伴い，英語は国際共通語としての役割を担う傾向が加速し，特定の分野だけではなく様々な分野で英語を使って意見を述べたり情報を発信したりすることが重要になってきている．

英語学習者の分野の広がりに伴い，学習者のニーズに合わせて対象分野を限定し，その分野特有の学習内容を提示することにより学習効果を高める学習法である ESP(English for Specific Purposes)が注目されている．さらに ESP においても，対象分野での英語での発信能力を向上させることが重要な課題の 1 つとなりつつある．

音声による英語での発信能力を向上させる手段の 1 つに，「英語対話システム」を用いた学習が考えられる．筆者らは，英語学習者がコンピュータと様々な課題について英語

* この研究の一部は ATR 音声言語コミュニケーション研究所で行われた．

で対話を行う際に，学習者の英語の能力測定を行い，能力に応じて課題を変更することにより，発信能力の向上を支援する「英語対話システム」の開発を進めている．「英語対話システム」において，目標分野毎に対話シナリオを作成することは，多くの人的資源，費用，時間を要するために，如何にして開発を効率的に行うかが課題になる．この課題に対して，既存の特定分野の英語対話システムを対象分野に移植する方法や日本語対話システムをベースに英語対話システムを開発することとし，その際の効率的な開発手段として，機械翻訳などの自然言語処理技術を使用することが考えられる．

現在，Web 上の翻訳サービスの普及や，大規模な対訳コーパスの開発を基盤とした統計翻訳システム，用例翻訳システムによる高品質な翻訳の実現により，機械翻訳が広く一般に使われるようになった．しかし，機械翻訳による翻訳品質の精度が向上したとはいえ，誤訳文を生成してしまう場合がある．このため，翻訳された文が適切な表現であるかどうかを検証することが重要となるが，全ての訳文の検証を人手で行うことは，機械翻訳による効率化の効果を損なうこととなる．

本研究では，翻訳文の品質を評価する手法として，Web 上に存在する文書群を利用した方法を検討する．Web 上のテキスト情報を利用して英文の品質を識別する手法は，隅田らにより，英語の多肢選択課題の自動作成手法について，その妥当性を検証するための手段として提案されている[1]．隅田らの手法では，英文の多肢選択課題で棄却候補を自動作成する際に，Web 上に存在するテキストは適切な英文であるとみなし，Web 上に存在するテキストと同一の表現の課題は多肢選択課題の棄却課題としては適切でないとして排除する．

同様に，機械翻訳システムからの翻訳文を Web 上で検索することができれば，その翻訳文は正しい表現と見なせる確率が高いと推測できる．正しい翻訳と見なせる文は適切な表現であるか否かの検証の対象外とすることにより，訳文の品質の検証過程を効率化できる．

筆者らは，このような考え方に基づき，Web 上で翻訳文がヒットするか否かにより翻訳文を正解訳と誤訳に分類する手法を提案した[2]が，この手法は，文の区切りが特定されないため，誤訳が正解訳に分類されるなどの問題があった．

本稿では，先に提案した手法の問題点の対処法として，単語クラスの導入，訳文が Web 上で部分文として一致しているかの検証，更に翻訳モデルを用いた翻訳品質評価を組み合わせる手法について検討した結果を報告する．

第 2 章では，先に提案した手法について簡単に説明し，実験結果を示すと共に，実験により示された提案手法の課題について説明する．第 3 章では，これらの課題に関して，Web 上の英文テキストとマッチングを行う手法の範囲内で可能な改善手法について検討する．第 4 章では，翻訳モデルを用いた翻訳品質の評価手法について検討する．第 5 章では，提案手法の有効性の評価を行う．第 6 章では，結論と今後の検討課題について述べる．

2 提案手法と検索ヒット率

既に述べたように，Web 上には様々な英文テキストを見つけることができ，翻訳文と一致する文を Web 上で検索することができれば，その意味内容が翻訳対象文として適合するかどうかは度外視したとしても，その翻訳文は正しい確率が高く，この文を正解訳と誤訳の評価対象から除くことにより効率化が可能と期待できる．

2.1 適切な英文の検索ヒット率の検証

まず本章では，本推測の妥当性の検証を行うために，ATR で収集された，旅行会話に関する大規模な日英対訳コーパスである BTEC（Basic Travel Expression Corpus）[3]の中からランダムに選択された約 2,300 文の英文を対象として，各種の検証結果を示す．

検索エンジン Google[4]により，評価対象文の英文と一致する文を Web から検索した際の検索ヒット率を図 1 に示す．図 1 の横軸は英文に含まれる単語数を，縦軸は単語数毎の検索ヒット率を示す．なお，図中の文数は，評価に用いた約 2,300 文の単語数の分布を示している．検索ヒット率とは，全ての翻訳文中の 1 件でも一致する文が存在する翻訳文の割合を示している．

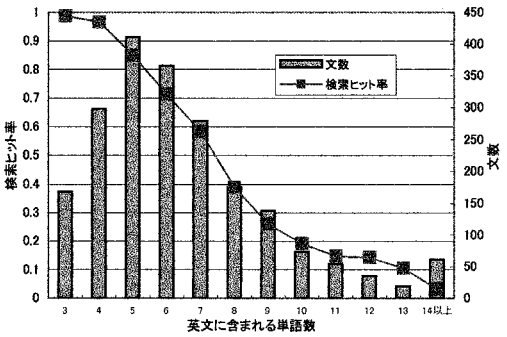


図 1：英文に含まれる単語数毎の検索ヒット率

評価対象文全体としての検索ヒット率は 68%であった．図 1 では省略しているが，1 文に含まれる単語数が 1 または 2 である文は，全体の 6.5%あり，検索ヒット率は 100%であった．図 1 から，英文に含まれる単語数が多くなるに伴い，検索ヒット率が低下していることがわかる．また，BTEC の英文の平均単語数は 6.2 であり，単語数が 6 のとき，検索ヒット率は 71%，単語数が 7 のとき，検索ヒット率は 58%と，平均単語数周辺で検索ヒット率が 50%を超えた．一方で，単語数が 10 以上となると，検索ヒット率が 20%以下にまで低下した．

以上から，英語として適切な表現と考えられる BTEC の英文に関しては，1 文に含まれる単語数が極端に多くない場合は，かなりの割合で検索ヒットし，Web 上に一致する文が存在し，英文の質を検証するためのリソースとして，Web 上の英文テキストを検索する手法は有効に機能すると考えられる．

一方，適切な英語表現の文では，Web 上に一致する文が

かなりの割合で存在するが，機械翻訳による誤訳も検索ヒットすることが予測される．この検証結果を次節で述べる．

2.2 機械翻訳による訳文の検索ヒットの検証

機械翻訳による翻訳文の検索ヒットの検証を行うに際して，BTECが対象とする旅行会話に関する文に適合していると考えられる4種類の異なる機械翻訳方式による翻訳文を用いた．具体的な翻訳文は，BTECを対象とした機械翻訳に関するワークショップであるIWSLT2004評価キャンペーン[6]における言語資源と参加した翻訳システムの翻訳結果を利用した．このキャンペーンは，旅行会話に関するBTECを用い，参加システムの翻訳結果を評価するものである．翻訳文の総数は，評価キャンペーンに使用されたBTECのテスト文500文に対する翻訳結果，計2,000文である．なお，図1に示した結果を得るために使用した約2,300文の英文は，その一部として本ワークショップで使用された500文の対訳を含んでいる．

翻訳文のWeb上での検索ヒットの検証手順は以下の通りである．

1. 日英バイリンガル3名により，翻訳文ごとの翻訳品質をFluency（流暢さ）とAdequacy（適切さ）の観点からの主観評価を，表1の評価基準に基づき行う．3人の評価値のメディアン値を評価値として採用する．翻訳品質の主観評価を用いて，翻訳文を正解訳（Adequacy，Fluency共に4以上）と3つの誤訳クラスに分類．誤訳クラスとしては，誤訳クラス$A^{-}F^{-}$（Adequacy，Fluency共に3以下），誤訳クラス$A^{+}F^{-}$（Adequacyは4以上，Fluencyは3以下），誤訳クラス$A^{-}F^{+}$（Adequacyは3以下，Fluencyは4以上）に分類．
2. 翻訳文を検索フレーズとして，翻訳文と一致する文の検索ヒット文数（Web上に一致する文が検索される場合の入力された翻訳文の数）を求める．

表1：翻訳品質の主観評価

Fluency *1		Adequacy *2	
5	Flawless English	5	All Information
4	Good English	4	Most Information
3	Non-native English	3	Much Information
2	Disfluent English	2	Little Information
1	Incomprehensible	1	None

*1 Fluency(流暢さ)：英語としてどのくらい滑らかか
*2 Adequacy(適切さ)：元の文の内容がどの程度正確に反映されているか

図2に翻訳文に含まれる単語数毎の正解訳と誤訳クラス$A^{-}F^{-}$，$A^{+}F^{-}$，$A^{-}F^{+}$に分類される誤訳の文数を示す．なお，以下の説明の簡単化のために，各翻訳文に含まられる単語数に応じて，表2に示すように翻訳文を，単語数が5未満(少単語数クラス)，単語数が5〜9(中単語数クラス)，単語数が10以上（多単語数クラス）に分類する．

表2：翻訳文の単語数に応じたクラス分類

クラス	単語数	占める割合
少単語数クラス	5未満	27%
中単語数クラス	5〜9	59%
多単語数クラス	10以上	14%

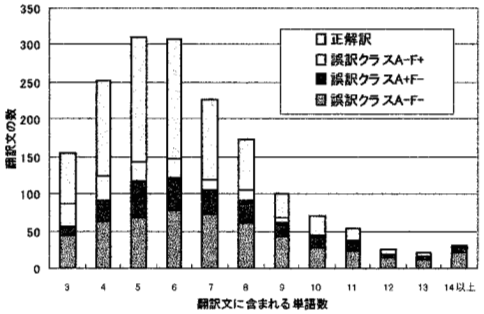


図2：単語数毎の正解訳と誤訳の文数

翻訳文と一致する文を，Googleを用いて検索した場合の，検索ヒットした翻訳文の文数を図3に示す．図3の横軸は翻訳文に含まれる単語数を，縦軸は検索ヒットした翻訳文の文数を表している．

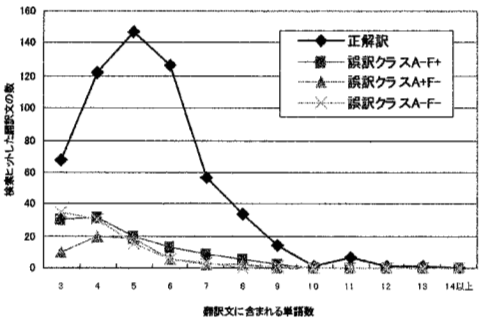


図3：単語数と検索ヒットした翻訳文の数の関連

図3から，翻訳文に含まれる単語数が中単語数クラスの場合は，正解訳と各誤訳クラスでは，検索ヒットした文数に顕著な差を確認できた．つまり，検索ヒットするか否かによる正解訳と誤訳の分類は，対象とする分野の会話文に適合した機械翻訳システムを用いた場合，かなりの割合で可能であると考えられる．

3 提案手法の課題と改善

先に提案したWeb上で検索ヒットするか否かで，正解訳と誤訳を分類する手法を2章で説明し，単語数がある程度以上の文の場合，かなりの割合で分類可能であることを示した．しかし，正解訳と誤訳の分類を可能とするために

は，まだ幾つかの課題が残されていた．以下，課題とその対処法について述べる．

図4に単語数と検索ヒット率の関係を示す．図4から明らかなように，比較的出現する文数の多い単語数が7～10の中程度の長さの文の場合，正解訳の検索ヒット率が低下する．更に，単語数が5未満の少単語数クラスにおいて，Fluencyが低い誤訳である誤訳クラス $A^{-}F^{-}$ 及び $A^{+}F^{-}$ の検索ヒット率が高い．本章では，これらの課題の対処法について述べる．

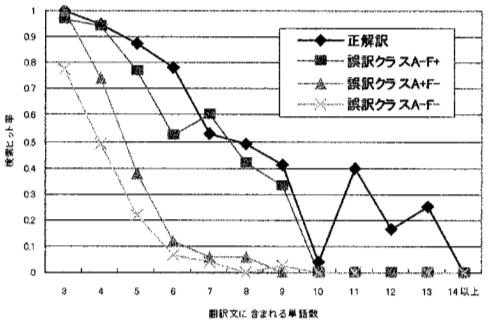


図4：単語数に対する検索ヒット率

3.1 単語クラスの導入などによるヒット率の向上

正解訳がWeb上で検索ヒットしない要因として，以下のようなことが考えられる．

1. 訳文を構成する単語に，数詞や固有名詞，頻度を示す副詞などのクラス内で置換可能な語が含まれている．
2. 訳文に，"please"などの会話特有な挿入表現が含まれる．
3. 訳文に，"I'd"などの短縮形が含まれている．

表3に，Web上で正解訳が検索ヒットしないと考えられる要因別の割合を示す．

表3：ヒットしない正解訳(201文)の要因別の割合

短縮形を含む文	45文/201文(22%)
会話特有な挿入表現を含む文	42文/201文(21%)
数詞を含む文	39文/201文(19%)
固有名詞を含む文	19文/201文(9%)

以下，これらの要因に対する検索ヒット率の改善手法について述べる．

① 要因1に対しては，汎用性の高い品詞に対して，同じ品詞内でクラスを作り，そのクラス内の単語が訳文に含まれた場合，その単語を，属するクラスに置き換えて検索する．

② 要因2に対しては，全体としての意味に影響しないと考えられる"please"，"hi"，"hello"などの会話特有な挿入表現を削除する．

③ 要因3に対しては，短縮形を，短縮形と短縮しない形の双方を用いて検索する．

上記の修正を行うことによる検索ヒット数の向上の結果を表4に示す．本検証での数詞をクラスに置換する方法は以下の例のような簡便な手法を用いた．

(例) *I'd like two brandies.* ⇒ *I'd like (one OR two OR ･･･ ten) brandies.*

上記の方法で検索を行うと，一文に数詞が2つ以上含まれた場合，Googleの検索語数の上限を超えてしまう．本検証では，このような場合は置換を行っていない．

表4：訳文修正で検索ヒット可能になった文の割合

短縮形の修正	18文/45文(40%)
会話特有な挿入表現除去	21文/42文(50%)
数詞をクラスに置換	18文/39文(46%)

翻訳文に対して，上記の改善手法①～③を全て実施した結果を図5に示す．図5には，正解訳の訳文修正前のヒット率と修正後のヒット率を示す．併せて，誤訳クラス $A^{-}F^{-}$ 及び $A^{+}F^{-}$ の修正前と修正後のヒット率を示す．

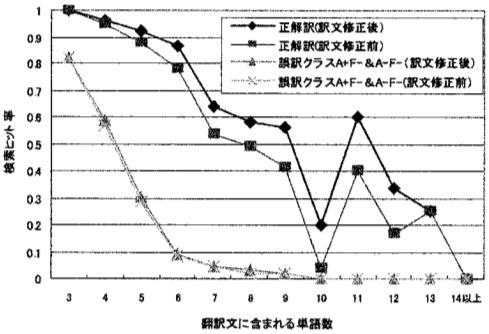


図5：検索ヒット率の改善手法によるヒット率の改善

中単語数クラス（単語数5～9）に関して，正解訳は訳文修正後にヒット率が約10%程度向上していることがわかる．一方で，誤訳クラス $A^{-}F^{-}$ 及び $A^{+}F^{-}$ は，ほぼヒット率は変化していない．以上から，本手法は，中単語数クラスで，Fluencyの低い誤訳のヒット率を変えずに，正解訳の検索ヒット率の向上を可能にしている．

3.2 部分文マッチの不許可によるヒット率の低減

Googleを使用する現在の検索手法では，検索する際に，文に含まれるピリオドや？マークは無視される．その結果，翻訳結果が文としてではなく，部分文として検索される場合がある．例えば，“このカメラはどうですか”という日本語文に対する翻訳結果として，“Is this camera how?”が得られたとする．このときWeb上では，“How old is this camera? How many megapixcels?”という様に，2文に渡った部分文として検索されてしまうケースがある．これによりFluencyの低い訳文も誤ってヒットする．この改善手法として，検索時にテキスト文そのものを獲得し，テキスト文の中から訳文を探し，訳文が部分文ではなく一文として一致しているかを検証する（以下，この検証を部分文マッチ許可/不許可と呼ぶ．）．図6に，正解訳に対し部分文マッチ許可/不許可を行ったときの検索ヒットした翻訳文の

数を示す．併せて，誤訳クラス A^-F^-及び A^+F^-に対し部分マッチ許可/不許可を行ったときの検索ヒットした翻訳文の数を示す．

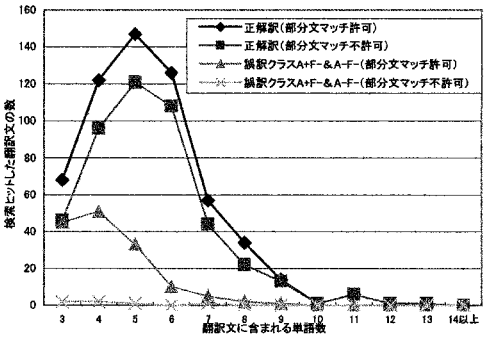


図6：部分文マッチの許可/不許可による検索ヒットした訳文の数の変化

誤訳クラス A^-F^-及び A^+F^-では，部分文マッチを不許可にした場合，検索した訳文の数がほぼ0になっている．一方，正解訳では，部分文マッチを不許可にした場合，検索ヒットした訳文の数は20%程度の低減で済んでいる．以上から，本手法は，正解訳の検索ヒットした訳文の数を多少下げるものの，Fluencyの低い誤訳の検索ヒットした訳文の数をほぼ0に低減することを可能にしている．

4 翻訳モデルの導入

翻訳文に含まれる単語数と翻訳文の検索ヒット率の関係を示した図4から明らかなように，Adequacyが低くFluencyが高い誤訳クラス A^-F^+の検索ヒット率が高い問題が存在する．このような誤訳は，原文である日本語文の意味を適切に伝えていないが，英文そのものは適切な表現であるので，Web上で検索ヒットしてしまう．つまり，誤訳クラス A^-F^+の翻訳文の識別は，Web上のテキストとマッチングを行う手法のみでは原理的に解決が困難である．

そこで，Web上のテキストとマッチングを行うことにより翻訳文の品質を評価した後に，提案する翻訳モデルを用いた品質評価を行うことによる，正解訳と誤訳の判別性能を向上させる手法について述べる．

4.1 翻訳モデルと単語翻訳確率

翻訳モデルは原言語単語列が生起する条件の下での，目的言語単語列の生成確率を示し，訳語の選択と単語の対応付けをモデル化する．翻訳モデルとしてIBM model1-5等があり[6]，IBM model4では，モデルパラメータとして単語翻訳確率，繁殖数確率，歪み確率，NULL生成確率の4つの確率を扱うモデルを提案している．

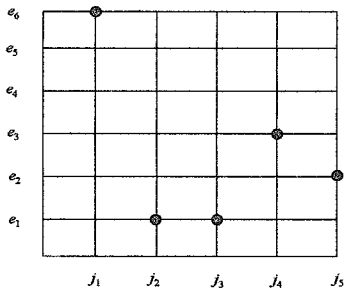
本稿では，翻訳文の品質評価の精度を向上させる手法として単語翻訳確率に着目する．単語翻訳確率とは，日本語のある単語 j が英語のある単語 e に翻訳される確率 $P(e|j)$ である．この単語翻訳確率を用いて，原文 $J(j_1|j_1j_2\cdots j_m)$に対する翻訳文 $E(e_k \mid e_1,e_2\cdots e_n)$の $P(E|J)$を求める．$P(E|J)$は，以下の手順に従って求める．なお，単語翻訳確率 $P(e|j)$はBTECの中の16万文を用いて学習した単語翻訳確率を使用している[7]．

1. 原文中の各日本語単語 j_i (i=1…n)について，翻訳文中の各英単語 e_k(k=1…m)の中から，$P(e_k|j_i)$の最大確率値を持つ英単語 e_k 求める．但し，$P(e_k|j_i)$が値を持つ単語 e_kが存在しない場合は0とする．
2. 各日本語単語 j_i (i=1…n)について求めた $P(e_k|j_i)$ の最大確率値を i=1～n まで乗算し，その結果を $P(E|J)$とする．

すなわち，$P(E|J)$は式(1)のように表される．

$$P(E \mid J) = \prod_{i=1}^{n} \max_{k} P(e_k \mid j_i) \quad (k = 1,2 \ldots m) \qquad (1)$$

図7に，$P(e_k|j_i)$が最大確率値となる日本語単語と英単語の対応関係の例を示す．



J={ パスポート(j_1) を(j_2) お(j_3) 見せ(j_4) 下さい(j_5)}

E={NULL(e_1) please(e_2) show(e_3) me(e_4) your(e_5) passport (e_6)}

図7：$P(e_k|j_i)$が最大確率値となる日本語単語と英単語の対応関係

図7に示す例では，日本語文の格助詞「を」や接頭辞「お」については，対応する英語の単語はなく，仮想的な単語NULLに対応している．また，単語NULL以外でも，複数の日本語単語が単一の英単語に対応することも可能である．

図8に，正解訳と誤訳クラス A^-F^+における，$P(E|J)$の分布を示す．縦軸は，正解訳と誤訳クラス A^-F^+の総数中の正解訳の割合を示し，横軸は式(2)に示すように日本語単語数で正規化された $P(E|J)$を示している．

$$\text{正規化}\ P(E \mid J) = \frac{\log P(E \mid J)}{n} \qquad (2)$$

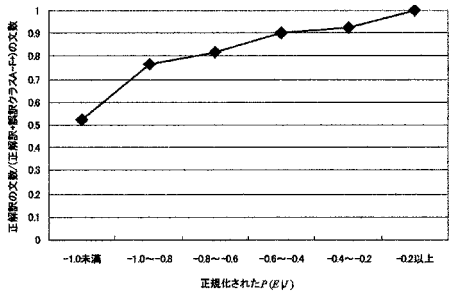


図8：正規化された$P(E|J)$の分布

図8から，正規化された$P(E|J)$が-1.0未満のとき，正解訳の占める割合は約50%である．また，正規化された$P(E|J)$の値が大きくなるに連れて，正解訳の占める割合も高くなっている．正規化された$P(E|J)$が-0.2以上の場合は，正解訳の占める割合は100%である．以上より，正規化された$P(E|J)$を用いることで，正解訳と誤訳クラスA^-F^+の識別が可能であると期待できる．

4.2 翻訳モデルを用いた誤訳の識別

4.1節で求めた単語翻訳確率$P(e_k|j_i)$を用いて，まず翻訳文を次のように分類する手法について検討を行う．

$P(E|J)>0$のときは，原文の日本語単語すべてが，対応する英単語を翻訳文中に持つので，このときの翻訳文は正解訳に分類する．逆に$P(E|J)=0$のときは，原文の日本語単語の中に，対応する英単語を翻訳文中に持たない単語が存在するので，このときの翻訳文は誤訳に分類する．

正解訳と誤訳クラスA^-F^+の翻訳文に対して，上記の分類手法を行った結果を図9に示す．図9には，正解訳、誤訳クラスA^-F^+それぞれについて，$P(E|J)>0$となる翻訳文の割合を示す．

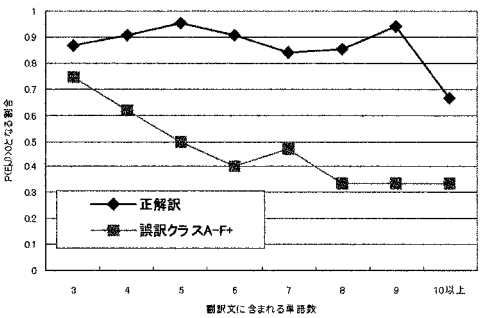


図9：$P(E|J)>0$となる翻訳文の割合

図9から，翻訳文に含まれる単語数に拘らず，$P(E|J)>0$となる割合は，正解訳の方が誤訳クラスA^-F^+より高い．また，出現する文数が比較的多い中単語クラスでは，その差は顕著である．以上から，上記の分類手法は，正解訳と誤訳クラスA^-F^+の識別に有効であると考えられる．

5 提案手法の評価

本章では提案手法の有効性を検証する．検証は，翻訳文毎に，表1に基づく翻訳品質の主観評価（正解訳，誤訳クラスA^-F^-，A^+F^-，A^-F^+）と提案手法による品質評価の一致度に基づく．

5.1 Web上の英文テキストとマッチングを行う手法の評価

Web上の英文テキストとマッチングを行う手法を3.1～3.2節に基づいて改良する前と後における，正解訳の検出割合と誤訳の正解訳への誤検出割合を表5に示す．

表5：正解訳の検出割合と誤訳の正解訳への誤検出割合1

		提案手法による評価（改良前）	提案手法による評価（改良後）
主観評価	正解訳	73% (577文/787文)	66% (523/787文)
	誤訳クラスのいづれか	28% (258文/936文)	9% (81文/936文)

提案手法を改良することにより，主観評価で正解訳と評価された訳文について，提案手法が正解訳に分類する割合は，約9/10に少し低下している．一方，提案手法が誤訳を正解訳と誤る割合は約1/3に低下している．また，誤訳を正解訳と誤る場合の85%（69文/81文）は誤訳クラスA^-F^+の訳文であった．

5.2 翻訳モデルを組み合わせる手法の評価

Web上の英文テキストとマッチングを行う手法に翻訳モデルを用いる手法を組み合わせる前と後における，正解訳の検出割合と誤訳の正解訳への誤検出割合を表6に示す．

表6：正解訳の検出割合と誤訳の正解訳への誤検出割合2

		提案手法による評価（組み合わせ前）	提案手法による評価（組み合わせ後）
主観評価	正解訳	66% (523/787文)	60% (474文/787文)
	誤訳クラスのいづれか	9% (81文/936文)	5% (44文/936文)

Web上の英文テキストとマッチングを行う手法に翻訳モデルを組み合わせることにより，主観評価で正解訳と評価された訳文について，提案手法が正解訳に分類する割合は約9/10に少し低下している．一方，提案手法が誤訳を正解訳と誤る割合は約1/2に低下している．また，翻訳モデルを組み合わせる手法を用いても，誤訳を正解訳に誤って分類した翻訳文は44文存在した．このような文は，「すいません，わかりません．」に対する「Excuse me, I don't know」や「私は会社員です．」に対する「I'm an office worker.」のように，日本人では正しい訳か間違っている訳かの判断が難しいが，ネイティブは使わない表現であるといった場合が多く見られた．

6 結論と今後の課題

本稿では，「英語対話システム」の開発の効率化のために，機械翻訳を用いることを想定し，機械翻訳の正解訳と誤訳を分類する手法として，先に提案した Web 上で翻訳文が検索ヒットするか否かによって正解訳と誤訳を分類する手法における課題と対策ついて述べた

改善手法として，単語クラスの導入などにより，単語数が 5〜9 の翻訳文で，Fluency の低い誤訳の検索ヒット率はそのままに，正解訳の検索ヒット率を 10%程度向上させることを示した．また，Fluency が低い文でも一定の割合でヒットする問題の対策として，翻訳文の部分文マッチを不許可にすることで，Fluency の低い誤訳の検索ヒット率をほぼ 0 にすることを示した．最後に，Web 上の英文テキストとマッチングを行う手法のみでは解決が困難である Fluency の高い誤訳の識別に関しては，翻訳モデルを用いて品質を評価する手法を検討した．本手法を用いることで，Fluency の高い誤訳の 50%程度を誤訳として分類できることを示した．

この結果，誤訳の 5%程度を正解訳として誤って判断するが，正解訳の約 60%は正しく判断できる段階までに到達した．

今後は以下のような課題に関して検討を行う．

① Web 上の英文テキストとマッチングを行う現在の手法では，翻訳文に含まれる単語数が極端に多い場合，正解訳の検索ヒット率が大幅に低下する．この課題に対しては，英語構文解析を用いて，翻訳文を句単位や節単位で分割し，それぞれを Web 上の英文テキストとマッチングをして評価する対応を検討する．

② Adequacy が低く Fluency が高い誤訳である誤訳クラス $A^{-}F^{+}$の翻訳文の識別には，翻訳モデルを用いて品質を評価する手法を提案した．しかし本手法では，誤訳クラス $A^{-}F^{+}$の翻訳文すべてを誤訳として分類するまでには至っていない．今後は，翻訳モデルのパラメータ学習の際の対訳データを増やすと共に，フロアリング等の手法と正規化された $P(E|J)$の利用等を検討する．

③ 英語対話システムの対象分野毎の対話シナリオの開発の際に，市販されている機械翻訳を使用することも考えられる．市販ソフトの機械翻訳により会話文などの口語表現を翻訳した場合，誤訳率が高くなることが想定されるため，誤訳の割合が高いと考えられる場合の対応を検討する．

謝辞

本研究を進めるにあたり，有意義なコメントを頂いた ATR 音声言語コミュニケーション研究所，隅田英一郎室長，ルパージュ・イブ主任研究員に感謝致します．

本研究は，科学研究費補助金（基盤研究B）（課題番号 16300048）による助成研究の一部である．

参考文献

(1) E. Sumita, F. Sugaya, S. Yamamoto：“Measuring non-native speaker's profiency of Englsih by using a test with automatically-generated fill-in-the-blank questions”, Proc. 2nd Workshop on Building Applications using NLP, pp. 61-68 (2005).

(2) 宮下 広平, 安田 圭志, 山本 誠一, 柳田 益造: “WWW 上のテキスト情報を利用した翻訳品質評価法の検討”, 情報処理学会 第 171 回自然言語処理研究会, Vol.2006-NL-171, pp89-94

(3) T. Takezawa, E. Sumita, F. Sugaya, H. Yamamoto, S. Yamamoto：“Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the rea world”, Proc. LREC-02, pp.147-152 (2202).

(4) “http://www.google.com/apis/index.html”

(5) Y. Akiba, M. Federico, N. Kando, H. Nakaiwa, M. Paul, J. Tsujii.：“2004. Overview of the IWSLT04 evaluation campaign”, Proc. IWSLT2004, pp.1-12 (21004).

(6) Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L.Mercer. “The mathematics of statistical machine translation: parameter estimation.” , *Computational Linguistics*, Vol. 19, No. 2, pp. 263–311, 1993.

(7) Eiichiro Sumita, Yasuhiro Akiba, Takao Doi, Andrew Finch, Kenji Imamura, Hideo Okuma, Michael Paul, and Mitsuo Shimohata. “EBMT, SMT, Hybrid and More: ATR Spoken Language Translation System.” IWSLT 2004 (International Workshop on Speech Language Translation), Proc. of IWSLT2004, pp.13-20