TR-S-0023

対訳コーパスを用いた表層的類似度に基づく翻訳能力自動評価法

An Automatic Evaluation Method of Translation Quality Using Translation Answer Candidates Queried from a Parallel Corpus

安田 圭志 Keiji Yasuda
菅谷 史昭 Fumiaki Sugaya
竹澤 寿幸 Toshiyuki Takezawa

2001.3.30

翻訳品質の自動評価法について提案し,本手法を日英ATR-MATRIXの評価に適用した結果について報告する．本手法は,対訳コーパスから翻訳正解候補を追加し,システムの翻訳文と翻訳正解候補とで,DPマッチングにより表層的類似度に基づく評価を行うものである．本手法での評価結果と正解候補を追加しない場合の自動評価結果を,主観評価の結果を用いた判別分析により比較した．この結果,提案手法を用いた場合が,正解候補を追加しない場合より,有効な判別が行え,本手法の有効性が確認された．また,正解文追加類似度計算法を用いた翻訳システム能力のTOEICスコア換算の自動化法について説明し,ATR-MATRIXの評価に適用した結果について示す.

目 次

1 はじめに

これまでにも，ATR 音声言語通信研究所では，ATR-MATRIX [1] の評価を通じて翻訳システムの評価方法について研究を行っており，評価者が主観で翻訳の質を 4 ランクに割り当て，翻訳の質のランクを決定する「翻訳ランク評価法」 [2] や，システムと人間能力との比較を通じてシステムの翻訳能力を評価し TOEIC スコアに換算する「翻訳一対比較法」 [3] による評価を行ってきた．

しかしながら，何れの方法も，主観による判定が必要であり，それに要するコストは少なくない．システムの性能改善の効率化には客観的な評価が必要であることが一般に認識されるようになった．

本稿では，DP マッチングをベースとしながら，未登録の正解を対訳コーパスで補う方法を提案する．更に，提案手法を日英 ATR-MATRIX の評価に適用した結果について示し，提案手法の評価能力と従来の DP マッチングによる評価手法の評価能力の比較を行っている．また，翻訳システムの持つ翻訳能力の TOEIC スコアへの換算を自動化する方法を提案し，この方法を使った換算結果と，翻訳一対比較法での換算結果との比較を行っている．

2　従来の翻訳評価法と提案手法

本節では，人手による主観評価である翻訳ランク評価法と，従来の自動評価法であるDPマッチングによる評価法について簡単に説明し，次に，提案手法である正解文追加類似度計算法についての説明を行う．

2.1　翻訳ランク評価法

従来の人手による評価手法である翻訳ランク評価法では，評価者は主観により，次の基準で翻訳の質のランク付けを行っている．

Aランク（完全訳）：訳文だけで全く問題なし．
Bランク（部分訳）：訳文は少し情報が欠けている．
Cランク（可能訳）：訳文はかなり情報が欠けている．
Dランク（不可訳）：訳文からは，情報が想像もできない．

2.2　従来のDPマッチングによる自動評価法

従来のDPマッチングによる自動評価法では，1つの原言語テスト文に対し，翻訳の正解として，1つの正解目的言語文を用意しておき，DPマッチングを用いて，以下に定義する類似度 (Similarity) を計算し，この値を評価結果としている．

$$\sigma = \frac{T - S - I - D}{T} \tag{1}$$

式 (1) において，σ は類似度を表している．T は正解目的言語文の総語数，S は正解目的言語文とシステムによる翻訳結果をDPマッチングにより比較した時の置換語数，I は同様に比較した時の挿入語数，D は同様に比較した場合の脱落語数である．

2.3　正解文追加類似度計算法

従来のDPマッチングによる類似度の問題点は，ある1つの原言語テスト文に対して，正解目的言語文を1つだけしか用意していない点であった．正解文追加類似度計算法では，従来のDPマッチングによる類似度と同様に，システムによる翻訳結果と正解目的言語文との表層的な単語一致度に注目して評価を行うが，従来のDPマッチングによる類似度の問題点を解決するため，1つの原言語テスト文に対して，複数の正解目的言語文を対訳コーパスを用いて収集している．

正解文追加類似度計算法の処理の流れを図1に示す．図中の原言語コーパス（Source language corpus）と目的言語コーパス（Target language corpus），及び，原言語テスト文（Source language test sentence）と目的言語テスト文（Target language test sentence）は対訳関係になっている．図1では，まず，DPマッチングにより，原言語側コーパスの中から，原言語側テスト文の類似文を検索する．類似度がある一定の閾値以上となるものを類似文とする．以降，ここで得られた類似文を類似原言語文，閾値を類似文検索閾値と呼ぶ．類似文検索の際の類似度は，式 (1) で定義した類似度において，正解翻訳文を原言語テスト文

に，システムによる翻訳結果を原言語コーパス内の文に置き換えて計算した類似度である．次に，ここまでに得られた類似原言語文の目的言語側と，テスト文の目的言語側をあわせて正解群とする．最後に，システムによる翻訳結果と正解群の中の各文とで，DP マッチングにより翻訳結果のスコアリングを行い，類似度を求める．この結果として，正解群に含まれる文の数だけ類似度が求まるが，その最大類似度を正解群類似度 (answer set similarity) とし，これを翻訳文の評価尺度とする．

3 正解文追加類似度計算法の ATR-MATRIX の評価への適用

本節では，正解文追加類似度計算法を，日英 TDMT の評価に適用した結果について述べる．次に，これらの評価結果を用いた判別分析について述べる．最後に，音声認識サブシステムと言語翻訳サブシステムを統合した音声翻訳システム ATR-MATRIX の評価に適用した結果について示し，音声翻訳システムの評価に適用する場合の問題点について述べる．

TDMT 単体の評価では，テストセットを TDMT へテキスト入力してえられた翻訳結果の評価を行う．音声翻訳システムの評価では，テストセットを SPREC に音声入力し，認識結果を TDMT で翻訳した結果の評価を行う．音声認識正解率は，88.5%である．

正解文追加類似度計算において，用いた対訳コーパスは，ATR で構築された 618 会話（16110 文）からなるバイリンガル旅行対話データベース [6] [7] であり，テストセットはこの内の 23 会話（330 文）である．この 23 会話は，音声認識部，言語翻訳部に対してオープンである．また，正解文追加類似度計算における原言語側（日本語）の類似文検索では，形態素単位での DP マッチングを行い，目的言語側（英語）でのスコアリングでは単語単位での DP マッチングを行っている．

3.1 翻訳ランクと正解群類似度の関係

本小節では，正解文追加類似度計算法による TDMT の評価結果について述べる．図 2 は，類似文検索閾値と類似文検索により得られる目的言語類似文数の平均の関係を示している．図中の○は，重複を許した場合ののべ文数，●は，重複を許さない場合の異なり文数を表している．

図 2 より類似文検索閾値の減少にともない，追加される正解文の数が増加していることがわかる．表 1 は類似文検索閾値が 0.6 の場合に，類似文検索により追加される目的言語類似文の一例である．表中の"/"は，日本語における形態素境界を表している．異なる言い回しであるが，同義の目的言語文が得られていることがわかる．

図 3 に，翻訳ランクと正解群類似度の関係を示す．類似文検索閾値は 0.6 としている．図 3 の横軸は翻訳ランクを表しており，縦軸は正解群類似度を表している．○は各テスト文を表しており，●は各翻訳ランク内での正解群類似度の平均を表している．

図 3 より，主観評価で高いランクである翻訳結果ほど，正解群類似度も大きくなる傾向があることがわかる．

3.2 判別分析によるランク決定

本小節では，正解文追加類度計算法の評価性能について調べるため，翻訳ランクと正解群類似度を用いた判別分析を行った結果について述べる．

3.2.1 判別方法

2 クラス分けの判別と，4 クラス分けの判別を行った．2 クラス分けの判別では， A ランクと B,C,D ランクの 2 クラス分け， A,B ランクと C,D ランクの 2 クラス分け，A,B,C ランクと D ランクの 2 クラス分けを行った．4 クラス分けについては，翻訳ランク評価法の各ランクを，そのままクラスとしている．

判別は，各クラスの正解群類似度の平均を求め，最近傍則に従って行う．
以下に判別率（Discriminant ratio）を定義する．

$$D = \frac{n_{correct}}{n_{total}} \tag{2}$$

式 (2) において，D は判別率，$n_{correct}$ は正しく判別された文の数，n_{total} は，文の総数である．

3.2.2 判別結果

図 4 に TDMT の評価結果について判別を行った結果を示す．図 4 では，類似文検索閾値毎の判別率と，従来の DP マッチングによる類似度を用いて判別を行った場合の判別率を示している．図 4 において，縦軸は判別率であり，横軸は類似文検索閾値または，従来の DP マッチングによる類似度を用いた方法を表している．また，図中の凡例の"/" は，クラス分けの境界を表している．例えば，A/B,C,D の記述では，A と B,C,D の 2 クラス分けを表している．

図 4 では，クラス分けの方法により判別率が最大となる類似文検索閾値は変化するが，全ての類似文検索閾値と全てのクラス分けにおいて，正解群類似度を用いた判別の方が，従来の DP マッチングによる類似度を用いた場合より，判別率が 10%程度改善されている．特に A と B,C,D の 2 クラス分けの判別では効果が大きく，判別率が 68.5%から 83.5%となり 15 ポイントの改善がみられている．

3.3 音声翻訳システムの評価への適用

本小節では，正解文追加類似度計算法を，TDMT の入力側に音声認識サブシステムを加えた音声翻訳システムの評価に適用した結果について示す．

図 5 に，翻訳ランクと正解群類似度の関係を示す．類似文検索閾値は 0.6 としている．図 5 は図 3 同様，横軸は翻訳ランクを表しており，縦軸は正解群類似度を表している．○は各テスト文を表しており，●は各ランク内での正解群類似度の平均を表している．

図 3 と図 5 での結果を比較すると，図 5 の評価結果では，D ランクと評価されているが，正解群類似度の値が大きいものがある．表 2 がその例である．このような結果となる原因は，ランク評価の評価単位と DP マッチングのマッチング単位のズレであると考えられる．本稿での目的言語側での DP マッチングのマッチング要素は，分かち書きされた単語である．そのため，数字の連鎖からなる電話番号や料金，固有名詞では，音声認識により誤りが生じても，翻訳の構造を損なうことなく言語翻訳がなされ，その正解群類似度も高い．しかしながら，翻訳ランク評価法では，翻訳結果の数字などの重要な情報が誤っていることから D ランクと判定されてしまう．この問題を解決するために，自動翻訳評価に適したマッチング単位を選択する必要があると考えられる．この点については，今後の検討課題とする．

4 正解文追加類似度計算法と翻訳ランク評価法との併用による評価コスト削減の検討

前節で述べたように，翻訳ランクのAランクとBCDランクの2クラス分けでは，正解群類似度を用いた判別分析により，83.5%とある程度高い判別率が得られた．これは，従来のDPマッチングによる類似度を用いて判別分析を行った場合より，判別率が15ポイント改善されているが，実用を考えた場合，更に高精度な評価能力を必要とする場合がある．

そこで，本節では，正解群類似度を用いた判別分析と，翻訳ランク評価法を併用することにより，誤りを小さくし，人手による評価コストを削減する方法について述べる．本節では，まず，正解文追加類似度計算法と翻訳ランク評価法との併用方法について説明し，削減される評価コストと誤りの定義を行う．最後にTDMTの評価に適用した場合の結果について述べる．

4.1 正解文追加類似度計算法と翻訳ランク評価法との併用方法

正解文追加類似度計算法と翻訳ランク評価法との併用方法は，Aランクの判別を自動で行い，その他のクラスと自動判別されたものについては，翻訳ランク評価法により評価を行う．以降，Aランクをクラス1，B,C,Dランクをクラス2と呼ぶ．クラス1とクラス2の判別においては，翻訳ランク評価法での結果がAである場合（$class\,1$）と，B,C,Dランクである場合（$class\,2$），それをクラス1であると自動判別する場合（$CLASS\,1$）と，クラス2であると自動判別する場合（$CLASS\,2$）の4つの組み合わせがあり，表3に示す4種類の確率が定義できる．

表3おいて，$P(CLASS\,1\,|\,class\,1)$はクラス1をクラス1として正しく受理する確率（クラス1受理率：Correct acceptance ratio），$P(CLASS\,1\,|\,class\,2)$はクラス2をクラス1として誤って受理する確率（クラス2誤り受理率：false acceptance ratio），$P(CLASS\,2\,|\,class\,1)$はクラス1をクラス2として棄却する確率（クラス1棄却率：False rejection ratio），$P(CLASS\,2\,|\,class\,2)$はクラス2をクラス2として棄却する確率である．

4.2 コストと誤りの定義

自動判定されるテスト文のテストセット全体に対する割合$P(CLASS\,1)$を，削減される評価コストとする．$P(CLASS\,1)$は次式で求めることが出来る．

$$P(CLASS\,1) = P(CLASS\,1\,|\,class\,1) \times P(class\,1) + P(CLASS\,1\,|\,class\,2) \times P(class\,2) \tag{3}$$

式(3)において，$P(class\,1)$は全テストセットの内クラス1に属する文の割合で，$P(class\,2)$は全テストセットの内クラス2に属する文の割合である．今回の評価対象であるTDMTでは，$P(class\,1)$が0.48，$P(class\,2)$が0.52である．

式(3)の右辺第一項の$P(CLASS\,1\,|\,class\,1) \times P(class\,1)$は，クラス1をクラス1として正しく受理する文数の，テストセット全体に対する割合である．また，右辺第2項の$P(CLASS\,1\,|\,class\,2) \times P(class\,2)$は，クラス2を誤ってクラス1として受理してしまう文数の，テストセット全体に対する割合である．

削減される評価コスト内の誤り率を，C とすると，C は次式で求めることができる．

$$C = P(CLASS\,1 \mid class\,2) \times P(class\,2) \div P(CLASS\,1) \tag{4}$$

4.3 TDMT への適用結果

図 6 は，判別する際にクラスの境界とする正解群類似度（判別閾値）と，クラス 2 誤り受理率及びクラス 1 棄却率の関係である．図 6 において横軸は，正解群類似度を表しており，縦軸は，クラス 2 誤り受理率，またはクラス 1 棄却率を表している．これらは，類似文検索閾値は 0.6 とした場合の結果である．図 7 は，従来の DP マッチングによる類似度を用いた場合について，図 6 と同様にプロットした図である．図 6 及び図 7 において，○はクラス 1 棄却率を表しており，●はクラス 2 誤り受理率を表している．

図 8 は，図 6 及び図 7 から求めた，クラス 1 受理率とクラス 2 誤り受理率との関係である．横軸はクラス 2 誤り受理率，縦軸はクラス 1 受理率である．図 8 の各点は，判別閾値を 0.1 から 1.0 まで，0.1 きざみで変化させた場合の結果である．○は正解群類似度を用いた場合の結果であり，●は従来の DP マッチングによる類似度を用いた場合の結果である．従来の DP マッチングによる類似度を用いた場合と比較し，正解群類似度を用いた場合の方が，全ての点で優れていることが分かる．

図 9 は，削減される評価コストである $P(CLASS\,1)$ と，削減される評価コスト内の誤り率 $P(CLASS\,1 \mid class\,2) \times P(class\,2) \div P(CLASS\,1)$ の関係を表している．縦軸が $P(CLASS\,1)$ で，横軸が $P(CLASS\,1 \mid class\,2) \times P(class\,2) \div P(CLASS\,1)$ である．図中の○は，正解群類似度を用いた場合の結果であり，●は，従来の DP マッチングによる類似度を用いた場合の結果である．

図 9 を，自動判別された内の誤り率である横軸を固定してみた場合，正解群類似度を用いた結果の方が，従来の DP マッチングによる類似度を用いた場合と比べ，縦軸の値が 0.1～0.2 程度大きくなっている．これは，正解群類似度を用いた場合，従来の DP マッチングによる類似度を用いた場合に比べて，全体に対する誤りを同じとしたまま，テストセット全体に対して 10%～20%多く評価コストを削減できることを表している．例えば，削減された評価コスト内の誤りを 5%許容したとすると，提案手法ではテストセット全体に対して約 30%，従来の DP マッチングでは，約 20%の評価コストの削減が可能となっている．

5 TOEICスコア評価の自動化

本節では，まず，人手により翻訳システム能力をTOEICスコアへの換算する方法である翻訳一対比較法 [3] について簡単に説明する．次に，2節で述べた正解文追加類似度計算法を用いることによって，翻訳システム能力のTOEICスコアへの換算を自動化する手法について説明し，ATR-MATRIXに適用した結果について述べる．以降，翻訳システム能力のTOEICスコアへの自動換算法を「TOEICスコア自動換算法」と呼び，翻訳一対比較法，及びTOEICスコア自動換算法によるTOEICスコア換算結果を「TOEIC換算点」と呼ぶ．

本節では，前節同様，TDMT単体の評価と，音声認識サブシステムを含めたATR-MATRIX全体の評価を行う．また，本節で用いた正解文追加類似度計算法での評価結果は，類似文検索閾値を0.6とした場合の結果である．

5.1 翻訳一対比較法

本小節では，従来の人手による翻訳一対比較法により，翻訳システムの能力をTOEICスコアに換算する方法について簡単に説明する．TOEICスコアの推定には，システムの評価に用いたテストセット23会話（330文）をTOEIC受験者により翻訳させた結果を用いる．TOEIC受験者のTOEICスコアは，300点から900点まで，100点台毎に5名で計30名である．TOEIC受験者は，895点のTOEIC受験者が2名である以外は，全て異なるスコアである．

評価者は，各テスト文に対するTOEIC受験者による翻訳結果と，翻訳システムから出力された翻訳結果とを一対比較により優劣を決定する．これにより，それぞれのTOEIC受験者のTOEICスコア毎に，TOEIC受験者優勢の文の数と，翻訳システム優勢の文の数が求まる．翻訳システム能力をTOEICスコアに換算するには，この結果を用いて回帰分析を行い，330文の内の半数である165文となり，翻訳システムと能力が均衡するTOEICスコアを求める．

5.2 TOEICスコア自動換算法

本小節では，TOEICスコア自動換算法について説明する．

翻訳一対比較法でTOEICスコアの推定に用いられた各TOEIC受験者の翻訳結果について，正解文追加類似度計算法により，各文の正解群類似度を求める．次に，TOEIC受験者ごとにテストセット330文についての平均正解群類似度を求める．この結果が図10である．図10の縦軸がTOEIC受験者のTOEICスコアで，横軸が平均正解群類似度である．図中の●は，TOEIC受験者を表しており，図中の直線は回帰直線である．TOEICスコア自動換算には，この回帰直線に，翻訳システムによるの翻訳結果330文の平均正解群類似度を代入し，翻訳システム能力に相当するTOEICスコアを求める．図10中の▲はTDMT単体の評価に正解文追加類似度計算法を適用して得られた平均正解群類似度であり，■はそのTOEIC換算点を表している．同様に△は，TDMTの入力側に音声認識サブシステムを加えたATR-MATRIXの評価での平均正解群類似度であり，□はそのTOEIC換算点である．

5.3 TOEIC スコアへの換算結果

文献 [3] によれば，翻訳一対比較法による TDMT 単体の TOEIC 換算点は 708 点であり，図 10 よりえられる TDMT 単体の TOEIC 換算点は 705 点である．両手法による TOEIC 換算点の差は，わずか 3 点であり，提案手法により翻訳一対比較法と近い結果が得られている．ATR-MATRIX においては，提案手法による TOEIC 換算点が 618 点，翻訳一対比較法による TOEIC 換算点が 548 点である．両手法による換算点の差は 70 点となり，音声認識誤りが含まれない場合よりも，両手法間の TOEIC 換算点の差が大きくなっている．

表 4 は，これらの結果をまとめたものである．

6 まとめ

翻訳システムの新たな評価手法として，正解文追加類似度計算法を提案し，ATR 音声翻訳通信研究所で開発された ATR-MATRIX の評価に適用した．

3 節では，正解文追加類似度計算法を ATR 音声翻訳通信研究所で開発された音声翻訳システム ATR-MATRIX の言語翻訳サブシステムである TDMT の評価と，TDMT の入力側に音声認識サブシステムを加えた，音声翻訳システムである ATR-MATRIX の評価に適用した．正解文追加類似度計算法を TDMT の評価に適用した結果を用いて判別分析を行ったところ，従来の DP マッチングによる評価結果を用いた場合と比較して，10%前後判別率の改善がみられた．特に，A ランクと B,C,D ランクの 2 クラス分けでは，判別率が 68.5%から 83.5%となり，15 ポイントの改善がみられた．しかし，図 3 からもわかるように，提案手法を用いても，翻訳ランク評価法では，A ランクと評価されているにもかかわらず，正解群類似度の値が小さいままのものがあり，この事が誤りの原因となっていると考えられる．

音声翻訳結果で，翻訳システムへの入力側に音声認識誤りが含まれる場合は，D ランクであるにもかかわらず，正解群類似度が大きい場合があることが分かった．原因は，3.3 節で述べたように，ランク評価の評価単位と DP マッチングのマッチング単位のズレであると考えられる．この問題を解決するために，自動翻訳評価に適したマッチング単位を選択する必要があると考えられる．この点については，今後の検討課題とする．

4 節では，正解文追加類似度計算法と，翻訳ランク評価法を併用することにより，誤りを小さくしたまま，人手による評価コストを削減する方法について述べた．これを TDMT の評価に適用した結果，削減された評価コスト内の誤りを 5%許容したとすると，提案手法ではテストセット全体に対して約 30%の評価コストの削減が可能であることが示された．また，正解文追加類似度計算法による正解群類似度の代わりに，従来の DP マッチングによる類似度を用いた場合は，正解文追加類似度計算法による正解群類似度を用いた場合より，削減される評価コストが，テストセット全体に対して 10%～20%少なくなることが確認された．

5 節では，正解群類似度を用いて，翻訳システム能力の TOEIC スコアへの自動換算法を提案し，TDMT と ATR-MTARIX の評価に適用した．この結果，TDMT の評価では，翻訳一対比較法とほぼ同様の結果が得らた．音声翻訳サブシステムを含めた ATR-MATRIX の評価では，提案手法による TOEIC 換算点が，翻訳一対比較法による TOEIC 換算点より 70 点高くなるという結果が得られた．これは，3 節で述べたように，翻訳システムの入力側に音声認識誤りが含まれる場合は，翻訳品質が低いにもかかわらず，正解群類似度が高くなることが起因しているためと考えられる．

謝辞

本研究の機会を与えられ，常に暖かいご指導，御鞭撻を頂いた，ATR 音声言語通信研究所　山本誠一社長，匂坂芳典第二研究室室長, 同志社大学工学部　柳田益造教授に深く感謝いたします．

本研究を進めるにあたり，実験等の支援して頂いた林輝昭氏, 津山佳子氏，吉岡由紀氏に心より感謝致します．また，本研究を進めるにあたり，本研究の内容について熱心に討論に応じていただいた，ATR 音声言語通信研究所の諸氏に厚くお礼申し上げます．

参考文献

[1] T. Takezawa, T. Morimoto, Y. Sagisaka, N. Campbell, H. Iida, F. Sugaya, A. Yokoo, S. Yamamoto, "A Japanese-to-English speech translation system: ATR-MATRIX", Proc. ICSLP, pp.2779-2782, 1998.

[2] E. Sumita, S. Yamada, K. Yamamoto, M. Paul, H. Kashioka, K. Ishilawa, S. Shirai, "Solutions to Problems Inherent in Spoken language Translation : The ATR-MATRIX Approach", Proc. MT Summit, pp.229-235, 1999.

[3] F. Sugaya,T. Takezawa, A. Yokoo, Y. Sagisaka, S. Yamamoto, "Evaluation of the ATR-MATRIX Speech Translation System with Pair Comparison Method Between the System and Humans", Proc. ICSLP, pp.1105-1108, 2000.

[4] K. -Y. Su, M. -W. Wu, and J. -S. Chang, "A new quantitative quality measure for machine translation systems", Proc. COLING, pp.433-439, 1992.

[5] T. Takezawa, F. Sugaya, A. Yokoo, S. Yamamoto, "A New Evaluation Method for Speech Translation Systems and a Case Study on ATR-MATRIX from Japanese to English", Proc. MT Summit, pp.299-307, 1999.

[6] T. Morimoto, N. Uratani,T. Takezawa, O. Furuse, Y. Sobashima, H. Iida, A. Nakamura, Y. Sagisaka, N. Higuchi, and Y. Yamazaki "A Speech and Language Database for Speech Translation Research", Proc. ICSLP, pp.1791-1794, 1994.

[7] T. Takezawa, "Buildng a Biligual Travel Conversation for Speech Translation Research", Proc. 2nd International Worshop on East-Asian Language-Resources and Evaluation —Oriental COCOSDA Workshop '99, pp.17-20, 1999.

[8] 古瀬蔵，山本和英，山田節夫，"構成素境界解析を用いた多言語話し言葉翻訳", 自然言語処理, vol.6, no.5, pp.63-91, 1999.

表目次

図目次

表１　正解として追加される文の例

Source language test sentence	Target language test sentence
はい/分かり/まし/た/お/調べ/し/ます/ので/少々/お/待ち/ください	All right. Please hold the line and I will check.

Source language similar texts	Target language similar texts
かしこまりました/お/調べ/いたし/ます/ので/少々/お/待ち/ください	Okay, let me check. Just a moment please.
はい/お/調べ/し/ます/少々/お/待ち/ください/ませ	Okay, could you wait for a moment while I check.
分かり/まし/た/確認/し/ます/ので/少々/お/待ち/ください	Okay, I 'll check for you please hold on a moment.
お/調べ/いたし/ます/ので/少々/お/待ち/ください	One moment please. I 'll check on availability.
ただいま/お/調べ/し/ます/ので/少々/お/待ち/ください/ませ	Could you hold on a minute while I check please.

表２　D ランクであるが正解群類似度が大きくなる例

Example 1	Source language test sentence	コネクティングルームが一泊五万七千円となっております
	Recognition result	コネクティングルームが一泊五〇七千円となっております
	Correct answer	A connecting room is fifty seven thousand yen per night .
	Translation result	A connecting room is five zero seven thousand yen per night .
	Answer set similarity	0.8
Example 2	Source language test sentence	今ワシントンのワシントンホテルに滞在しています
	Recognition result	今ワシントンの足のホテルに滞在しています
	Correct answer	I 'm staying at the washington hotel in Washington .
	Translation result	I 'm staying at the foot hotel in Washington now .
	Answer set similarity	0.78
Example 3	Source language test sentence	二一三五四三の一七五五
	Recognition result	二三五四三の一七五五
	Correct answer	Two one three five four three one seven five five .
	Translation result	Two three five four three , one seven five five .
	Answer set similarity	0.9

表3　2クラス判別の4つの確率

		Subjective evaluation	
		class 1	*class 2*
Automatic discrimination	*CLASS 1*	*P(CLASS 1 | class 1)*	*P(CLASS 1 | class 2)*
	CLASS 2	*P(CLASS 2 | class 1)*	*P(CLASS 2 | class 2)*

表4　TOEIC 換算点

	ATR-MATRIX	TDMT
Paired comparison method (manual)	548	708
Proposed method (automatic)	618	705

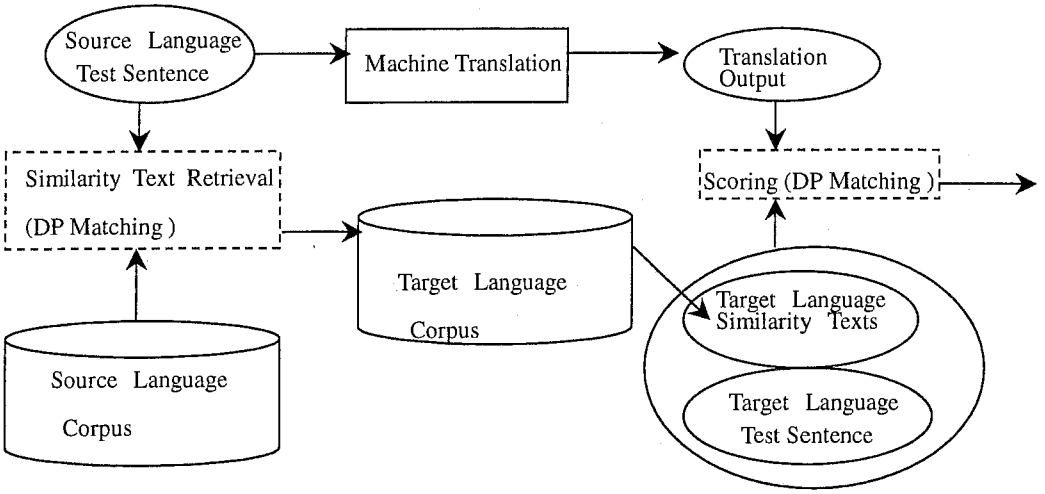

図１　正解文追加類似度計算法の処理の流れ

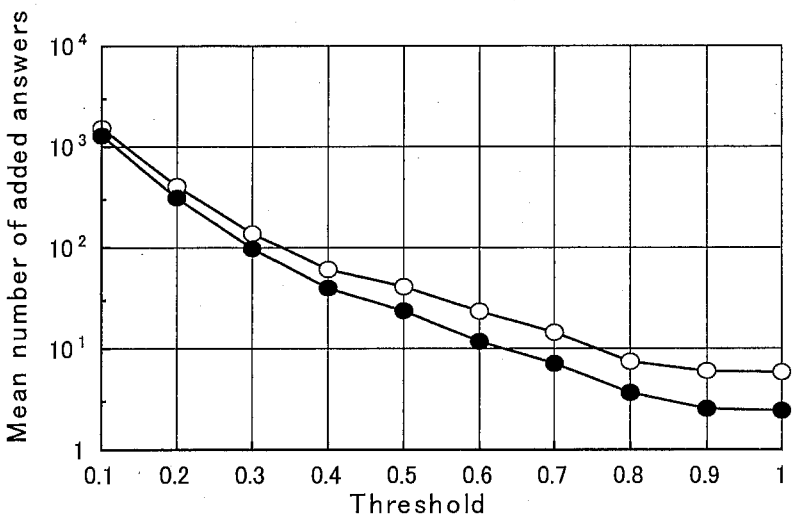

図２　類似文検索閾値と追加される正解文数の関係

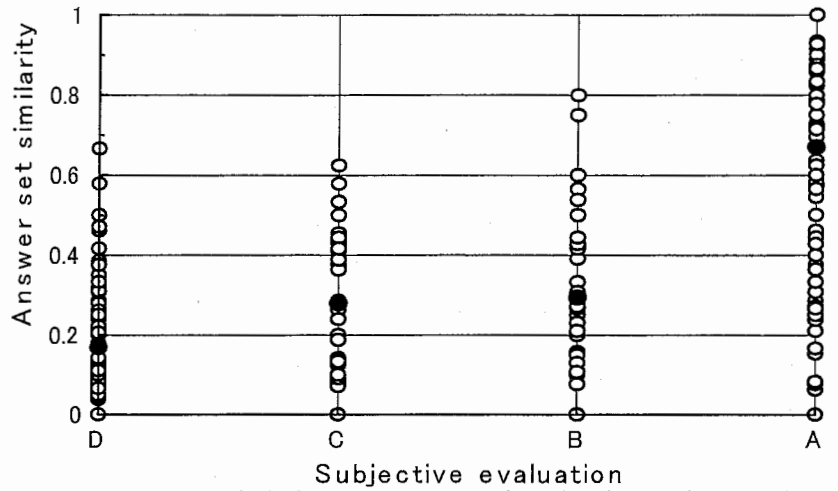

図３　言語翻訳部の翻訳ランクと正解群類似度の関係

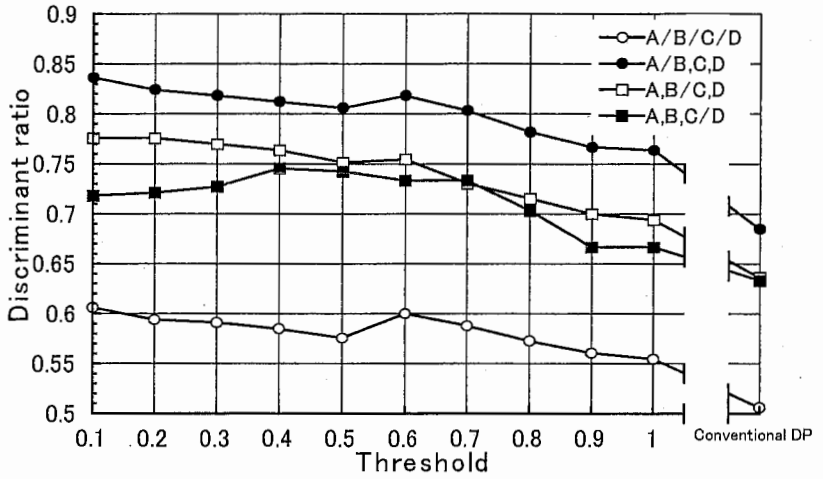

図４　類似文検索閾値と判別率の関係

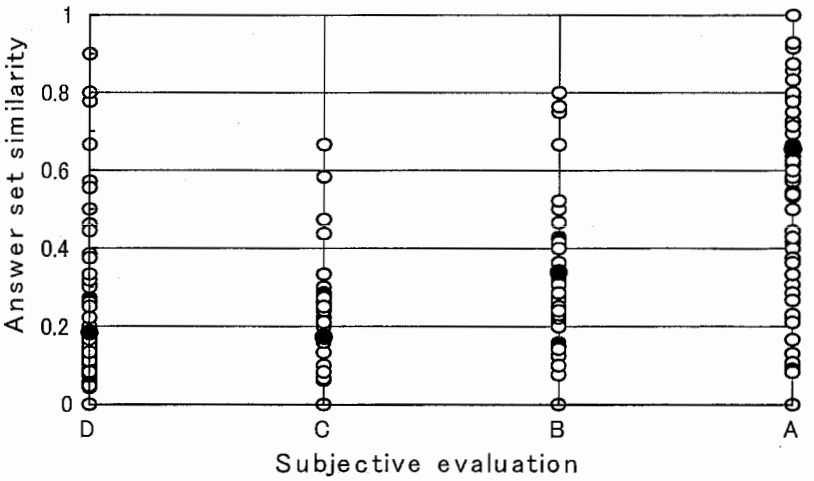

図５　音声認識部を含めた場合の翻訳ランクと正解群類似度の関係

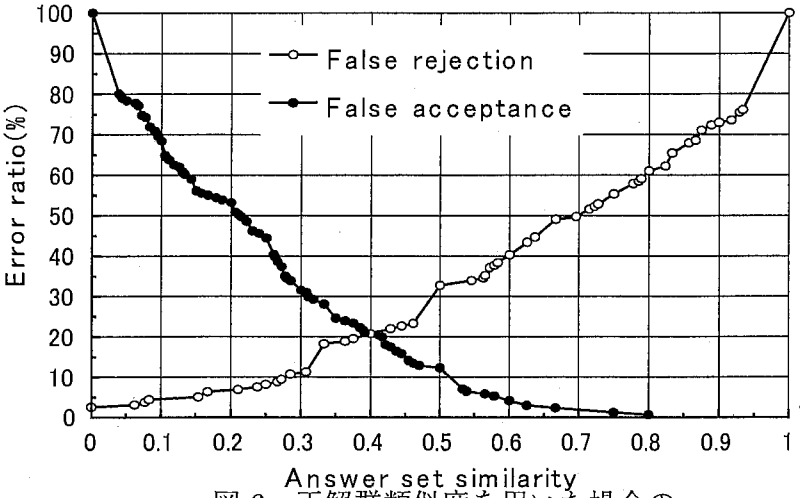

図 6　正解群類似度を用いた場合の
判別閾値と誤り率の関係

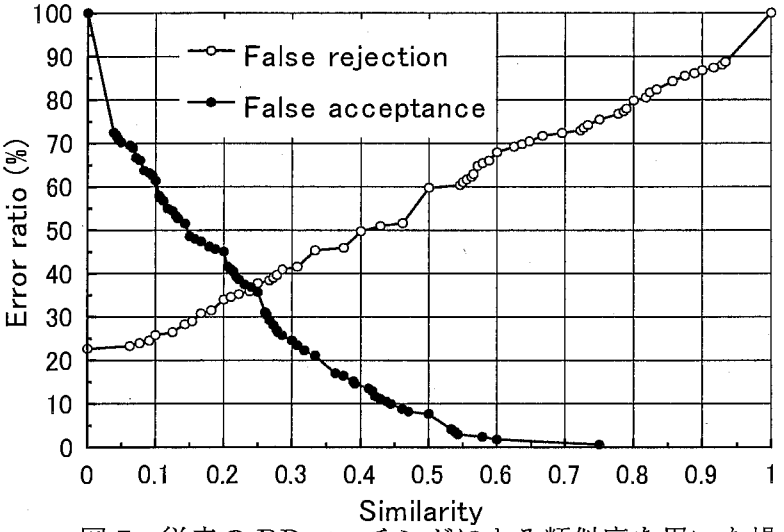

図 7　従来の DP マッチングによる類似度を用いた場合の
判別閾値と誤り率の関係

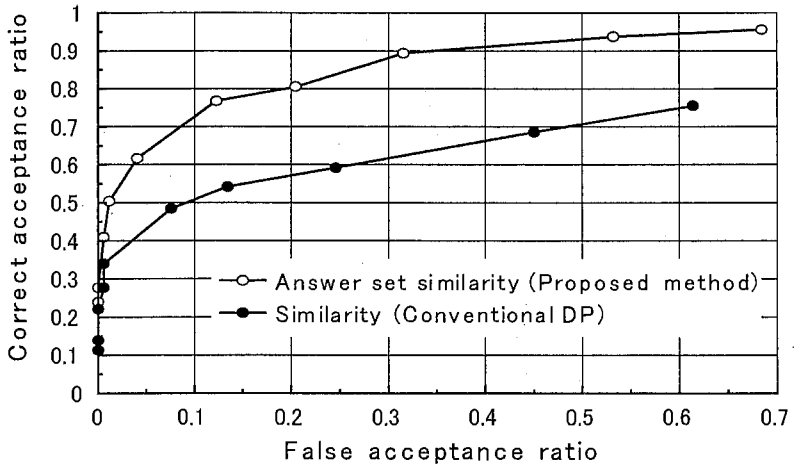

図 8　クラス 1 受理率とクラス 2 誤り受理率の関係

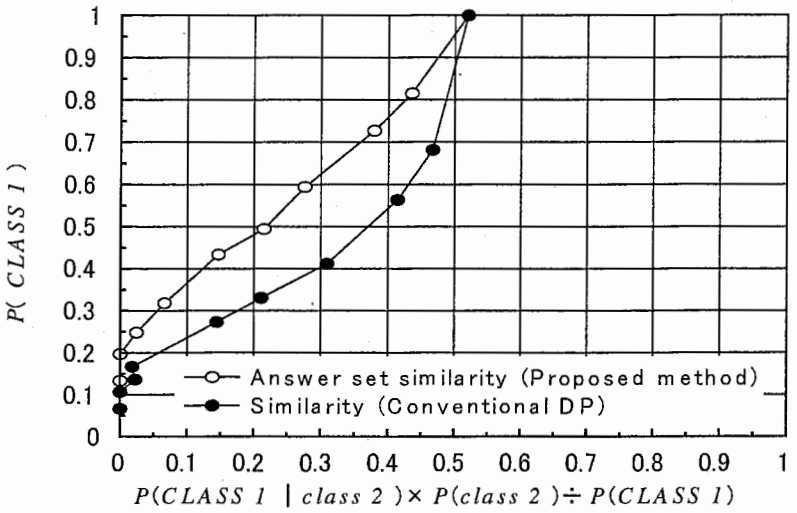

図 9　削減される評価コストと誤りの関係

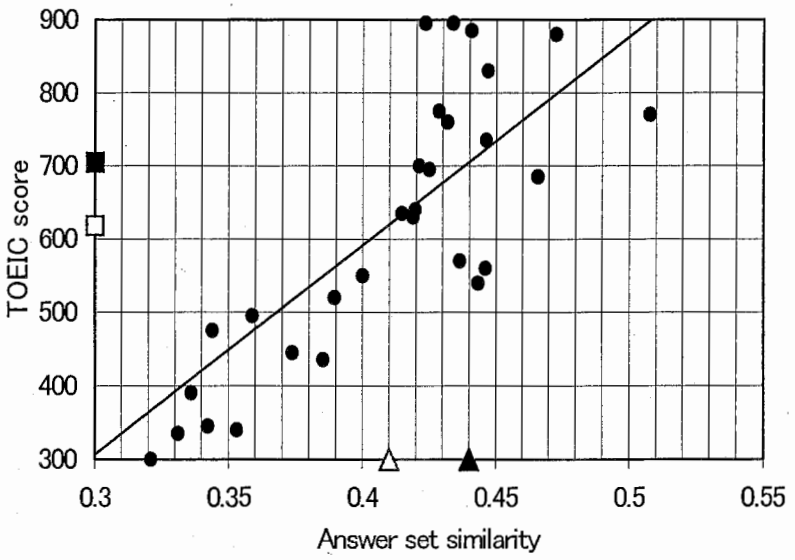

図 10　TOEIC スコアと正解群類似度の関係