概要

現在，機械翻訳の翻訳品質の評価において，数多くの自動評価方法が提案されている．多くの評価方法は，翻訳された最尤の出力文と参照文から単語の順列や出現頻度を見て評価を行う．つまり，最尤の出力文 1 文に対して評価を行っている．しかし，実際に翻訳を行う際には，複数の出力文の中から最適な文を選び，翻訳を行うことがある．一方，情報検索においては最尤の文だけでなく，複数の文を使用し検索精度を評価する．評価指標の一つとして MRR がある．そこで，本研究では，機械翻訳において，MRR を参考にして，複数の出力文を使用する自動評価方法を提案した．

そして，日英翻訳と英日翻訳の 2 種類，単文と重文複文の 2 種類，評価方法 2 種類，翻訳システム 7 種類の合計 56 種類の実験を行った．人手評価に対する自動評価の相関係数を調査した結果，提案手法の結果と 1 文出力の結果を比較しても，人手評価に対する自動評価の相関係数に差はあまり見られなかった．しかし，単文の英日翻訳において，提案手法の結果は 1 文出力の結果と比較すると，人手評価に対する自動評価の相関係数が向上した．

目 次

図 目 次

表 目 次

第1章　はじめに

現在，機械翻訳の翻訳品質の評価において，数多くの自動評価方法が提案されている．

多くの評価方法は，翻訳された最尤の出力文と参照文から単語の順列や出現頻度を見て評価を行う．つまり，最尤の出力文1文に対して評価を行っている [1]．しかし，実際に翻訳を行う際には，複数の出力文の中から最適な文を選び，翻訳を行うことがある．翻訳品質の評価において，複数の出力文を使用する評価方法は見当たらない．

一方，情報検索においては，入力文1文に対し，複数の文を出力する．出力された複数の文に対し，評価指標を使用して評価を行う．評価指標の一つとして MRR[2] がある．そこで，本研究では，機械翻訳において，MRR を参考にして，複数の出力文を使用する自動評価方法を提案した．人手評価に対する自動評価の相関係数を調査した結果，提案手法の結果と1文出力の結果を比較しても，人手評価に対する自動評価の相関係数に差はあまり見られなかった．しかし，単文の英日翻訳において，提案手法の結果は1文出力の結果と比較すると，人手評価に対する自動評価の相関係数が向上した．ここで，本論文の構成を以下に示す．第2章で，翻訳システムの説明を行う．第3章で，自動評価と人手評価の説明を行う．第4章で，MRR の説明を行う．第5章で，提案手法の説明を行う．第6章で，実験環境の説明を行う．第7章で，実験の結果を示す．第8章で，本研究の考察を述べる．第9章で，結論を述べる．

第2章　翻訳システム

2.1　ルールベース翻訳

本研究では，ルールベース翻訳を”RBMT”と表記する．

RBMT は，人手によって構築された変換規則を元に翻訳を行うシステムである．長所として，規則を厳密に定義するので，規則が存在する翻訳において，精度が高い．しかし，短所として，規則が存在しない翻訳において，精度が低い．さらに人手によって規則を構築するため，開発コストが高い．

一般的な RBMT の手順を以下に示す．

手順 1 辞書の品詞などから原言語の構文解析を行う．

手順 2 目的言語の語順に変換する．

手順 3 再度辞書を参照し，助詞，助動詞などの不足語を補い，目的言語の出力文を生成する．

2.2 句に基づく統計翻訳

本研究では，句に基づく統計翻訳を，”PSMT”と表記する．

統計翻訳は，文法構造が近い言語間で翻訳精度が高い．しかし，文法構造の異なる言語間で翻訳精度が低い．

PSMT の概略を図 2.1 に示す．

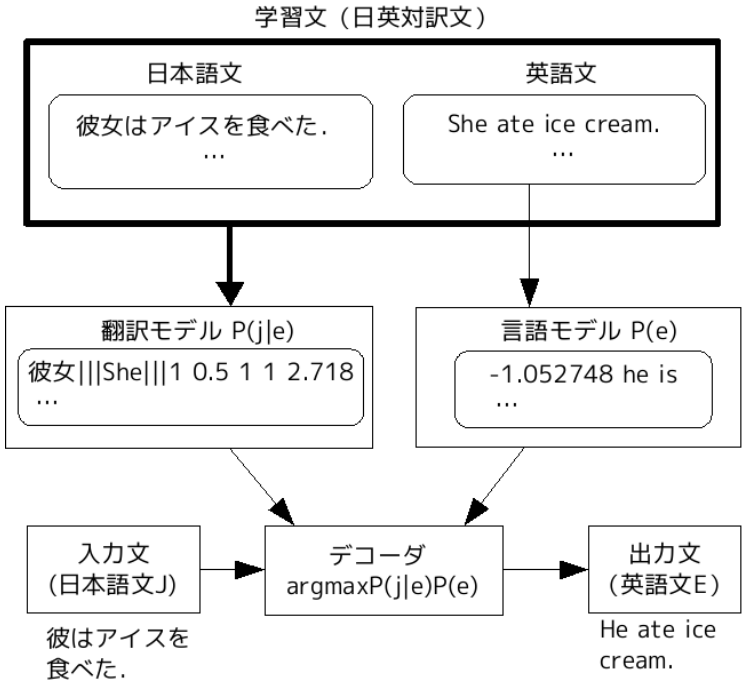

図 2.1 句に基づく統計翻訳

日英統計翻訳システムは，入力文（日本語文 J）が与えられたとき，デコーダを用いて翻訳モデル $P(j|e)$ と言語モデル $P(e)$ の確率を組み合わせ，確率が最大となる英語文 E を求めて翻訳を行う．$P(j|e)$ は e が j に翻訳される確率である．式をベイズの定理を用いて以下に示す．

$$E = argmax_e P(e|j) \tag{2.1}$$

$$= argmax_e \frac{P(j|e)\,P(e)}{P(j)} \tag{2.2}$$

$$= argmax_e P(j|e)\,P(e) \tag{2.3}$$

2.2.1 翻訳モデル

翻訳モデルは，単語または単語列の翻訳確率を組み合わせるモデルである．日英翻訳において，日本語の単語列から英語の単語列へ確率的に翻訳を行うため用いる．また，翻訳モデルは，表 2.1 のようなフレーズレーズテーブルで管理されている．

表 2.1 フレーズテーブルの例

おもしろい 本 ||| interesting book ||| 1 0.157258 1 0.180402 2.718
おもしろい 話 ||| funny stories ||| 0.5 0.0112613 0.2 0.000721527 2.718
おもちゃ 箱 ||| toy box||| 0.125 0.037 0.142 0.201 2.718
タイ 政府 ||| Thai government ||| 0.5 0.2778 0.5 0.093438 2.718

左から，日本語フレーズ，英語フレーズ，フレーズの英日翻訳確率 $P(j|e)$，英日方向の単語翻訳確率の積，フレーズの日英方向の翻訳確率 $P(e|j)$，日英方向の単語翻訳確率の積，フレーズペナルティ(一定) となっている．

2.2.2 言語モデル

言語モデルは，単語または単語列に対して，生成確率を付与するモデルである．日英翻訳では，言語モデルを用いて，生成された翻訳候補から英語を選出する．統計翻訳では一般に，N-gram モデルを用いる．N-gram モデルの例を表 2.2 に示す．なお，表 2.2 は，2-gram(2 単語間) である．

表 2.2 N-gram モデルの例

-1.782704 ||| I am ||| -0.04873917
-1.610493 ||| that is ||| -0.01120672
-2.346281 ||| train goes ||| -0.09572452
-1.868116 ||| woman and ||| -0.1343922

表 2.2 において，一番上の行は，左から，“I”の後に “am”が続く確率を常用対数で表した値 “$\log_{10}(P(am|I)) = -1.782704$”，2-gram で表現された単語列 “I am”，バックオフスムージングにより推定された “I”の後に “am”が続く確率を常用対数で表した値 “$\log_{10}(P(am|I)) = -0.04873917$”である．

バックオフスムージングとは，高次の N-gram の値が存在しない場合，低次の N-gram の値から推定する手法である．低次の確率を改良したスムージングの手法は，Kneser-Ney スムージングである．言語モデルの N-gram の作成においては，一般的に Kneser-Ney スムージングが用いられる．

2.2.3 デコーダ

デコーダは翻訳モデル $P(j|e)$ と言語モデル $P(e)$ を組み合わせて，確率が最大となる翻訳候補を探索し，出力する．デコーダの動作例を図 2.2 に示す．

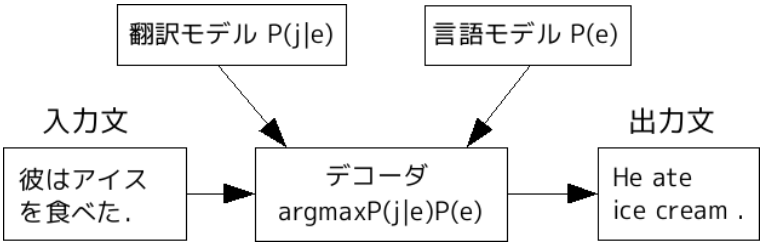

図 2.2 デコーダの動作例

日英翻訳において，$\arg\max_e P(j|e)P(e)$ の確率が最大となる英語文を出力するために，日本語と英語の単語対応を適切な順序で選択する必要がある．しかし，全探索をおこなうには，膨大な計算量と時間が必要となる．そこで，計算量と時間を削減するために，ビームサーチ法を用いる．

2.2.4 パラメータチューニング

パラメータチューニングは，デコーダで用いるパラメータの最適化を行う．一般的に評価関数 (BLEU) を最大にする翻訳結果が選ばれるように，パラメータ調整を行う．なお，パラメータ調整に，試し翻訳を行うデータとして，ディベロップメントデータを用いる．各文に対して上位 100 個程度の翻訳候補を出力し，重みを変えて翻訳候補が上位にくるようにパラメータを調整する．

2.3 階層型統計翻訳

本研究では，階層型統計翻訳を，”HSMT”と表記する．

HSMT とは，階層句を用いて翻訳を行う統計翻訳システムである．句を階層にすることで構文の評価が可能となる．また階層型統計翻訳は，語の並び替えを文脈自由文法で表現する．HSMT の概略を図 2.3 に示す.

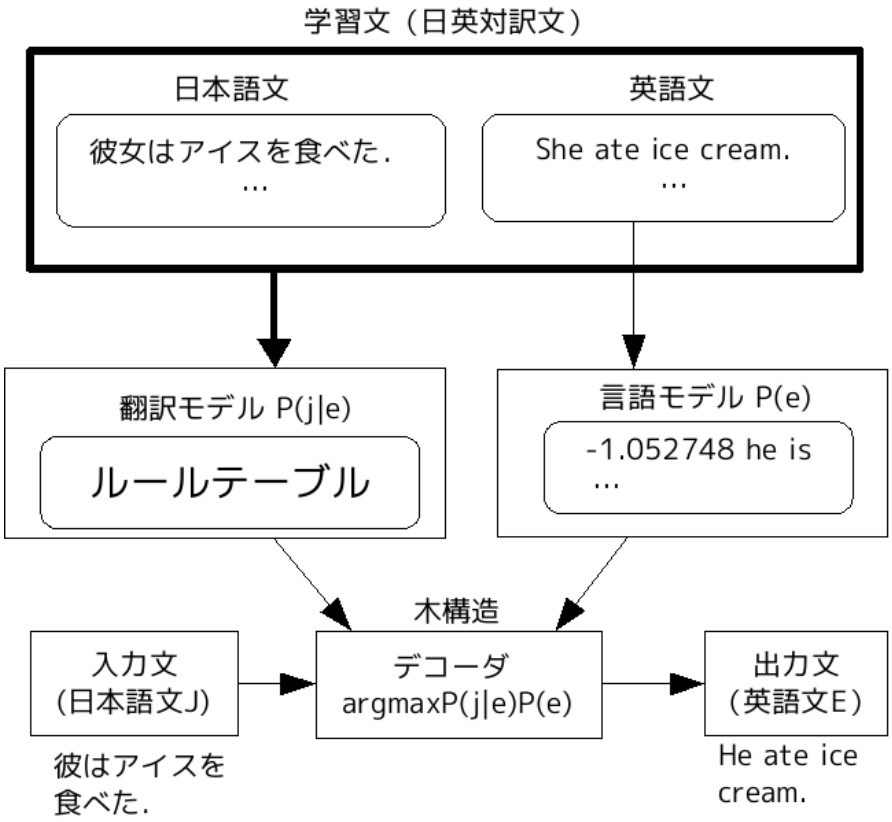

図 2.3 階層型統計翻訳システムの枠組み

HSMT は，翻訳モデルにルールテーブルを用いる点で，PSMT と異なる．さらにデコーダで，木構造を用いて翻訳を行う点も異なる．

2.3.1 翻訳モデル

階層型統計翻訳における翻訳モデルには，ルールテーブルを用いる．ルールテーブルの例を表 2.3 に示す.

表 2.3 日英ルールテーブルの例

[X] ||| これ は ||| This is ||| 0.138021 0.0102929 0.140684 0.0110294
[X] ||| これ は [X,1][X,2] ||| This is a [X,1][X,2] ||| 0.5879 0.03111 0.1734 0.00075
[X] ||| これ は [X,1][X,2] です 。 ||| This is a [X,1][X,2] . ||| 0.038 0.00018 0.7 0.0076
[X] ||| 新しい ||| a new ||| 0.165997 0.39984 0.029316 0.0253706
[X] ||| 新しい ||| new ||| 0.460302 0.39984 0.79316 0.884752

左から，非終端記号，日本語ルール，英語ルール，ルールの英日翻訳確率 $P(j|e)$，英日方向の単語翻訳確率の積，ルールの日英方向の翻訳確率 $P(e|j)$，日英方向の単語翻訳確率の積である．

2.3.2 デコーダ

デコーダは翻訳モデル $P(j|e)$ と言語モデル $P(e)$ を組み合わせて，確率が最大となる翻訳候補を探索し，出力する．

日英翻訳において，$\arg\max_e P(j|e)P(e)$ の確率が最大となる英語文を出力するために，言語モデルと翻訳モデルを用いて翻訳を行う．しかしデコーダにおいて，木構造で翻訳を行うため，適切な英語文を決定させるために，膨大な計算量と時間が必要となる．

HSMT におけるデコーダの動作例を示す．なお，階層句を表 2.4 に示し，デコーダの動作の例を図 2.4 に示す．

表 2.4 階層句の例

X1 found that X2	X1 は X2 だとわかった。
She is X3	彼女が X3 だ
a music teacher	音楽の先生
My mother	私の母

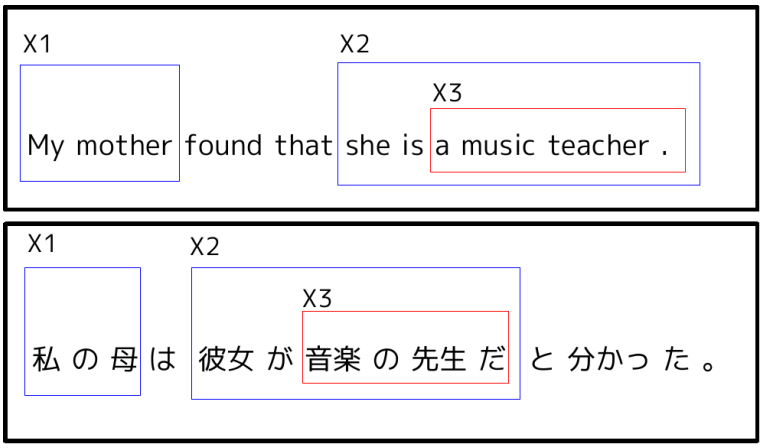

図 2.4 デコーダの動作例

2.4 ハイブリッド翻訳

2.4.1 概要

ハイブリッド翻訳は，前処理にルールベース翻訳を用いる．そして後処理に統計翻訳を使用する翻訳システムである．

後処理に句に基づく統計翻訳を用いる場合，“RBMT+PSMT”と表記する．また，後処理に階層型統計翻訳を用いる場合，“RBMT+HSMT”と表記する．本研究では，英英統計翻訳を英’英統計統計翻訳と定義する．RBMT+PSMT の概略を図 2.5 に示す．

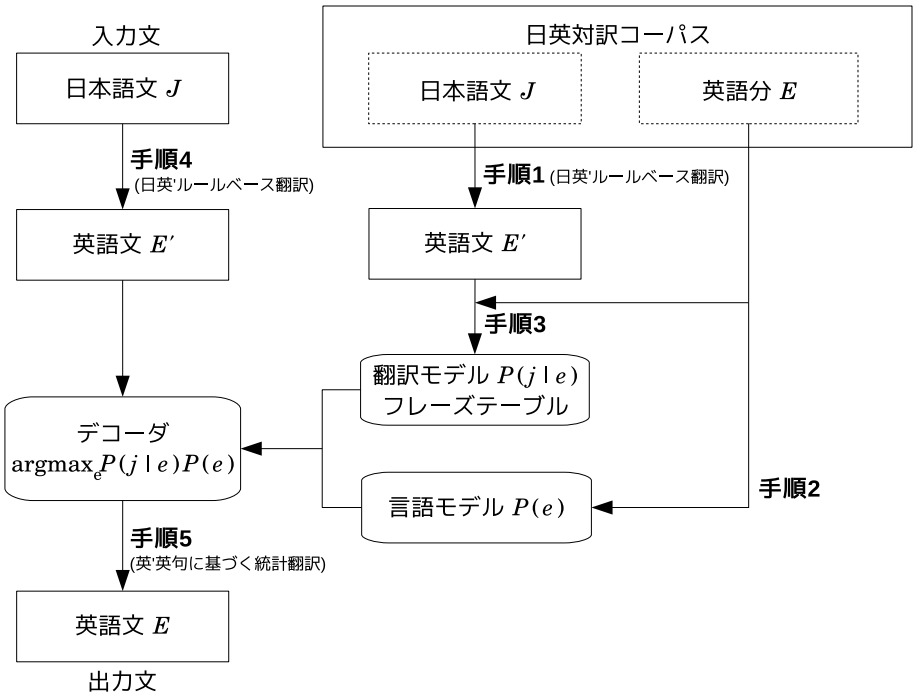

図 2.5 日英ハイブリッド翻訳の枠組

2.4.2 手順

学習の手順を以下に示す

手順 1 ルールベース翻訳を用いて，日英対訳コーパスの日本語文を英' 語文に翻訳する．翻訳例を以下に示す．

入力文 (日本語文)	あの 人 の 家 は すぐ 見つかっ た 。
出力文 (英' 語文)	That person 's house was found immediately .
参照文	I soon found that person 's house .
入力文 (日本語文)	列車 が 着い て いる 。
出力文 (英' 語文)	The train has reached .
参照文	The train is in .
入力文 (日本語文)	コーヒー が 飲み たい 。
出力文 (英' 語文)	I would like to drink coffee .
参照文	I 'd like some coffee .

手順 2 手順 1 で作成した英’語文と日英対訳コーパスの英語文を用いて，翻訳モデルを作成する．英’英フレーズテーブルの例を以下に示す．

I polished	||| I polished	||| 1 0.0388231 1 0.0748095 2.718
got injured .	||| was fatally wounded .	||| 1 0.0479841 1 2.72564e-05 2.718
is dancing	||| dancing	||| 0.037037 0.00659718 0.166667 0.333333 2.718

手順 3 日英対訳コーパスの英語文を用いて，言語モデルを作成する．N-gram モデルの例を以下に示す．

-3.425136	His computer
-3.494154	our TV
-0.1251315	due to

翻訳の手順を以下に示す

手順 1 ルールベース翻訳を用いて，テスト文の日本語文を英’語文に翻訳する．翻訳例を以下に示す．

入力文 (日本語文)	ウイスキー を 1 杯 もらお う 。
出力文 (英’語文)	I will get whiskey one cup .
参照文	I ’ll have a whiskey .
入力文 (日本語文)	この 理論 は くずれる だろ う 。
出力文 (英’語文)	This theory will collapse.
参照文	This theory won ’t hold water .
入力文 (日本語文)	手続き は 個人 でし て 下さい 。
出力文 (英’語文)	Procedure is an individual and please give it to me .
参照文	Carry out the procedure by yourselves , please .

手順 2 手順 1 で作成した英’語文を入力文として，英’英統計翻訳を行う．なお，翻訳モデル，言語モデルは手順 2，手順 3 で作成されたものを使用する．翻訳例を次のページに示す．

入力文 (英’ 語文)	I will get whiskey one cup .
出力文 (英語文)	Let ’s get whiskey a cup .
参照文	I ’ll have a whiskey .
入力文 (英’ 語文)	This theory will collapse.
出力文 (英語文)	This theory will fail .
参照文	This theory won ’t hold water .
入力文 (英’ 語文)	Procedure is an individual and please give it to me .
出力文 (英語文)	There is an individual Please give it to me .
参照文	Carry out the procedure by yourselves , please .

2.5 各システムの翻訳例

RBMT, PSMT, HSMT, RBMT(a)+PSMT, RBMT(a)+HSMT, RBMT(t)+PSMT, RBMT(t)+HSMT の翻訳例を表 2.5, 表 2.6, 表 2.7 に示す.

表 2.5 各システムの翻訳例 1

入力文	電気 コンロ の コイル が 焼き 切れ た 。
RBMT	The coil of the electric cooker was able to be burned off .
PSMT	The The buckets or of electricity .
HSMT	The electric The of the cooking stove .
RBMT+PSMT	The company of the electric cooker was burned out .
RBMT+HSMT	The heavy coil in the electric cooker was burned out .
参照文	The heater coil is burnt out .

表 2.6 各システムの翻訳例 2

入力文	もっと 右 へ 寄っ て ください 。
RBMT	Please come to visit the right more .
PSMT	more to the right .
HSMT	more to the right .
RBMT+PSMT	Please come visit to the right .
RBMT+HSMT	Please come visit to the right .
参照文	Please move over more to the right .

表 2.7 各システムの翻訳例 3

入力文	彼 の 考え方 は 極端 すぎる 。
RBMT	His view is too going too far.
PSMT	His way of thinking is too the extreme .
HSMT	His way of thinking is too the extreme .
RBMT+PSMT	His way of thinking is too going too far .
RBMT+HSMT	His way of thinking is too going too far .
参照文	His way of thinking goes too far .

第3章　評価手法

3.1　人手評価

人手評価は，利点として，文法や意味を正確に評価可能である．しかし欠点として，時間と人件費が膨大にかかるため，大量の文の評価は難しい．

人手評価には，様々な評価方法がある．大きく分けて2種類ある．絶対評価と相対評価である．絶対評価には，了解度と正確さの観点から9段階で評価を行う手法，Adequacy(内容としての適切さ) と Fluency(言語としての流暢さ) の観点からそれぞれ5段階で評価を行う手法，さらに10点満点で評価を行う手法などがある．相対評価には，2つの翻訳システムの翻訳結果を比較して，翻訳品質が高いほうを良い評価とする，対比較評価などがある．

例として，Adequacy と Fluency の5段階評価の評価基準を表3.1に示す．

表 3.1　Adequacy と Fluency の評価基準

ランク	Adequacy	Fluency
5	入力文の意味が全て伝わっている．	かなり読みやすい．
4	入力文の意味がほとんど伝わっている．	少し読みやすい．
3	入力文の意味は伝わっている．	ほとんど変わらない．
2	入力文の意味が少ししか伝わっていない．	少し読みにくい．
1	入力文の意味が全く伝わっていない．	かなり読みにくい．

3.2 自動評価

3.2.1 BLEU

BLEU[3] は，機械翻訳システムの自動評価において，現在主流な評価法である．BLEU は，N-gram 適合率で評価を行う．実験では 4-gram を用いる．BLEU は 0 から 1 のスコアを算出し，スコアが大きい方が良い評価である．BLEU の計算式を以下に示す．

$$BLEU = BPexpW_n \sum_{n=1}^{N} (log_e P_n) \tag{3.1}$$

$$W_n = \frac{1}{N} \tag{3.2}$$

$$P_n = \frac{\sum_i \text{出力文中} i \text{と参照文} i \text{で一致した} N-gram \text{数}}{\sum_i \text{出力文中} i \text{の中の全} N-gram \text{数}} \tag{3.3}$$

ここで，BP は短い翻訳文が高い評価にならないように補正を行うパラメータである．また W_n は N-gram の重みである．具体的な計算例を以下に示す．

例

日本語文：お先 に 失礼 し ます 。
参照文：Excuse me , I must be going now .
出力文：Excuse me , but I mest be going now .

計算方法

参照文と出力文の N-gram より計算を行うと

$$P_1 = \frac{9}{10}, P_2 = \frac{7}{9}, P_3 = \frac{5}{8}, P_4 = \frac{3}{7}, W_1 = 1, W_2 = \frac{1}{2}, W_3 = \frac{1}{3}, W_4 = \frac{1}{4} \tag{3.4}$$

これらのスコアを計算式に代入すると

$$BLEU \text{ スコア} = e^{W_4(logP_1+logP_2+logP_3+logP_4)} \tag{3.5}$$

$$= e^{\frac{1}{4}(log\frac{9}{10}+log\frac{7}{9}+log\frac{5}{8}+log\frac{3}{7})} \tag{3.6}$$

$$= 0.6580 \tag{3.7}$$

また BLEU は，英語とフランス語などの文法構造が近い言語間において，人手評価と評価が一致する場合が多い．しかし，英語と日本語などの文法構造が異なる言語間において，人手評価と評価が一致しない場合がある．原因として，BLEU は部分的な単語列の一致数を調べ，スコアを求めていることが挙げられる．そのため，参照文との比較に

おいて，同一の単語列を局所的に含む出力文が高いスコアを算出する．したがって，出力文において，文法的な誤りが存在しても高いスコアを算出してしまう．表 3.2 に具体的な例文を示す．なお，表 3.2 に対応する BLEU スコアを表 3.3 に示す．

表 3.2 翻訳例

入力文	その 機械 の 構造 に は 欠陥 が ある 。
出力文 1	The structure of the machine has a defect .
出力文 2	The structure of **the is** a fault in the machine .
参照文	There is a fault in the machine 's construction .

表 3.3 1 文における BLEU スコア

出力文 1	BLEU = 0.000
出力文 2	BLEU = 0.367

表3.3 より，出力文1 と出力文2 を比較すると，1 文における BLEU スコアは，出力文 2 が良い評価 となる．しかし出力文 2 は “the is” と出力されているので，文法的に誤っている．

3.2.2 METEOR

METEOR[4] は，単語属性が正しい場合に高いスコアを出す．実験では *uni*-gram を用いる．METEOR は 0 から 1 までのスコアを出力し，スコアの大きい方が評価が良い評価である．計算式を以下に示す．

$$F\text{値} = \frac{P \times R}{\alpha \times P + (1-\alpha) \times R} \tag{3.8}$$

$$Pen = \gamma \times (\frac{c}{m})^{\beta} \tag{3.9}$$

$$METEOR = F \times (1 - Pen) \tag{3.10}$$

METEOR は F 値，ペナルティ関数 Pen を用いて計算される．F 値は適合率 P と再現率 R の調和平均で求められる．そしてペナルティ関数 Pen において，m は参照文と出力文の間で一致した単語数を示す．また c は，一致した単語を対象として，参照文と一致

する単語列を１つのまとまりに統合した際のまとまりの数を示す．したがって，参照文と出力文が同一文である場合は c=1 となる．なお α，β，γ の値はパラメータである．具体的な計算例を以下に示す.

例

日本語文：お先 に 失礼 し ます 。
参照文：Excuse me , I must be going now .
出力文：Excuse me , but I mest be going now .

計算方法

参照文Bと出力文A，AとBの重複部分Cとする．またパラメータ $\alpha = 0.8, \beta = 2.5, \gamma = 0.4$ とする.

$$\text{適合率 } P = \frac{C}{A} = \frac{9}{10} \tag{3.11}$$

$$\text{再現率 } R = \frac{C}{B} = \frac{9}{9} \tag{3.12}$$

$$F \text{ 値} = \frac{P * R}{\alpha * P + (1 - \alpha) * R} = \frac{45}{46} \tag{3.13}$$

$$\text{ペナルティ関数 } Pen = \gamma * (\frac{c}{m})^{\beta} = 0.4 * (\frac{2}{9})^{2.5} = 0.00931169\cdots \tag{3.14}$$

$$METEOR \text{ スコア} = F * (1 - Pen) \tag{3.15}$$

$$= \frac{45}{46} * (1 - 0.0093) \tag{3.16}$$

$$= 0.9692 \tag{3.17}$$

3.2.3 RIBES

RIBES[5] は，参照文と出力文との間で，共通単語の出現順序を順位相関係数で評価を行う評価法である．計算式を以下に示す．

$$RIBES = NSR \times P^{\alpha} \tag{3.18}$$

$$RIBES = NKT \times P^{\alpha} \tag{3.19}$$

$$NSR = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} \tag{3.20}$$

$$NKT = \frac{\sum_{i=1}^{n-1} K_i - \sum_{i=1}^{n-1} L_i}{\frac{n(n-1)}{2}} \tag{3.21}$$

ここで，NSR はスピアマンの順位相関係数であり，NKT はケンドールの順位相関係数である．P は共通単語が少ない場合のペナルティである．また α はペナルティに対する重みとして使用され，$0 \leqq \alpha \leqq 1$ の値である．n は文の単語数であり，d_i は参照文の i 番目の単語と出力文の i 番目の単語の語順の差分である．
K_i は，以下の 2 つの単語列の共起数である．

- 出力文における，出力文の i 番目の単語以降の単語列．
- 参照文における，出力文の i 番目の単語以降の単語列．

L_i は，以下の 2 つの単語列の共起数である．

- 出力文における，出力文の i 番目の単語以降の単語列．
- 参照文における，出力文の i 番目の単語以前の単語列．

RIBES は，単語の出現順を順位相関係数を用いて評価することで，文全体の語順に着目することができる．なお，RIBES は 0 から 1 のスコアを出力し，スコアが大きい方が良い評価である．具体的な計算例を次のページに示す．

例

日本語文：雨 に 濡れた ため ， 彼 は 風邪 を 引い た 。
参照文：He caught a cold because he got soaked in the rain .
出力文：He got soaked in the rain because he caught a cold .

計算方法

$$
\begin{aligned}
d_1 &= 1-8=-7, d_2=2-9=-7, d_3=3-10=-7, \\
d_4 &= 4-11=-7, d_{本}\text{研究では，後処理に句に基づく統計翻訳を用いる場合,“}RBMT+P\\
d_7 &= 7-2=5, d_8=8-3=5, d_9=9-4=5, \\
d_{10} &= 10-5=5, d_{11}=11-6=5
\end{aligned}
$$

$$
\begin{aligned}
NSR &= 1-\frac{6(6*(-5)^2+(-2)^2+4*(-7)^2)}{11^3-11}=-0.59 \\
RIBES &= \frac{-0.59+1}{2}=0.2050
\end{aligned}
$$

$$
\begin{aligned}
K_1 &= 5, K_2=4, K_3=3, K_4=2, K_5=1, \\
K_6 &= 0, K_7=0, K_8=3, K_9=2, K_{10}=1
\end{aligned}
$$

$$
\begin{aligned}
L_1 &= 5, L_2=5, L_3=5, L_4=5, L_5=5, \\
L_6 &= 5, L_7=4, L_8=0, L_9=0, L_{10}=0
\end{aligned}
$$

$$
\begin{aligned}
NKT &= \frac{21-34}{\frac{11*10}{2}}=-0.23 \\
RIBES &= \frac{-0.23+1}{2}=0.3850
\end{aligned}
$$

3.2.4 TER

TER[6] は，Translation Edit Rate の略で翻訳の誤り率を求める評価法である．計算式を以下に示す.

$$TER = \frac{\sum_i (\text{挿入語数}\, i + \text{置換語数}\, i + \text{削除語数}\, i + \text{シフト語数}\, i)}{\sum_i (\text{参照文}\, i \,\text{の平均単語数})} \tag{3.26}$$

分子は参照文と出力文の比較における編集操作数のことである．TER の編集操作は挿入，置換，削除，シフトの 4 種類の編集を行うことである．なお，TER はスコアが小さい方が良い評価である．具体的な計算例を以下に示す.

例

日本語文：お先 に 失礼 し ます 。
参照文：Excuse me , I must be going now .
出力文：Excuse me , but I mest be going now .

計算方法

例では挿入語数= 1 より，分子の編集操作数=1 である．また分母は参照文の平均単語数 =9 である.

$$\begin{aligned} TER\,\text{スコア} &= \frac{\sum_i (\text{挿入語数}\, i + \text{置換語数}\, i + \text{削除語数}\, i + \text{シフト語数}\, i)}{\sum_i (\text{参照文}\, i \,\text{の平均単語数})} && (3.27) \\ &= \frac{1}{9} && (3.28) \\ &= 0.1111 && (3.29) \end{aligned}$$

3.2.5 STR

STR(Sentence Translation Ratio)[7] は, 出力文と参照文の文一致数で評価を行う評価法である．文一致とは，出力文と参照文の完全一致である．文一致数とは，出力文と参照文が文一致した数である．

文一致の例を図 3.1 に示す．図 3.1 中の出力文と参照文の，太字で示した”He ruined his health .”が文一致している．

図 3.1 文一致の例

第4章　MRR(Mean Reciprocal Rank)

MRR とは 1 つの正解について正解が出現した順位を評価する指標である. 検索課題毎に正解が最初に現れた順位の逆数を求め (RR(Reciprocal Rank)), 全検索課題の RR を平均すること (MRR) で, システムの検索精度を評価する.
計算式を以下に示す.

r : 正解が出現した順位

N : 全検索課題数

$$RRi = \frac{1}{r} \tag{4.1}$$

$$MRR = \frac{1}{N}\sum_{i}^{N} RRi \tag{4.2}$$

第5章　提案手法

5.1　提案手法の概要

本研究では，MRR を参考にして，複数の出力文を使用して評価を行う．計算式を以下に示す．また，提案手法の枠組みを図 5.1 に示す．

r : 文が出現した順位

e : 評価値

N : 入力文 1 文に対し出力される文数

M : 入力文数

$$RR_{ki} = \frac{e}{r} \tag{5.1}$$

$$\text{評価値} = \frac{1}{M}\sum_{k}^{M}\frac{1}{N}\sum_{i}^{N} RR_{ki} \tag{5.2}$$

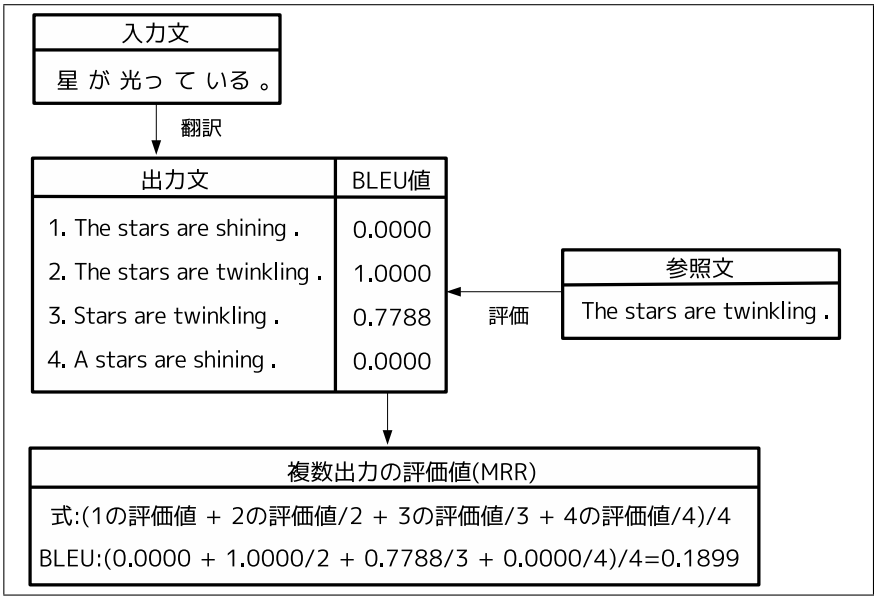

図 5.1　提案手法の枠組み

5.2 提案手法の手順

提案手法の手順を以下に示す.

手順 1 入力文 1 文を翻訳し，最尤の出力文から上位 4 文の出力文を得る.

手順 2 出力文 4 文の 1 文ずつに対し，評価を行う.

手順 3 5.1 節の式に対し，得られた評価値を使用し，複数出力に対する評価値を得る.

第6章　実験環境

6.1　翻訳システム

本研究の実験に使用した翻訳システムの実験環境を，以下で説明する．

6.1.1　ルールベース翻訳

ルールベース翻訳には，東芝の Tauras[8] を使用する．

6.1.2　句に基づく統計翻訳

6.1.2.1　翻訳モデルの学習

翻訳モデルの学習に，“train-model.perl[9]” を使用する．

6.1.2.2　言語モデルの学習

言語モデルの学習に，“SRILM[10]” の “ngram-count” を使用する．本研究では，N-gram モデルは 5-gram とする．またスムージングに，“Kneser-Ney discount”を使用する．

6.1.2.3　デコーダ

デコーダは”Moses[9]”を使用する．

6.1.2.4　パラメータチューニング

デコーダの Moses において，パラメータは，“mert-moses.pl[9]” を使用し，チューニングを行う．また，Moses[9] の設定ファイル “moses.ini”の修正を行う．“distortion-limit” の値は，パラメータチューニングで変更されない．よって，手作業で “distortion-limit” の値を，-1(無制限) に変更する．“distortion-limit”はフレーズの並び替えにおける制約である．

6.1.3 階層型統計翻訳

6.1.3.1 翻訳モデルの学習

翻訳モデルの学習に，“train-model.perl[9]” を使用する．

6.1.3.2 言語モデルの学習

言語モデルの学習に，“SRILM[10]” の “ngram-count” を使用する．本研究では，N-gram モデルは 5-gram とする．またスムージングに，“Kneser-Ney discount”を使用する．

6.1.3.3 デコーダ

デコーダは”Moses[9]”を使用する．

6.1.4 ハイブリッド翻訳

本研究では，前処理に Atlas を使用し，後処理に句に基づく統計翻訳を使用する場合，“RBMT(a)+PSMT”と表記し，後処理に階層型統計翻訳を用いる場合，“RBMT(a)+HSMT”と表記する．また，前処理に Tauras を使用し，後処理に句に基づく統計翻訳を使用する場合，“RBMT(t)+PSMT”と表記し，後処理に階層型統計翻訳を用いる場合，“RBMT(t)+HSMT”と表記する．

6.2 実験データ

本実験では 2 種類のコーパスを使用する．

6.2.1 単文コーパス

単文コーパス [12] は，日本語が単文である対訳コーパスである．コーパスの文は辞書の例文から抽出している．本実験では学習データとして 100,000 文，ディベロップメントデータとして 1,000 文を用いる．日英翻訳にはテスト文として 3,744 文，英日翻訳にはテスト文として 1,122 文を用いる．単文コーパスの例を表 6.1 に示す．

表 6.1 単文コーパスの例

日本語文	コンピューター は 2 進 法 の 2つ の 数 を 用いる 。
英語文	A computer employs the two digits of the binary system .
日本語文	彼 の 姿 は 暗闇 の 中 で 見え なかっ た 。
英語文	He was hidden by the darkness .
日本語文	その ビル は 倒壊 の 危険 が ある 。
英語文	The building is in danger of collapsing .

6.2.2 重文複文コーパス

重文複文コーパス [12] は，日本語が重文複文である対訳コーパスである．コーパスの文は，辞書の例文から抽出している．本実験では，学習データとして 100,000 文，ディベロップメントデータとして 1,000 文を用いる．日英翻訳にはテスト文として 5,435 文，英日翻訳にはテスト文として 1,280 文を用いる．重文複文コーパスの例を表 6.2 に示す．

表 6.2 重文複文コーパスの例

日本語文	腹 の 皮 が よじれる ほど 笑っ た 。
英語文	I almost split my sides laughing .
日本語文	ゆっくり と 踊り ながら 部屋 に 入っ て いっ た 。
英語文	She danced slowly into the room .
日本語文	彼 は それ を 聞い て 大層 心配 し て い た 。
英語文	He became deeply concerned at the news .

6.3 評価方法

6.3.1 人手評価

人手評価は，入力文と出力文と参照文を比較し，翻訳品質を 1 から 5 の数値でランクづけする手法で行う．ランク 1 がもっとも悪い評価で，ランク 5 がもっとも良い評価である．評価対象は，各翻訳システムの出力文から，ランダムに 400 文抽出した文である．評価基準を表 6.3 に，評価の例を表 6.4 から表 6.7 に示す．

表 6.3 評価基準

ランク	Adequacy
5	入力文の意味が全て伝わっている．
4	入力文の意味がほとんど伝わっている．
3	入力文の意味は伝わっている．
2	入力文の意味が少ししか伝わっていない．
1	入力文の意味が全く伝わっていない．

単文の日英翻訳の人手評価の例を以下に示す.

表 6.4 単文 日英翻訳 人手評価の例

	評価文	人手評価
入力文	この ボール は よく はずむ 。	5
出力文	The ball bounces well .	
参照文	This ball bounces well .	
入力文	彼ら は 残り の 食糧 を 等分 し た 。	5
出力文	They divided the rest of food equally .	
参照文	They divided the rest of the food equally .	
入力文	わたし は ときどき 自分 で すし を 握り ます 。	5
出力文	I sometimes make sushi myself .	
参照文	I sometimes make sushi myself .	
入力文	下水 が よく 流れ ない 。	4
出力文	The drain is not flow .	
参照文	The drain does not flow well .	
入力文	彼 の 信仰心 が ひどく 動揺 し た 。	4
出力文	His faith was shaken .	
参照文	His faith was severely shaken .	
入力文	この ボール は よく はずむ 。	4
出力文	The ball bounces .	
参照文	This ball bounces well .	
入力文	彼女 は 古い 着物 を はやり の 色 に 染め直した 。	3
出力文	She redyed the old kimono a popular color .	
参照文	She redyed an old kimono in a color that is in fashion .	
入力文	彼女 は 係長 に なっ た 。	3
出力文	She was a chief clerk .	
参照文	She became a supervisor .	
入力文	その 車 は 9 秒 で 最高 速度 に 達する 。	3
出力文	The car is at maximum speed 9 speeds .	
参照文	The car gets to speed in just nine seconds .	

	評価文	人手評価
入力文	日本 は 毎年 6 月頃 よく 雨 が 降る 。	2
出力文	In Japan are I often every year around June rains .	
参照文	It rains a lot in Japan around June of each year .	
入力文	カウンター パンチ を はね のけ た 。	2
出力文	Take puch the counter .	
参照文	He deflected the counterpunch.	
入力文	二度と 失敗 は 許さ れ ない 。	2
出力文	The failure again .	
参照文	We can’t afford another mistake .	
入力文	その 機械 の 使い方 が だれ に も わから なかっ た 。	1
出力文	Nobody knew how to use the machine .	
参照文	No one was able to guess how to use the machine .	
入力文	部屋 は そのまま に なっ て い た 。	1
出力文	We the room .	
参照文	I found the room as it had been when I left it .	
入力文	2 の 3 乗 は 8 で ある 。	1
出力文	The square is the two .	
参照文	The cube of 2 is 8 .	

単文の英日翻訳の人手評価の例を以下に示す.

表 6.5 単文 英日翻訳 人手評価の例

	評価文	人手評価
入力文	A carpenter made a wooden desk .	5
出力文	大工 は 木製 の 机 を 作っ た 。	
参照文	大工 が 木 の 机 を 作っ た 。	
入力文	It has gotten dark outside .	5
出力文	外 は 暗く な っ て いる 。	
参照文	外 が 暗く な っ た 。	
入力文	I put a kettle on the heater .	5
出力文	私 は ストーブ に やかん を かけ た 。	
参照文	ストーブ に やかん を かけ た 。	
入力文	The lecture was very boring and I slept through most of it .	4
出力文	講義 は とても 退屈 なので 私 は その 大部分 で 眠った 。	
参照文	その 講義 は たいへん 退屈 な もの だっ た ので 私 は その 大半 を 眠っ て い た 。	
入力文	I have half as many books as you .	4
出力文	あなたの 半分 の 本 を 持っ て いる 。	
参照文	私 は あなたの 半分 だけ の 本 を 持っ て い ます 。	
入力文	I bought much fruit .	4
出力文	果物 を 買っ た 。	
参照文	私 は 沢山 の 果物 を 買っ た 。	
入力文	It has gotten dark outside .	3
出力文	外 は 暗い なった 。	
参照文	外 が 暗く な っ た 。	
入力文	The Great Wall of China is over 2 , 400 km long .	3
出力文	外 は 暗い なった 。	
参照文	万里 の 長城 は ２ ４ ０ ０ キロ 以上 の 長 さ が ある 。	
入力文	Your actions are not in accordance with common sense .	3
出力文	あなた の 行動 は 常識 によって ない 。	
参照文	君 の 行動 は 常識 を はずれ て いる 。	

	評価文	人手評価
入力文	My bicycle chain came off on my way to school .	2
出力文	私 の 自転車 が 鎖 登校 の 途中 で 飛び立っ た 。	
参照文	登校 の 途中 で 自転車 の チェーン が 外れた 。	
入力文	That house is being refoofed now .	2
出力文	その 家 は 今 屋根 を し て いる 。	
参照文	あの 家 は 今 屋根 を 葺き 替え て いる 。	
入力文	My bicycle chain came off on my way to school .	2
出力文	私 の 自転車 が 鎖 登校 の 途中 で 飛び立っ た 。	
参照文	登校 の 途中 で 自転車 の チェーン が 外れた 。	
入力文	Civil aviation in Japan is very backward .	1
出力文	後方 に は 日本 の だ 。	
参照文	日本 の 民間 飛行 は さっぱり 振 わ ない 。	
入力文	I am full up to the throat .	1
出力文	私 は 喉 。	
参照文	おなか が 一杯 で 動け ない 。	
入力文	The scratch smarts .	1
出力文	雑記 用 の 才覚 。	
参照文	すり 傷 が ぴりぴり 痛む 。	

重文複文の日英翻訳の人手評価の例を以下に示す.

表 6.6 重文複文 日英翻訳 人手評価の例

	評価文	人手評価
入力文	大阪 へ 向かう 列車 の 中 で 友達 に 会っ た 。	5
出力文	I met a friend in the train to Osaka .	
参照文	I ran into a friend on the train for Osaka .	
入力文	生徒 を 手助け する の は 先生 の 仕事 だ 。	5
出力文	It is a teacher's work that helps a student .	
参照文	It is a teacher 's business to help his pupils .	
入力文	何 か 面倒 が ある 様子 だ 。	5
出力文	There seems to be some trouble .	
参照文	There seems to be some trouble .	
入力文	彼 は 信念 の ある 記事 を 書く 。	4
出力文	He writes an account of belief .	
参照文	He writes edgy articles .	
入力文	わが家 に は 台所 を 含め 、 全部 で 7つ の 部屋 が ある 。	4
出力文	My house is seven rooms in all including the kitchen .	
参照文	There are a total of seven rooms including a kitchen in my house .	
入力文	其の 内 雨 が 降る と 思う 。	4
出力文	I think it will rain .	
参照文	I think it will rain sooner or later .	
入力文	その 材料 を 組み合わせ て きわめて 軽い 飛行機 を 作る 。	3
出力文	The materials are together and made an extremely light airplane .	
参照文	Those materials are combined to create airplanes that are extremely light.	
入力文	これ を する こと は 経営 者 側 の 重大 な 義務 で ある 。	3
出力文	This is an obligation of the management .	
参照文	It is a serious duty of the management to do this .	
入力文	いかなる 大力 で も この 石 を 持ち上げる の は 骨 だ 。	3
出力文	It is a hard to lifts this stone .	
参照文	Whatever strength one may have , one will find it hard to lift this stone .	

	評価文	人手評価
入力文	人間 と 獣 を 分け て いる の は 知性 で ある 。	2
出力文	It is intelligence to the beasts man .	
参照文	It is intelligence that differentiates man from the beasts .	
入力文	部長 は 腰 の 低い 人 だ 。	2
出力文	The manager is a man .	
参照文	The manager is a modest man .	
入力文	音楽家 を 扱う に は こ つが 必要 で ある 。	2
出力文	He knows how to a musician .	
参照文	It requires tact to deal with musicians .	
入力文	針 一 本 落ち て も 聞こえ そう だ 。	1
出力文	A the sound of a book .	
参照文	A pin might be heard to drop .	
入力文	父 が 目 の 前 に いる ので 彼 は たばこ を 吸う の を 我慢 した 。	1
出力文	He is smoking in front of my father .	
参照文	The presence of his father inhibited him from smoking .	
入力文	彼 は 度 を 過ごす こと が 嫌い だ 。	1
出力文	He spends the times .	
参照文	He is abhorrent of excess .	

重文複文の英日翻訳の人手評価の例を以下に示す.

表 6.7 重文複文 英日翻訳 人手評価の例

	評価文	人手評価
入力文	I cannot believe him any longer .	5
出力文	もう これ 以上 彼 の 言う 事 は 信じ られ ない 。	
参照文	あの 人 の 言う こと は もう 本当に は でき ない 。	
入力文	He is wanting in that knowledge which is requiness at once .	5
出力文	彼 は 教師 として 必要 な 知識 を 欠い て いる 。	
参照文	彼 は 教師 たるもの に 必要 な 知識 を 欠い て いる 。	
入力文	I cannot believe him any longer	5
出力文	洗濯 機 の 脱水 機 は 遠心 力 を 利用 し て いる 。	
参照文	洗濯 機 の 脱水 機 は 遠心 力 を 利用 し た もの で ある 。	
入力文	A clown wobbled past on a unicycle .	4
出力文	ピエロ は 、 一輪車 で 通り過ぎ て いっ た 。	
参照文	道化師 が 一輪車 に 乗っ て よろよろ 通り過ぎ た 。	
入力文	I made the surface of the stone smooth by polishing it .	4
出力文	私 は 石 の 表面 を 、 それ を 磨く こと により 滑らか に なら せた 。	
参照文	石 の 表面 を 磨い て 滑らか に し た 。	
入力文	The firm wants to expand internationally .	4
出力文	同社 は 国際 的 に 拡大 し たい と 思っ て いる 。	
参照文	その 会社 は 国際 的 に 拡張 し たい と 願っ て いる 。	
入力文	He is wanting in that knowledge which is requisite to a teacher .	3
出力文	彼 は 先生 に は 必要 な 知識 が に は 足りない 。	
参照文	彼 は 教師 たる もの に 必要 な 知識 を 欠い て いる 。	
入力文	We have enough people here to play baseball .	3
出力文	我々 に は 、 プレー 野球 に対して 十分 に 人々 が ここ に いる 。	
参照文	野球 を する に は 人数 が じゅうぶん だ 。	
入力文	She has high ideals in life .	3
出力文	彼女 は 高い 理想 で 生活 を し て いる 。	
参照文	彼女 は 人生 に 高い 理想 を 抱い て いる 。	

	評価文	人手評価
入力文	It is bad for health to exercise hard right after a meal .	2
出力文	それ は 身体 に 悪い 食事 を した 後 に 右 の 運動 を し て い た 。	
参照文	食後 すぐ に 激しい 運動 を する の は 体 に 良く ない 。	
入力文	I somehow feel that it is wrong .	2
出力文	どう 間違っ て いる と 感じ て い た 。	
参照文	そんな こと を し て は なんだか 悪い よう な 気 が する 。	
入力文	I felt I was stepping on air .	2
出力文	私 は 空気 を 刺し て いる と 思っ た 。	
参照文	足 が 地 に つか ぬ 思い だっ た 。	
入力文	Out of five partners , one dropped out , leaving four .	1
出力文	５ 人 が 共同 して 、４ 落とし た 。	
参照文	５ 人 の 仲間 から ひとり 抜け て ４ 人 に なっ た 。	
入力文	I do not see any wrong with that .	1
出力文	に も わから ない 。	
参照文	私 は 別に 悪い と 思わ ない 。	
入力文	The cliff setn an echo back to us .	1
出力文	崖 が 送付 踏まえ の 後ろ を 送っ た 。	
参照文	崖 に当たって こだま が かえっ て き た 。	

6.3.2 自動評価

本研究では，自動評価に BLEU[3]，METEOR[4] ，RIBES[5], TER[6], STR[7] を使用する．

第7章　実験結果

実験は，翻訳システム 7 種類，実験データ 2 種類，翻訳の種類 2 種類，評価方法 2 種類の計 56 種類の実験を行う．出力文 1 文に対して評価を行った結果を表 7.1，表 7.4，表 7.7，表 7.10 に示す．また，提案手法の結果を表 7.2，表 7.5，表 7.8，表 7.11 に示す．出力文 1 文と提案手法の，人手評価に対する自動評価の相関係数を表 7.3，表 7.6，表 7.9 に示す．自動評価と人手評価の散布図を図 7.1 から図 7.8 に示す．

表 7.1　単文　日英翻訳　1 文出力

	BLEU	METEOR	RIBES	TER	STR	人手評価
RBMT	0.120	0.450	0.712	0.761	0.008	**3.34**
PSMT	0.124	0.438	0.691	0.743	0.009	2.01
HSMT	0.125	0.443	0.697	0.748	0.008	2.13
RBMT(a) +PSMT	**0.163**	**0.492**	0.731	**0.700**	**0.027**	2.81
RBMT(a) +HSMT	0.161	0.488	**0.734**	0.702	0.025	2.79
RBMT(t) +PSMT	0.151	0.477	0.730	0.713	0.016	3.03
RBMT(t) +HSMT	0.149	0.478	0.729	0.715	0.015	3.04

表 7.2　単文　日英翻訳　提案手法

	BLEU	METEOR	RIBES	TER	STR	人手評価
RBMT	0.048	0.207	0.353	0.446	0.008	**1.66**
PSMT	0.063	0.226	0.359	0.390	0.012	1.05
HSMT	0.063	0.229	0.362	0.392	0.011	1.11
RBMT(a) +PSMT	0.081	**0.253**	0.380	0.368	**0.033**	1.46
RBMT(a) +HSMT	**0.083**	0.253	**0.381**	**0.367**	0.030	1.46
RBMT(t) +PSMT	0.076	0.246	0.379	0.374	0.019	1.56
RBMT(t) +HSMT	0.077	0.248	0.379	0.373	0.017	1.57

表 7.3　単文　日英翻訳　人手評価に対する自動評価の相関係数

	BLEU	METEOR	RIBES	TER	STR
1 文出力	0.32	0.53	0.72	-0.21	0.24
提案手法	0.13	0.14	0.35	0.15	0.21

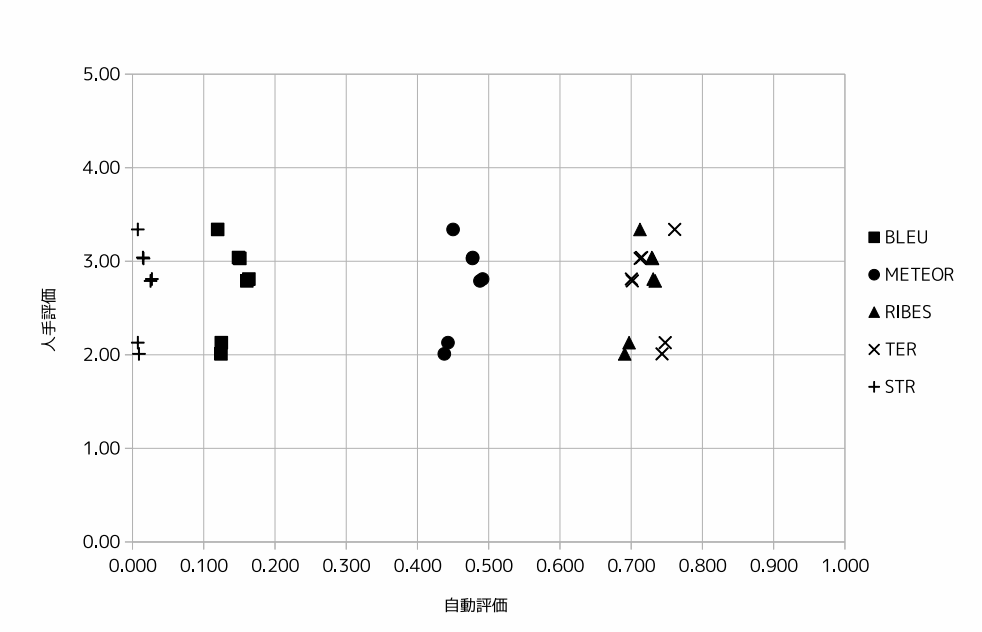

図 7.1 日英翻訳 単文 1 文出力 散布図

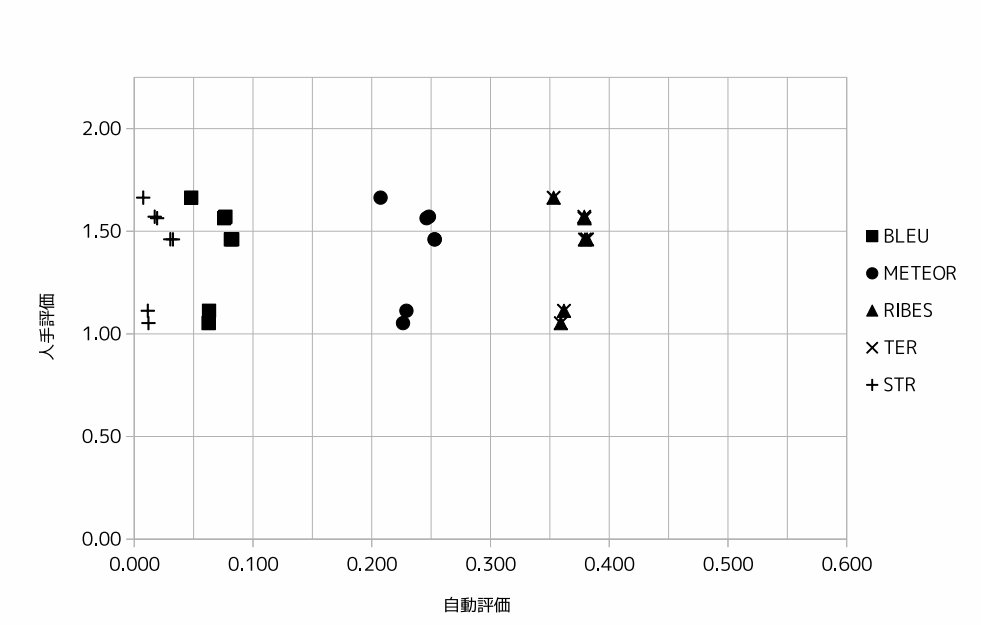

図 7.2 日英翻訳 単文 提案手法 散布図

表 7.4 単文　英日翻訳　1 文出力

	BLEU	METEOR	RIBES	TER	STR	人手評価
RBMT	0.119	0.369	0.737	0.906	0.003	**3.08**
PSMT	0.152	0.409	0.740	0.824	0.013	2.21
HSMT	0.159	0.416	0.751	0.816	0.017	2.97
RBMT(a) +PSMT	0.198	0.470	0.765	0.758	**0.026**	3.00
RBMT(a) +HSMT	**0.200**	**0.472**	**0.778**	**0.750**	0.025	3.00
RBMT(t) +PSMT	0.188	0.460	0.768	0.761	0.022	3.00
RBMT(t) +HSMT	0.195	0.465	0.773	0.757	0.022	2.92

表 7.5 単文　英日翻訳　提案手法

	BLEU	METEOR	RIBES	TER	STR	人手評価
RBMT	0.053	0.177	0.374	0.497	0.004	1.49
PSMT	0.076	0.211	0.384	0.433	0.017	1.14
HSMT	0.080	0.216	0.390	0.426	0.022	1.18
RBMT(a) +PSMT	0.101	0.244	0.398	0.397	**0.031**	1.50
RBMT(a) +HSMT	**0.104**	**0.245**	**0.406**	**0.391**	0.027	**1.56**
RBMT(t) +PSMT	0.095	0.237	0.398	0.399	0.026	1.53
RBMT(t) +HSMT	0.100	0.241	0.403	0.395	0.025	1.52

表 7.6 単文　英日翻訳　人手評価に対する自動評価の相関係数

	BLEU	METEOR	RIBES	TER	STR
1 文出力	0.30	0.33	0.48	-0.19	0.17
提案手法	0.41	0.38	0.48	-0.26	0.24

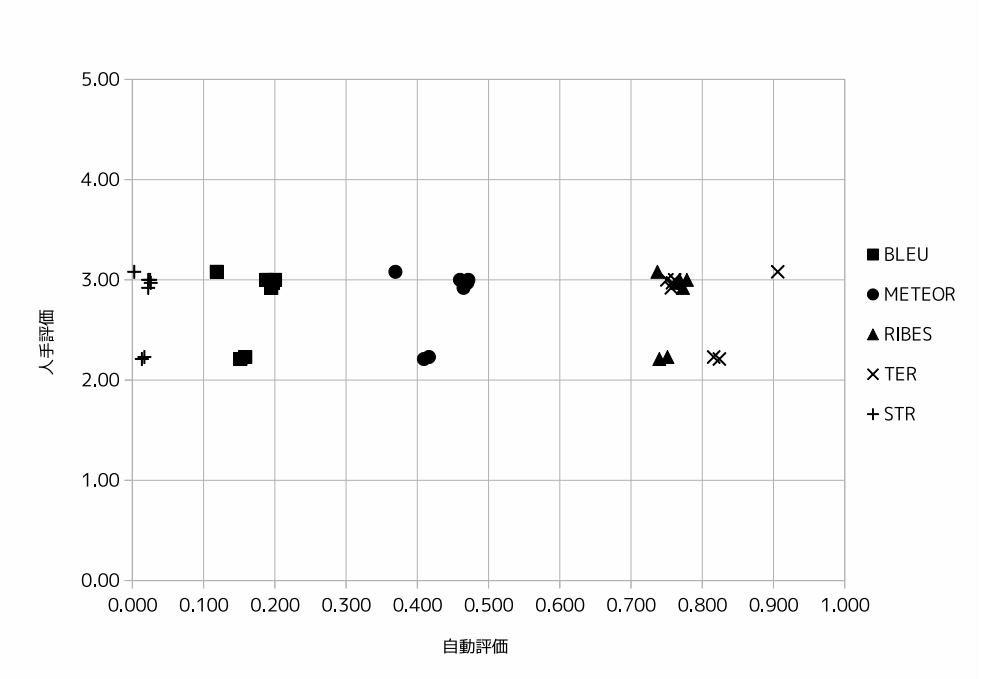

図 7.3 英日翻訳 単文 1 文出力 散布図

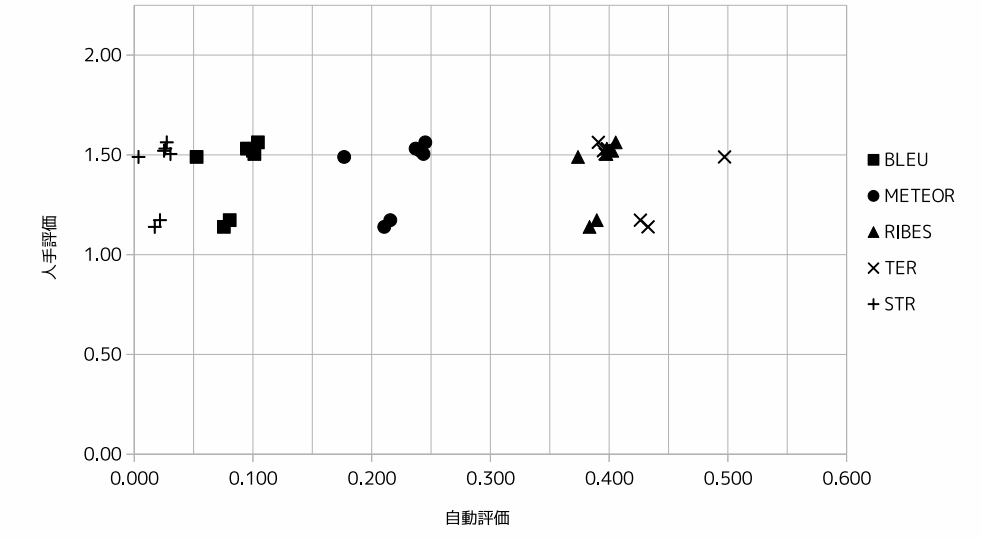

図 7.4 英日翻訳 単文 提案手法 散布図

表 7.7 重文複文　日英翻訳　1 文出力

	BLEU	METEOR	RIBES	TER	STR	人手評価
RBMT	0.091	0.393	0.665	0.857	0.002	**3.00**
PSMT	0.114	0.406	0.665	0.788	0.004	2.25
HSMT	0.117	0.419	0.671	0.776	0.004	2.32
RBMT(a) +PSMT	0.133	0.437	0.677	0.759	**0.009**	2.52
RBMT(a) +HSMT	0.135	0.437	0.684	0.759	**0.009**	2.47
RBMT(t) +PSMT	**0.140**	**0.446**	0.701	0.750	0.006	2.54
RBMT(t) +HSMT	0.137	0.441	**0.703**	**0.748**	0.006	2.54

表 7.8 重文複文　日英翻訳　提案手法

	BLEU	METEOR	RIBES	TER	STR	人手評価
RBMT	0.039	0.183	0.333	0.485	0.002	**1.52**
PSMT	0.058	0.210	0.345	0.413	0.005	1.17
HSMT	0.060	0.217	0.349	0.406	0.004	1.21
RBMT(a) +PSMT	0.067	0.226	0.351	0.398	**0.011**	1.31
RBMT(a) +HSMT	0.069	0.227	0.355	0.397	0.010	1.31
RBMT(t) +PSMT	**0.071**	**0.231**	0.364	0.397	0.008	1.32
RBMT(t) +HSMT	**0.071**	0.229	**0.366**	**0.391**	0.007	1.32

表 7.9 重文複文　日英翻訳　人手評価に対する自動評価の相関係数

	BLEU	METEOR	RIBES	TER	STR
1 文出力	-0.46	-0.31	-0.02	0.65	-0.34
提案手法	-0.49	-0.52	-0.32	0.70	-0.29

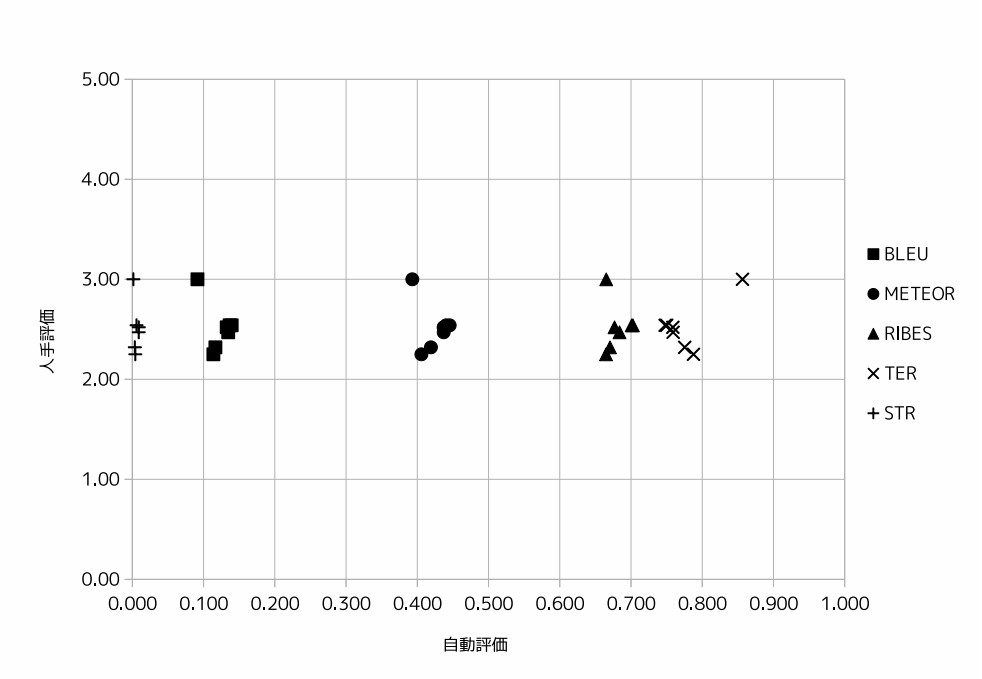

図 7.5 日英翻訳 重文複文 1 文出力 散布図

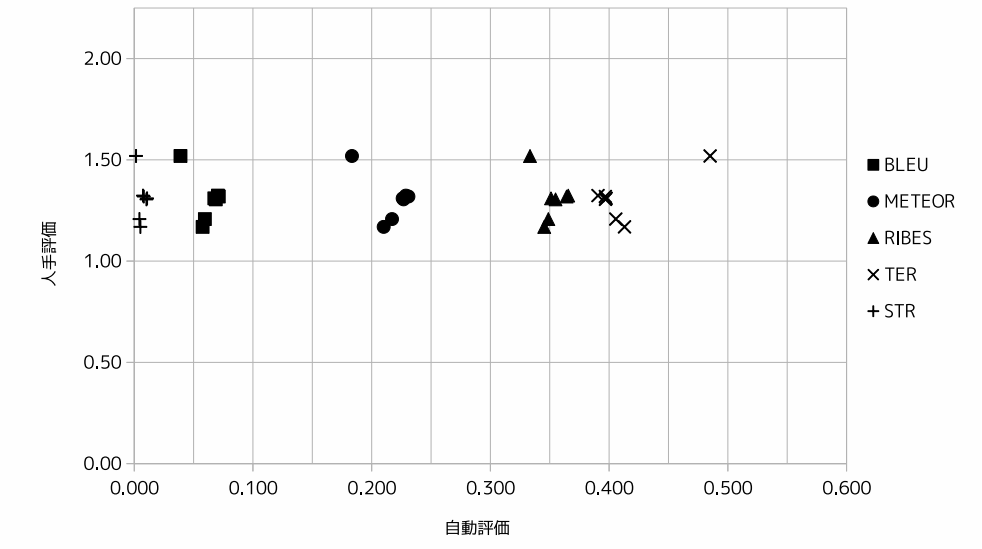

図 7.6 日英翻訳 重文複文 提案手法 散布図

表 7.10 重文複文　英日翻訳　1 文出力

	BLEU	METEOR	RIBES	TER	STR	人手評価
RBMT	0.113	0.372	0.700	0.898	0.001	**3.05**
PSMT	0.141	0.388	0.699	0.860	0.006	2.16
HSMT	0.142	0.397	0.707	0.841	0.005	2.24
RBMT(a) +PSMT	**0.190**	**0.456**	0.747	0.771	**0.009**	2.80
RBMT(a) +HSMT	0.188	0.452	**0.757**	**0.770**	0.005	2.73
RBMT(t) +PSMT	0.182	0.450	0.749	0.778	0.006	2.75
RBMT(t) +HSMT	0.179	0.449	0.750	0.779	0.005	2.82

表 7.11 重文複文　英日翻訳　提案手法

	BLEU	METEOR	RIBES	TER	STR	人手評価
RBMT	0.052	0.181	0.356	0.488	0.001	**1.53**
PSMT	0.071	0.200	0.362	0.450	0.007	1.13
HSMT	0.073	0.206	0.368	0.440	0.005	1.17
RBMT(a) +PSMT	**0.097**	**0.236**	0.389	**0.403**	**0.012**	1.46
RBMT(a) +HSMT	0.096	0.233	**0.393**	0.404	0.007	1.42
RBMT(t) +PSMT	0.094	0.234	0.390	0.406	0.008	1.42
RBMT(t) +HSMT	0.093	0.233	0.390	0.407	0.006	1.46

表 7.12 重文複文　英日翻訳　人手評価に対する自動評価の相関係数

	BLEU	METEOR	RIBES	TER	STR
1 文出力	0.18	0.31	0.44	-0.19	-0.22
提案手法	0.20	0.27	0.41	-0.16	-0.10

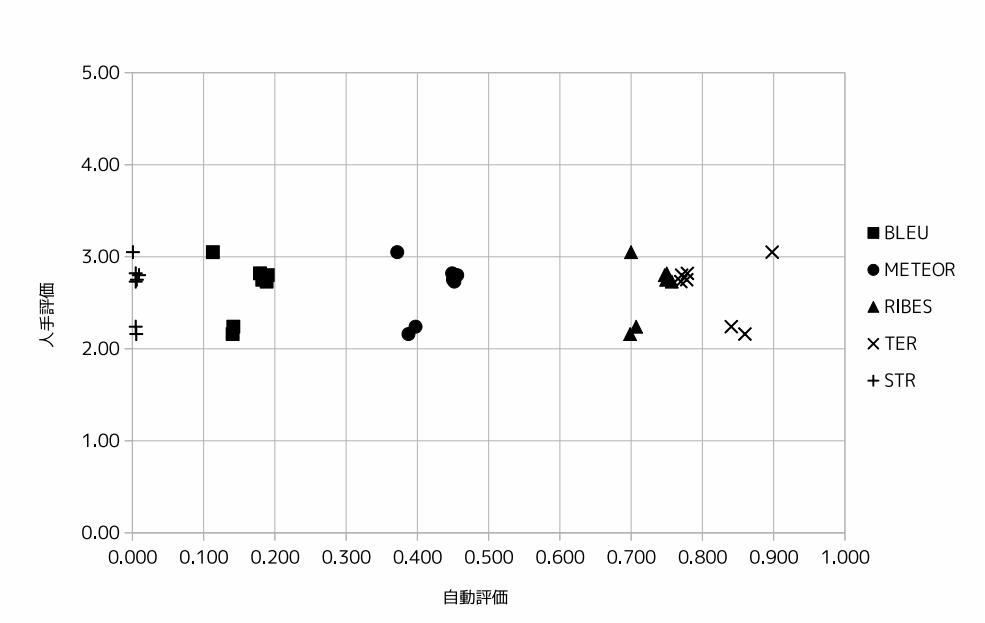

図 7.7 英日翻訳 重文複文 1 文出力 散布図

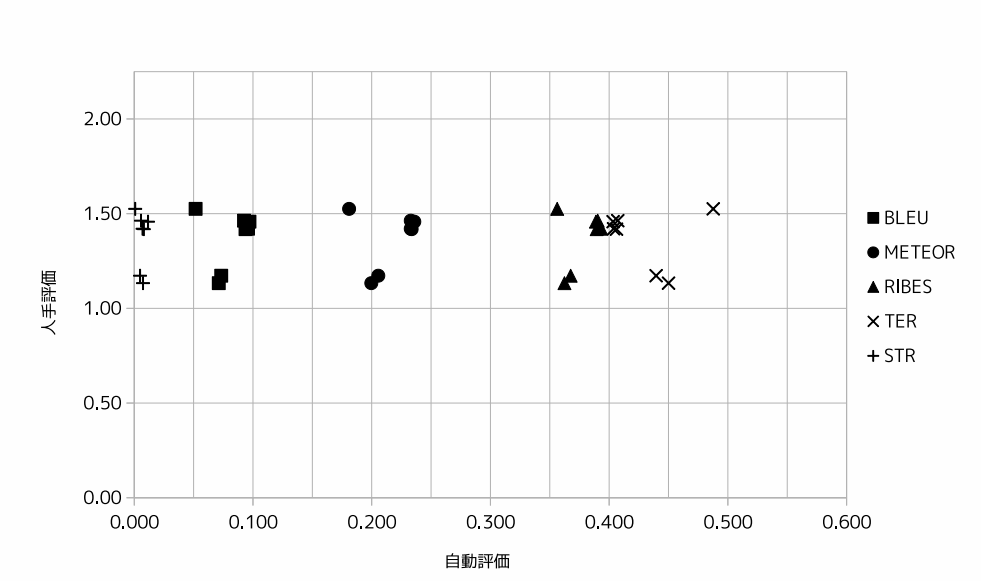

図 7.8 英日翻訳 重文複文 提案手法 散布図

表 7.6 より，単文の英日翻訳において，提案手法の結果は 1 文出力の結果と比較すると人手評価に対する自動評価の相関係数が向上した．しかし，他の実験では提案手法の結果と 1 文出力の結果を比較しても，人手評価に対する自動評価の相関係数に差はあまり見られなかった．

第8章　考察

8.1　提案手法と人手評価の差

提案手法の結果と1文出力の結果を比較しても，人手評価に対する自動評価の相関係数に差はあまり見られなかった．しかし，単文の英日翻訳においては，提案手法の結果は1文出力の結果と比較すると人手評価に対する自動評価の相関係数が向上した．この原因として，単文の英日翻訳の人手評価の結果において，1文出力ではRBMTが最も高い値を示しているが，提案手法ではRBMT(a)+HSMTが最も高い値を示している．そのため，単文の英日翻訳において提案手法は1文出力と比較すると，人手評価に対する自動評価の相関係数が高くなったと考えている．

しかし，重文複文の英日翻訳の実験では，提案手法の結果と1文出力の結果を比べても人手評価の相関係数にあまり差は見られない．単文と重文複文で違いが出たのは人手評価に原因があるとも考えられる．今後，評価者を増加させて，人手評価の信頼性を高める必要がある．

8.2　評価指標

本研究では，検索精度の評価に使用される評価指標の MRR に着目した．計算式が簡単だったため MRR を使用したが，他の評価指標を用いた方法も考えていきたい．

8.3　テスト文数

本研究では，ルールベース翻訳 (Tauras) において，10,000 文の入力文に対し，4 文以上出力した文をテスト文として使用した．ルールベース翻訳において，複数文を出力する文が少なかったため，テスト文の数も少なくなった．今後，テスト文を増やした調査の必要がある．

第9章　おわりに

本研究では，機械翻訳において，MRRを参考にして，複数の出力文を使用する自動評価方法について調査した．提案手法の結果と1文出力の結果を比較しても，人手評価に対する自動評価の相関係数に差はあまり見られなかった．しかし，単文の英日翻訳においては，提案手法の結果は1文出力の結果と比較すると人手評価に対する自動評価の相関係数が向上した．

今後は，人手評価の再調査を行っていきたい．また，ルールベース翻訳が複数文を出力する数が少なかったため，テスト文の数も少なくなった．テスト文の数を増やした調査も必要である．

謝辞

最後に,1 年間に渡ってご指導いただきました鳥取大学工学部知能情報工学科計算機 C 研究室の村田真樹教授, 村上仁一准教授, 徳久雅人講師そして計算機工学講座 C 研究室の方々に心から御礼申し上げます. また, 参考にさせていただいた論文の著者の方々に深く感謝致します.

参考文献

[1] Ondrej Bojar, Christian Buck, Chris Callison-Burch, Christian Federmann, Barry Haddow, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia, “Findings of the 2013 Workshop on Statistical Machine Translation ”,Proceedings of the Elighth Workshop on Statistical Machine Translation, pages 1-44, Sofia, Bulgaria, August 8-9,2013.

[2] E.M. Voorhees , “Proceedings of the 8th Text Retrieval Conference”, TREC-8 Question Answering Track Report,pp. 77-82, 1999.

[3] BLEU, “Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics”, 2002.

[4] Alon Lavie, Abhaya Agrwal, “METEOR: An Automatic Metric for MT Evaluation with High Level of Correlation with Human Judgments”, Proceedings of the ACL 2007 Workshop on Statistical Machine Translation, 2007.

[5] 平尾努, 磯崎秀樹, Kevin Duh, 須藤克仁, 塚田元, 永田昌明, “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 年次大会, D5-2, pp.1111-1114, 2011.

[6] Richard Schwartz, Linnea Micciulla, John Makhoul. “A Study of Translation Edit Rate with Targeted Human Annotation”, AMTA, 2006.

[7] 石原 雅文, ” 文一致数を用いた機械翻訳の自動評価”, 平成 24 年度卒業論文,2013.

[8] Shinya Amano, Hideki Hirakawa, Yoshinao Tsutsumi: “TAURAS: The Toshiba machine translation system”, Manuscr Program MT Summit, pp.15–23, 1987.

[9] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Alexandra Constantin, Evan Herbst,“Moses: Open Source Toolkit for Statistical

Machine Translation", Proceedings of the ACL 2007 Demo and Poster Sessions, pp.177-180, 2007.

[10] Andreas Stolcke: "SRILM - am Extensible Language Modeling Toolkit", 7th International Conference on Spoken Language Processing, pp.901–904, 2002.

[11] 英日・日英翻訳ソフト: "ATLAS"：http://software.fujitsu.com/jp/atlas/

[12] 村上仁一, 徳久雅人, "日英対訳データベースの作成のための 1 考察", 言語処理学会第 17 回年次大会, D4-5, pp.979-982, 2011.

[13] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: "Applying Conditional Random Fields to Japanese Morphological Analysis", Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230–237, 2004.

付 録A　人手評価

日英翻訳と英日翻訳の 2 種類，単文と重文複文の 2 種類，翻訳システム 7 種類の合計 28 種類に対し，出力文の中からランダムに抽出した 400 文ずつに人手評価を行った．左側の数字が人手評価の値である．