

概要

現在，文法構造が大きく異なる言語間での機械翻訳においてハイブリッド翻訳や階層型統計翻訳が注目されている．

ハイブリッド翻訳は，ルールベース翻訳と句に基づく統計翻訳を組み合わせる翻訳を行う．福田らは，日英翻訳において，句に基づく統計翻訳とハイブリッド翻訳を比較した．その結果，ハイブリッド翻訳が自動評価・人手評価ともに高いスコアを出力した [1]．

階層型統計翻訳は，文脈自由文法をモデルとして，統計的に翻訳を行う．久保田らは，日英翻訳において，句に基づく統計翻訳と階層型統計翻訳を比較した．その結果，階層型統計翻訳が自動評価・人手評価ともに高いスコアを出力した [2]．

そこで本研究では，ハイブリッド翻訳に用いられる句に基づく統計翻訳の代わりに階層型統計翻訳を用いて日英翻訳を行った．日英翻訳において，句に基づく統計翻訳よりも翻訳精度の高い階層型統計翻訳を用いることで，翻訳精度の向上を試みた．結果として，従来のハイブリッド翻訳と比較して，自動評価・人手評価ともに差がなかった．よって，提案手法は従来手法と差がなく，有効ではなかった．原因を調査したところ，前処理のルールベース翻訳において，多くの文が文法的に正しく翻訳されていた．ハイブリッド翻訳は前処理に大きく依存するため，後処理の統計翻訳の手法を変えても差が生じなかったのだと考えている．

また，ルールベース翻訳とハイブリッド翻訳を比較したところ，ルールベース翻訳が人手評価において，優れた翻訳を出力していた．今後は，ハイブリッド翻訳の精度向上のために，より正確な句の対応付けが可能な統計翻訳の方法を考えていく．

目次

第1章	はじめに	1
第2章	翻訳システム	3
2.1	ルールベース翻訳 (RMT)	3
2.2	統計翻訳	4
2.2.1	統計翻訳システムの概要	4
2.2.2	単語に基づく統計翻訳	7
2.2.3	句に基づく統計翻訳	8
2.2.4	階層型統計翻訳	9
2.3	ハイブリッド翻訳 (RMT+PSMT)	11
2.3.1	ハイブリッド翻訳の概要	11
2.3.2	ハイブリッド翻訳の手順	12
第3章	評価手法	14
3.1	自動評価	14
3.1.1	BLEU	14
3.1.2	METEOR	15
3.1.3	RIBES	15
3.2	人手評価	16
第4章	提案手法	17
4.1	提案手法の概要 (RMT+HSMT)	17
4.2	提案手法の手順	18
第5章	実験環境	19
5.1	ルールベース翻訳	19
5.2	翻訳モデルの学習	19

5.3	言語モデルの学習	19
5.4	デコーダ	19
5.5	ベースライン	19
5.6	評価	20
5.7	実験データ	20
5.7.1	単文コーパスを用いた実験	20
5.7.2	重文複文コーパスを用いた実験	21
第6章	実験結果	22
6.1	自動評価	22
6.2	人手評価	23
6.3	翻訳例	23
6.3.1	ベースライン (RMT+PSMT) の例	24
6.3.2	提案手法 (RMT+HSMT) の例	26
6.3.3	差なしの例	28
6.3.4	同一出力の例	30
第7章	考察	32
7.1	ベースラインと提案手法に差のない原因	32
第8章	各手法の比較	35
8.1	自動評価	35
8.2	人手評価	36
8.2.1	RMT+PSMT 対 RMT	36
8.2.2	RMT+HSMT 対 RMT	36
8.2.3	HSMT 対 RMT+HSMT	37
第9章	おわりに	38
付録A	各手法の対比較評価例	42
A.1	RMT+PSMT 対 RMT (翻訳例)	42
A.1.1	RMT の例	42
A.1.2	RMT+PSMT の例	44
A.1.3	差なしの例	45

A.1.4	同一出力の例	47
A.2	RMT+HSMT 対 RMT (翻訳例)	48
A.2.1	RMT の例	48
A.2.2	RMT+HSMT の例	50
A.2.3	差なしの例	51
A.2.4	同一出力の例	53
A.3	HSMT 対 RMT+HSMT (翻訳例)	54
A.3.1	HSMT の例	54
A.3.2	RMT+HSMT の例	56
A.3.3	差なしの例	58
A.3.4	同一出力の例	60

目 次

2.1	日英統計翻訳システムの枠組	4
2.2	デコーダ	6
2.3	階層型統計翻訳による翻訳の例	10
2.4	ハイブリッド翻訳システムの枠組	12
4.1	提案手法の枠組	17

表 目 次

2.1	ルールベース翻訳の出力例	3
2.2	N -gram モデルの例	5
2.3	日英フレーズテーブルの例	8
2.4	句に基づく統計翻訳の出力例	9
2.5	日英ルールテーブルの例	10
2.6	階層型統計翻訳の出力例	11
2.7	英' 英フレーズテーブルの例	13
2.8	ハイブリッド翻訳の出力例	13
3.1	対比較例評価出力例 1	16
3.2	対比較評価基準例 2	16
4.1	英' 英ルールテーブルの例	18
5.1	対比較評価基準	20
5.2	単文コーパス	20
5.3	単文実験データの例	21
5.4	重文複文コーパス	21
5.5	重文複文実験データの例	21
6.1	単文における自動評価	22
6.2	重文複文における自動評価	22
6.3	単文における対比較評価	23
6.4	重文複文における対比較評価	23
6.5	単文における RMT+PSMT の例	24
6.6	重文複文における RMT+PSMT の例	25
6.7	単文における RMT+HSMT の例	26
6.8	重文複文における RMT+HSMT の例	27

6.9	単文における差なしの例	28
6.10	重文複文における差なしの例	29
6.11	単文における同一出力の例	30
6.12	重文複文における同一出力の例	31
7.1	判断条件	32
7.2	判断条件で調査した結果	32
7.3	単文実験における RMT 文法 ×	33
7.4	重文複文実験における RMT 文法 ×	33
8.1	単文における自動評価比較	35
8.2	重文複文における自動評価比較	35
8.3	単文における対比較評価 (RMT+PSMT 対 RMT)	36
8.4	重文複文における対比較評価 (RMT+PSMT 対 RMT)	36
8.5	単文における対比較評価 (RMT+HSMT 対 RMT)	36
8.6	重文複文における対比較評価 (RMT+HSMT 対 RMT)	37
8.7	単文における対比較評価 (HSMT 対 RMT+HSMT)	37
8.8	重文複文における対比較評価 (HSMT 対 RMT+HSMT)	37
A.1	単文における RMT の例 (RMT+PSMT 対 RMT)	42
A.2	重文複文における RMT の例 (RMT+PSMT 対 RMT)	43
A.3	単文における RMT+PSMT の例 (RMT+PSMT 対 RMT)	44
A.4	重文複文における RMT+PSMT の例 (RMT+PSMT 対 RMT)	44
A.5	単文における差なしの例 (RMT+PSMT 対 RMT)	45
A.6	重文複文における差なしの例 (RMT+PSMT 対 RMT)	46
A.7	単文における同一出力の例 (RMT+PSMT 対 RMT)	47
A.8	重文複文における同一出力の例 (RMT+PSMT 対 RMT)	47
A.9	単文における RMT の例 (RMT+HSMT 対 RMT)	48
A.10	重文複文における RMT の例 (RMT+HSMT 対 RMT)	49
A.11	単文における RMT+HSMT の例 (RMT+HSMT 対 RMT)	50
A.12	重文複文における RMT+HSMT の例 (RMT+HSMT 対 RMT)	50
A.13	単文における差なしの例 (RMT+HSMT 対 RMT)	51
A.14	単文における差なしの例 (RMT+HSMT 対 RMT)	52

A.15 単文における同一出力の例 (RMT+HSMT 対 RMT)	53
A.16 単文における同一出力の例 (RMT+HSMT 対 RMT)	53
A.17 単文における HSMT の例 (HSMT 対 RMT+HSMT)	54
A.18 重文複文における HSMT の例 (HSMT 対 RMT+HSMT)	55
A.19 単文における RMT+HSMT の例 (HSMT 対 RMT+HSMT)	56
A.20 重文複文における RMT+HSMT の例 (HSMT 対 RMT+HSMT)	57
A.21 単文における差なしの例 (HSMT 対 RMT+HSMT)	58
A.22 重文複文における差なしの例 (HSMT 対 RMT+HSMT)	59
A.23 単文における同一出力の例 (HSMT 対 RMT+HSMT)	60
A.24 重文複文における同一出力の例 (HSMT 対 RMT+HSMT)	60

第1章 はじめに

従来の機械翻訳は、人手で作成した文法規則の定義にしたがって翻訳を行う“ルールベース翻訳”であった。しかし、ルールベース翻訳システムの構築には、言語の専門家の知識を必要とし、膨大な開発コストがかかる。そのため、近年は、言語対から自動的に翻訳モデルを学習する“句に基づく統計翻訳”が主流となっている。計算機の性能向上により、短時間での処理が可能になったため、様々な言語間での統計翻訳が研究されている。その一つとして、言語間の文法構造が類似しているイタリア-英語間において、句に基づく統計翻訳は翻訳精度が高いことが報告されている [3]。しかしながら、言語間の文法構造が大きく異なる日本語-英語間の翻訳においては、翻訳精度が低い。

そこで、現在、ハイブリッド翻訳が注目されている [4]。ハイブリッド翻訳は、まず入力文に対し、ルールベース翻訳を用いて目的言語の文法構造に近い文を出力する。次に、その出力文に対し、句に基づく統計翻訳で目的言語へと翻訳する。福田らは、ハイブリッド翻訳による日英翻訳は、句に基づく統計翻訳よりも翻訳精度が高いことを報告した [1]。

一方、Chiang が階層型統計翻訳を提案した [5]。階層型統計翻訳は文脈自由文法をモデルとして翻訳を行うため、正しい文法構造を持った文に翻訳される傾向がある。久保田らは日英翻訳において、階層型統計翻訳と句に基づく統計翻訳を比較した。その結果、自動評価、人手評価ともに階層型統計翻訳の翻訳精度が高いことを報告した [2]。

そこで、本研究では、ルールベース翻訳と階層型統計翻訳を組み合わせる日英翻訳を行い、翻訳精度の向上を試みた。その結果、提案手法は従来のハイブリッド翻訳と翻訳精度の差がなかった。原因として、ルールベース翻訳の文法構造変換が優れているため、階層型統計翻訳が文法構造を考慮する必要がなかったことが考えられる。また、提案手法や従来のハイブリッド翻訳の翻訳精度は、ルールベース翻訳単体よりも劣っていた。よって、より翻訳精度の高いハイブリッド手法の検討が必要である。

本論文の構成は以下の通りである。第2章において、各翻訳システムについての説明を行う。第3章において、評価手法についての説明を行う。第4章において、提案手法の手順についての説明を行う。第5章において、実験環境についての説明を行う。第6章において、自動評価と人手評価の結果を示す。第7章において、本研究の考察を述べる。

第 8 章において，実験で使⽤した各手法の比較を⾏う．第 9 章において，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 (RMT)

ルールベース翻訳は、人手によって構築された変換規則に従って翻訳を行うシステムである。本論文においてはルールベース翻訳を RMT と表記する。長所としては、翻訳する際に、適合する規則が存在する場合に翻訳精度が高いことが挙げられる。一方、適合する規則が存在しない翻訳においては、翻訳精度が低いという短所がある。また、人手によって規則を構築するため、開発に膨大なコストがかかる。

一般的なルールベース翻訳の手順を以下に示す。

手順 1 与えられた入力文に対し、辞書の品詞などを用いて構文解析を行う。

手順 2 構文解析を行った文の語順を、目的言語の語順に変換する。

手順 2 再度辞書を参照し、助詞や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。

ルールベース翻訳の出力例を表 2.1 に示す。なお、参照文とは、入力文の日本語に対する英語の正解文である。

表 2.1 ルールベース翻訳の出力例

例 1	
入力文	コンパスは常に北を指す。
RMT	Compass always points out north .
参照文	The compass always points to the north .
例 2	
入力文	彼女はフランスへ行った。
RMT	She went to France .
参照文	She went over to France .

例 3	
入力文	彼 が テニス クラブ に入る。
RMT	He goes into a tennis club .
参照文	He will join a tennis club .

2.2 統計翻訳

2.2.1 統計翻訳システムの概要

統計翻訳とは、翻訳する言語と目的言語の対訳文を大量に収集した対訳データを用いて、自動的に翻訳モデルを作成し翻訳を行う、機械翻訳手法の一つである。現在、統計翻訳システムとして、「単語に基づく統計翻訳」、「句に基づく統計翻訳」、「階層型統計翻訳」が提案されている。

図 2.1 に日英統計翻訳システムの枠組を示す。

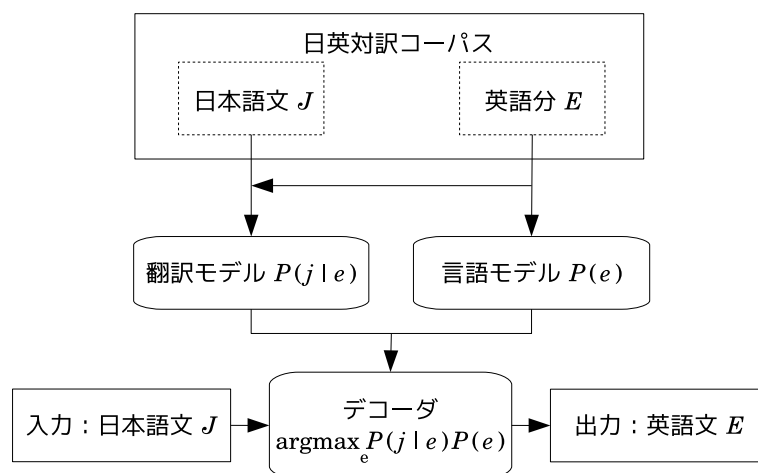


図 2.1 日英統計翻訳システムの枠組

日英統計翻訳は、まず日英対訳コーパスの学習によって、翻訳モデル $P(j|e)$ と言語モデル $P(e)$ を作成する。そして、入力文として日本語文 J が与えられたとき、 $P(j|e)$ と $P(e)$ の組み合わせから、確率が最大となる英語文 E を探索し翻訳を行う。統計翻訳の基本式を以下に示す。

$$E = \operatorname{argmax}_e P(e|j) \quad (2.1)$$

$$\approx \operatorname{argmax}_e P(j|e) P(e) \quad (2.2)$$

2.2.1.1 翻訳モデル

翻訳モデルとは、翻訳する言語の単語列から目的言語の単語列へ、確率的に翻訳を行うためのモデルである。翻訳モデルは各統計翻訳手法によって異なるため、詳細については各手法の項で説明する。

2.2.1.2 言語モデル

言語モデルは、単語列に対して、それらの単語列の生成確率を付与するモデルである。言語モデルを用いることで、目的言語の文として適切な単語列を選出する。統計翻訳では一般的に N -gram モデルを用いる。 N -gram モデルの例として 2-gram モデルの例を以下の表 2.2 に示す。

表 2.2 N -gram モデルの例

-0.920822	This is	-0.568185
-2.123066	This picture	-0.275901
-2.381344	This plan	-0.131124
-2.682374	This music	-0.083978
-3.638501	a pen	-0.100087
-3.296078	a pencil	-0.132691

一番上の行に関して説明すると以下となる。

-0.920822: “This” のあとに “is” がくる確率を常用対数で表した値
“ $\log_{10}(P(a|factory))$ ”
This is: 2-gram で表された単語列
-0.5681853: バックオフスムージングにより得られる，“This” のあとに “is”
がくる確率を常用対数で表した値 “ $\log_{10}(P(a|factory))$ ”

また、バックオフスムージングとは、高次の N -gram が存在しない場合、低次の N -gram を用いる手法である。この低次の確率を改良したスムージングの手法が Kneser-Ney スムージングである。言語モデルにおける N -gram 作成には、性能の良さから一般的に Kneser-Ney スムージングが用いられる。

2.2.1.3 N-gram モデル

N-gram モデルは，“単語列 w_1, w_2, \dots, w_n の i 番目の単語 w_i の生起確率 $P(w_i)$ は，直前の単語 ($N - 1$) に依存する” という仮定に基づくモデルである．計算式を以下に示す．

$$P(w_1 w_2 \dots w_n) = \prod_{i=1}^n P(w_i | w_{i-1}) \quad (2.3)$$

例として，“This is a pen” という文字列に対する 2-gram モデルを以下に示す．

$$P(e = \text{“This is a pen”}) \approx P(\text{This}) \times P(\text{is}|\text{This}) \times P(\text{a}|\text{is}) \times P(\text{pen}|\text{a})$$

また，3-gram モデルのときは $P(\text{a}|\text{This is})$ となる．このように，($N - 1$) 単語の次にくる単語が “is” や “pen” である確率を求める．

2.2.1.4 デコーダ

デコーダは翻訳モデル $P(j|e)$ と言語モデル $P(e)$ を用いて，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．

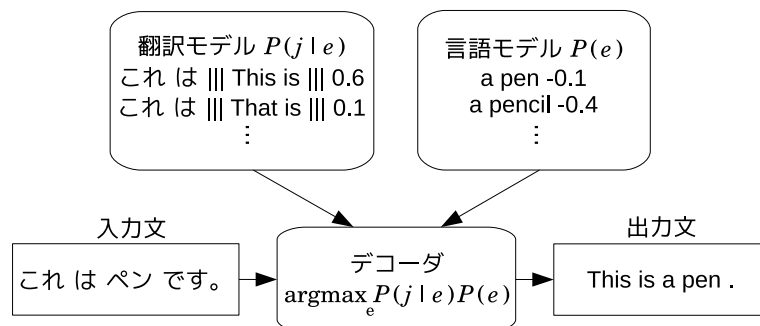


図 2.2 デコーダ

日英翻訳において， $\text{arg max}_e P(j|e)P(e)$ となる英語文を探索して翻訳を行う．しかし，この探索には膨大な計算量が必要となるため，ビームサーチ法を用いて候補をしぼることで計算量を減らす．ビームサーチ法は，翻訳確率の低い翻訳候補を無視することで，探索の範囲を限定する手法である．

2.2.1.5 パラメータチューニング

パラメータチューニングとは、デコーダで用いるパラメータの最適化を行うことである。パラメータチューニングには、MERT(Minimum Error Rate Training)[6] が用いられる。MERTは、自動評価 (BLEU) を最大にする翻訳結果が選出されるようにパラメータを調整する。その際、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 文程度の翻訳候補を出力し、最適な翻訳候補が上位にくるように重みを付与する。

2.2.2 単語に基づく統計翻訳

2.2.2.1 単語に基づく統計翻訳の概要

1990 年代前半に「単語に基づく統計翻訳」が提案された。単語に基づく統計翻訳は、単語の対応を翻訳モデルに用いる。そのため、いくつか問題点があった。一つとして挙げられるのが、翻訳モデルを作成する際に、翻訳する文の単語と目的文の単語の対応がない場合である。単語に基づく統計翻訳では、対応がない単語には NULL を対応させていた。よって、片方の単語のみで、単語の発生確率を計算するため、信頼性が低いという問題がある。

2.2.2.2 IBM 翻訳モデル

単語に基づく統計翻訳には、IBM 社の Brown ら [7] によって提案された IBM 翻訳モデルが用いられる。IBM 翻訳モデルは、Model1 から Model5 までの 5 つから構成され、順に複雑な計算を行う。仏英翻訳で使用することを想定しているため、本節では、仏英翻訳を前提に説明を行う。IBM 翻訳モデルでは、フランス語文 F と英語文 E の翻訳モデル $P(f|e)$ を計算するために、アライメント a を用いる。IBM 翻訳モデルの基本式を以下に示す。

$$P(f|e) = \sum_a P(f, a|e)$$

アライメント a は、仏単語 F と英単語 E の対応関係を意味する。IBM 翻訳モデルでは、英単語は 1:n の対応を持ち、仏単語は 1:1 の対応を持つと仮定する。また、仏単語に適切な対応関係を持つ英単語が存在しないときには、その仏単語と英語文の先頭の特殊文字 e_0 を対応させる。

2.2.2.3 GIZA++

GIZA++[8]とは、統計翻訳で用いることを前提として作られた、原言語と目的言語の対訳文において最尤な単語の対応を得るツールである。IBM 翻訳モデルを学習して、単語の対応関係の確率値を計算し、翻訳モデルの要素として付与する。

2.2.3 句に基づく統計翻訳

2.2.3.1 句に基づく統計翻訳の概要

2000年の初めに「句に基づく統計翻訳」が提案された。句に基づく統計翻訳は、句の対応を翻訳モデルに用いる。句を構成する単語の数が、翻訳する文の句と目的文の句で一致する必要がないため、単語に基づく統計翻訳の単語対応の問題を解決した。また、並べ替えにおいても、単語に基づく統計翻訳よりも優れている。そのため、近年では句に基づく統計翻訳が主流となっている。なお、本論文では、句に基づく統計翻訳を「PSMT」と表記する。

2.2.3.2 翻訳モデル

句に基づく統計翻訳の翻訳モデルはフレーズテーブルによって管理されている。フレーズテーブルの例を表 2.3 に示す。

表 2.3 日英フレーズテーブルの例

これは	This is a	0.525424 0.0311118 0.118774 0.000752339
これは	This is an	0.666667 0.0179866 0.0153257 0.000128339
これは	This is	0.127148 0.0102929 0.141762 0.0110294
これは私の	This is my	0.333333 0.000404608 0.5 0.000884628
これは私の	of this	0.00877193 4.32631e-06 0.5 0.0256013

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である。

句に基づく統計翻訳 (PSMT) の出力例を表 2.4 に示す。

表 2.4 句に基づく統計翻訳の出力例

例 1	
入力文	水が腐っている。
PSMT	The water is rotten .
参照文	The water is foul .
例 2	
入力文	彼の考え方は極端すぎる。
PSMT	His way of thinking is too the extreme .
参照文	His way of thinking goes too far .
例 3	
入力文	私は率直に話した。
PSMT	I talked with candor .
参照文	I spoke plainly .

2.2.4 階層型統計翻訳

2.2.4.1 階層型統計翻訳の概要

2005年に「階層型統計翻訳」が提案された。階層型統計翻訳は、句に基づく統計翻訳を拡張した手法である。階層型統計翻訳は、階層句の対応で、木構造を用いて翻訳を行うため、文法構造を考慮する。階層句とは、非終端記号を含む句である。そのため、文法構造の大きく異なる日本語-英語間での翻訳において、句に基づく統計翻訳よりも翻訳精度が高い。

階層型統計翻訳による翻訳の例を図 2.3 に示す。手順として、まず、学習文から階層句の対応を学習する。そして、入力文が与えられたとき、句の中に句が含まれる階層に分解し、それぞれの階層に階層句を当てはめて空白を埋めていくことで目的言語の出力文を生成する。なお、本論文では、階層型統計翻訳を「HSMT」と表記する。

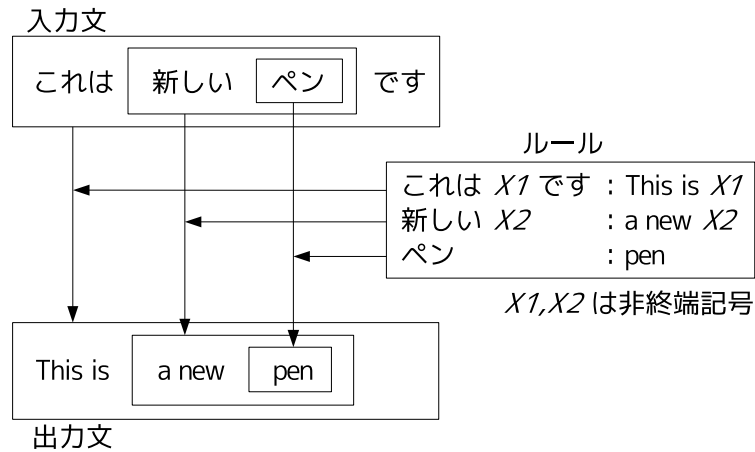


図 2.3 階層型統計翻訳による翻訳の例

2.2.4.2 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 2.5 に示す．

表 2.5 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．

階層型統計翻訳 (HSMT) の出力例を表 2.8 に示す．

表 2.6 階層型統計翻訳の出力例

例 1	
入力文	その計画は成功の見込みが十分ある。
HSMT	The plan has a chance of success .
参照文	The plan bids fair to succeed .
例 2	
入力文	組合はストライキに参加する。
HSMT	The union decided to go on strike .
参照文	The union will join in the strike .
例 2	
入力文	彼は健康を損なった。
HSMT	He ruined his health .
参照文	He ruined his health .

2.3 ハイブリッド翻訳 (RMT+PSMT)

2.3.1 ハイブリッド翻訳の概要

本論文におけるハイブリッド翻訳は、ルールベース翻訳を前処理に用いた句に基づく統計翻訳と定義する。また、RMT+PSMT と表記する。ハイブリッド翻訳は、PSMT の前処理に RMT を用いることで、文法構造を目的言語に類似させることを目的とした手法である。文法構造の大きく異なる日本語-英語間での翻訳において、句に基づく統計翻訳よりも翻訳精度が高い。

ハイブリッド翻訳の枠組を図 2.4 に示す。

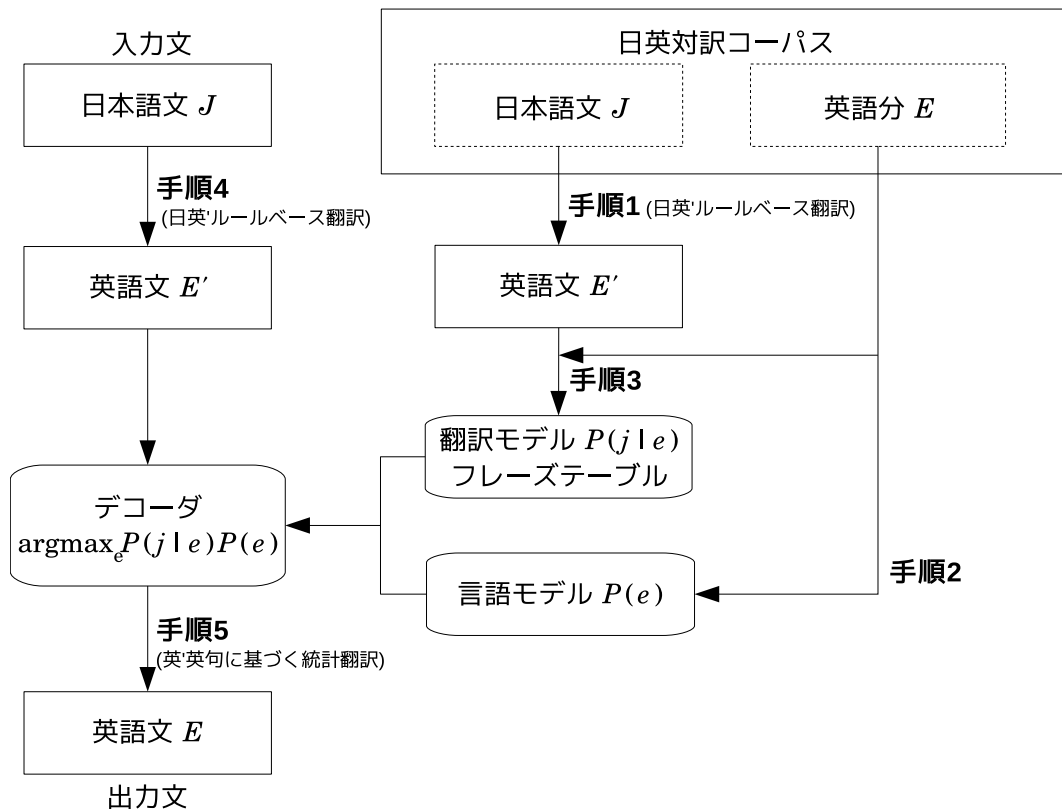


図 2.4 ハイブリッド翻訳システムの枠組

2.3.2 ハイブリッド翻訳の手順

ハイブリッド翻訳の手順を以下に示す．手順 1～3 で統計モデルを学習し，手順 4，5 でテストデータの翻訳を行う．

手順 1 前処理として，日英対訳コーパスの日本語文 J に対して，RMT で日英' 翻訳を行う．

手順 2 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．

手順 3 手順 1 で得られた英語文 E' と日英対訳コーパスの英語文を用いて，英' 英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の表 2.7 に示す．

手順 4 前処理として，テストデータ (入力文) の日本語文 J に対して，RMT で日英' 翻訳を行う．

表 2.7 英' 英フレーズテーブルの例

This is This is	0.443709	0.364442	0.650485	0.267549
This is This was	0.030303	0.00917643	0.00323625	0.00396328
This is not This is not	0.1875	0.246718	0.428571	0.0724435
This is not It is not	0.0172414	0.0038012	0.142857	0.000788545
This is my This is my	0.333333	0.08223	0.5	0.154497
This is my These are my	1	0.0080121	0.5	0.000729796

手順 5 手順 2 で作成した言語モデルと手順 3 で作成した翻訳モデルを用いて，手順 4 で得られた英語文 E' に対して PSMT で英' 英翻訳を行う．得られた英語文 E を最終出力とする．

ハイブリッド翻訳 (RMT+PSMT) の出力例を表 2.8 に示す．また，前処理 (RMT) の出力も示す．

表 2.8 ハイブリッド翻訳の出力例

例 1	
入力文	これらの計画には想像力が欠如している。
RMT+PSMT	These plans lack imagination .
RMT	These plans lack imaginative power .
参照文	These plans lack imagination .
例 2	
入力文	夢の世界に住んでいる。
RMT+PSMT	He lives in the world of a dream .
RMT	It lives in the world of the dream .
参照文	He lives in a dream world .
例 2	
入力文	豆腐は大豆より作る。
RMT+PSMT	Tofu is made from soy stringbeans .
RMT	Tofu is made from a soybean .
参照文	Tofu is made from soy beans .

第3章 評価手法

3.1 自動評価

自動評価とは、機械翻訳システムの翻訳精度を自動的に評価する方法である。一般的に、翻訳システムが出力した文と、あらかじめ用意した正解文との比較を行い評価する。現在、その比較の方法として様々な手法が研究されている。本研究では、自動評価法として BLEU[9]、METEOR[10]、RIBES[11] を用いる。本節では、それぞれの自動評価法の特徴を説明する。

3.1.1 BLEU

BLEU は、機械翻訳システムの自動評価において、現在主流となっている評価法である。N-gram 適合率を用いて、語順の正確さで評価を行う。実験では 4-gram を用いる。0 から 1 のスコアで評価し、語順が正解文に近いとき高いスコアを出力する。BLEU の計算式を以下に示す。

$$BLEU = BP \times \exp \left\{ \frac{1}{N} \sum_{n=1}^N \log_e P_n \right\} \quad (3.1)$$

$$P_n = \frac{\text{出力文と参照文で一致した } n\text{-gram 数}}{\text{出力文の } n\text{-gram 数}} \quad (3.2)$$

$$BP = \min \left(1, \exp \left(1 - \frac{\text{参照文の語数}}{\text{出力文の語数}} \right) \right) \quad (3.3)$$

BLEU は 1 ~ N-gram の適合率の幾何平均に BP をかけた式で表される。BP(Brevity Penalty) は、短い翻訳文が高いスコアになることを防ぐためのペナルティである。

3.1.2 METEOR

METEOR は、適合率と再現率の両方を考慮した評価方法である。実験では *uni-gram* を用いる。0 から 1 のスコアで評価し、1 に近いほど機械翻訳の評価が高いことを示す。METEOR の計算式を以下に示す。

$$METEOR = F \times (1 - Pen) \quad (3.4)$$

$$F \text{ 値} = \frac{P \times R}{\alpha \times P + (1 - \alpha) \times R} \quad (3.5)$$

$$Pen = \gamma \times \left(\frac{c}{m}\right)^\beta \quad (3.6)$$

METEOR は、適合率 P と再現率 R の調和平均である F 値と、単語の非連続性に対するペナルティ関数 Pen を用いて計算される。(5.6) 式において、 m は参照文と出力文の単語の一致率であり、 c は、一致した単語の列が参照文の単語の列と一致する数を表す。上式 (5.5) において、 α, β, γ の値はパラメータである。本研究では、 $\alpha = 0.8, \beta = 2.5, \gamma = 0.4$ とする。

また、一致率を計算する際、WordNet の類義語を参照するため、似た意味を持つ単語は同一であると判断される。さらに、単語の語形変化を無視して処理することが可能である。そのため、METEOR は単語属性で評価が可能である。なお、WordNet[12] は英語の概念辞書であり、英単語の品詞や単語の持つ意味概念の定義を管理するデータベースである。

3.1.3 RIBES

RIBES 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。0 から 1 のスコアで評価し、出現順序が正しいとき高いスコアを出力する。RIBES の計算式を以下に示す。

$$RIBES = \frac{\tau + 1}{2} \times P^\alpha \quad (0 \leq \alpha \leq 1) \quad (3.7)$$

$$\tau = \frac{\sum_{i=1}^{n-1} K_i - \sum_{i=1}^{n-1} L_i}{\frac{n(n-1)}{2}} \quad (3.8)$$

ここで、 τ はケンドールの順位相関係数である。まず、参照文と出力文の共通単語を抽出する。その共通単語の中で、参照文の単語の出現順に 1 から番号を与えたりリスト r を作成する。そして、出力文の中で、 r の単語に対応する単語が何番目に出てきたかを管理

するリスト h を作成する。 $j = i + 1, 2, \dots, n$ として、 h_i について、 $h_i < h_j$ となる場合の数を K_i 、 $h_i > h_j$ となる場合の数を L_i とする。 K_i や L_i を (5.8) 式に代入し τ を計算する。

また、 P は *uni-gram* 適合率であり、 α はペナルティに対する重みとして使用される。 RIBES は、単語の出現順序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。

3.2 人手評価

人手評価として、対比較評価がある。対比較評価は、二つの文を相対的に比較して、どちらがより正しい文であるかを人手で選択する評価方法である。二つの翻訳システムの出力で優劣を判断する場合に有効である。

対比較を行う例を表 3.1 に示す。日本語の入力文に対し手法 A と手法 B で日英翻訳を行う場合を考える。手法 A と手法 B を比較すると、正しい翻訳をしている文は手法 B である。この場合、評価者によって、表 3.2 の選択肢から「手法 B」が選択される。

人手評価は評価者によって評価に差が出てくる可能性がある。しかし、正解文を用意しておき、参照して対比較評価を行うことで、信頼性を高める。

表 3.1 対比較例評価出力例 1

例 1	
入力文	これは卒業論文です。
正解文	This is a graduation thesis.
手法 A	This is a student yearbook.
手法 B	This is a thesis of graduation.

表 3.2 対比較評価基準例 2

手法 A	手法 A の方が良い
手法 B	手法 B の方が良い
差なし	手法 A と手法 B の出力文の単語対応や伝わる意味に差がない
同一出力	手法 A と手法 B の出力文が完全一致

第4章 提案手法

4.1 提案手法の概要 (RMT+HSMT)

本論文における提案手法は、ルールベース翻訳を前処理に用いた階層型統計翻訳と定義する。また、RMT+HSMT と表記する。提案手法は、ハイブリッド翻訳における PSMT の代わりに、HSMT を用いる。PSMT よりも翻訳精度の高い HSMT を用いることで、翻訳精度の向上が期待できる。

提案手法の枠組を図 4.1 に示す。

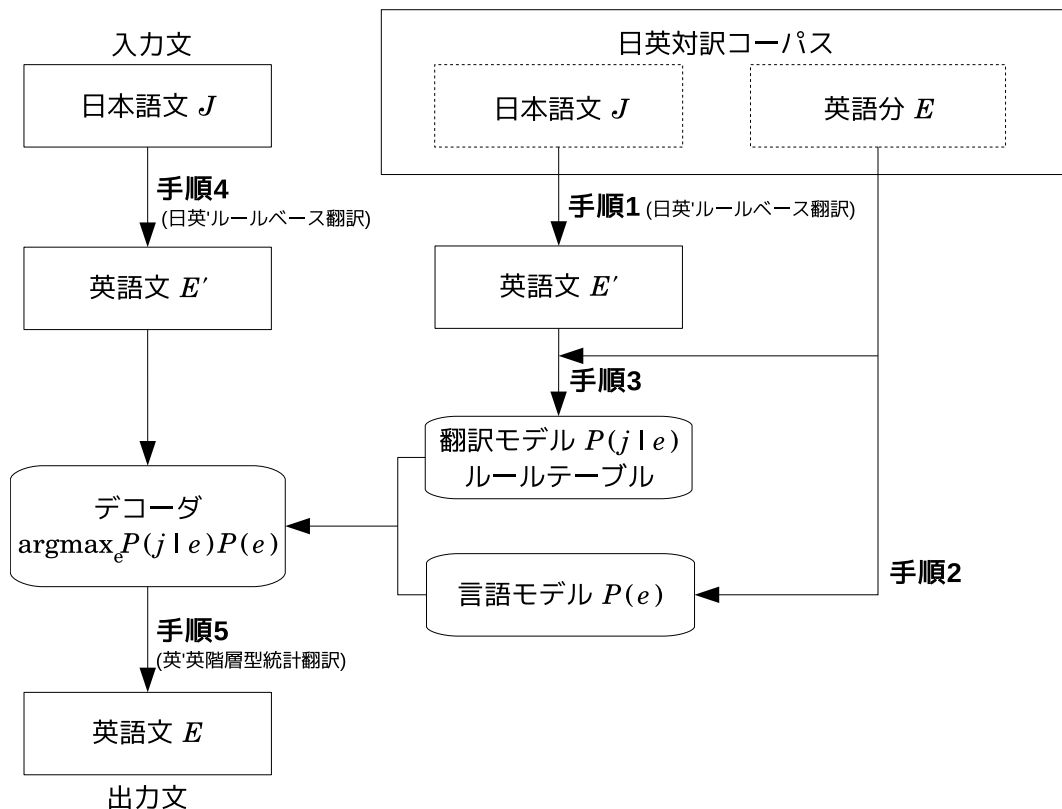


図 4.1 提案手法の枠組

4.2 提案手法の手順

提案手法の手順を以下に示す．手順 1～3 で統計モデルを学習し，手順 4，5 でテストデータの翻訳を行う．

手順 1 前処理として，日英対訳コーパスの日本語文 J に対して，RMT でそれぞれ日英’翻訳を行う．

手順 2 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．

手順 3 手順 1 で得られた英語文 E' と日英対訳コーパスの英語文を用いて，英’英翻訳モデルを作成する．英’英ルールテーブルの例を以下の表 4.1 に示す．

表 4.1 英’英ルールテーブルの例

[X]		This is		This is		0.51366	0.364442	0.644231	0.267549				
[X]		This is		This is a		0.0253376	0.364442	0.00320513	0.00926709				
[X]		This is	[X,1][X,2]	.		This is	[X,1][X,2]	.		0.5170	0.3639	0.6679	0.26
[X]		new		new		0.776831	0.73591	0.767031	0.806708				
[X]		new		a new		0.135188	0.73591	0.0260723	0.0279419				

手順 4 前処理として，テストデータ (入力文) の日本語文 J に対して，RMT で日英’翻訳を行う．

手順 5 手順 2 で作成した言語モデルと手順 3 で作成した翻訳モデルを用いて，手順 4 で得られた英語文 E' に対して HSMT で英’英翻訳を行う．得られた英語文 E を最終出力とする．

第5章 実験環境

5.1 ルールベース翻訳

ルールベース翻訳 (RMT) として, 東芝の TAURAS[13] を使用する.

5.2 翻訳モデルの学習

翻訳モデルの学習には, “train-model.perl[14]” を用いる.

5.3 言語モデルの学習

言語モデルの学習には, “SRILM[15]” の “ngram-count” を用いる. 本研究では, N -gram モデルは 5-gram とする.

5.4 デコーダ

本実験では句に基づく統計翻訳 (PSMT) のデコーダとして, “moses[14]” を用いる. また, 階層型統計翻訳 (HSMT) のデコーダとしても “moses” を用いる.

moses のパラメータは, “mert-moses.pl[14]” を用いてチューニングを行う. また, PSMT においては, moses の設定ファイル “moses.ini” の “distortion-limit” の値を “-1” に変更する. distortion-limit はフレーズの並び替えを制限する値である. -1 にすることで, 無制限とする.

5.5 ベースライン

本実験ではハイブリッド翻訳 (RMT+PSMT) をベースラインと定義する.

5.6 評価

自動評価として，BLUE，METEOR，RIBES を用いる．

また，人手評価として対比較評価を用いる．対比較評価の対象を，提案手法とベースラインの出力文からランダムに抽出した各 100 文とする．対比較評価の結果の表記方法を表 5.1 に示す．

表 5.1 対比較評価基準

RMT+PSMT	ベースラインの出力文の方が良い
RMT+HSMT	提案手法の出力文の方が良い
差なし	ベースラインの出力文と提案手法の出力文の単語対応や伝わる意味に差がない
同一出力	ベースラインの出力文と提案手法の出力文が完全一致

5.7 実験データ

本実験では，単文コーパスを用いた実験と重文複文コーパスを用いた実験を行う．

5.7.1 単文コーパスを用いた実験

辞書の例文より抽出した単文コーパス 182,899 文 [16] から，表 5.2 に示す文数を実験に用いる．

表 5.2 単文コーパス

英語・日本語学習文	各 100,000 文
英語・日本語テスト文	各 10,000 文
英語・日本語ディベロップメント文	各 1,000 文

統計翻訳の前処理として，各コーパスの日本語文に対して，“chasen[17]”を用いて形態素解析を行う．また，英語文に対して，“tokenizer.perl[14]”を用いて，わかち書きを行う．単文の実験データの例を表 5.3 に示す．

表 5.3 単文実験データの例

例 1	
日本語文	警察は容疑者を洗っている。
英語文	The police are investigating the suspect .
例 2	
日本語文	彼は仕事でアメリカへ飛んだ。
英語文	He flew to the United States on business .
例 2	
日本語文	学校にも行った。
英語文	I went to a school also .

5.7.2 重文複文コーパスを用いた実験

辞書の例文より抽出した重文複文コーパス 122,719 文 [16] から，表 5.4 に示す文数を実験に用いる．

表 5.4 重文複文コーパス

英語・日本語学習文	各 100,000 文
英語・日本語テスト文	各 10,000 文
英語・日本語ディベロップメント文	各 1,000 文

統計翻訳の前処理として，各コーパスの日本語文に対して，“chasen[17]”を用いて形態素解析を行う．また，英語文に対して，“tokenizer.perl[14]”を用いて，わかち書きを行う．重文複文の実験データの例を表 5.5 に示す．

表 5.5 重文複文実験データの例

例 1	
日本語文	その家を借りる事に決めてきた。
英語文	We have decided to rent that house .
例 2	
日本語文	この町は鉄道がないので不自由だ。
英語文	This town is inconvenient because of its lack of railways .
例 2	
日本語文	彼は離婚して一人になった。
英語文	He became single after he got divorced .

第6章 実験結果

6.1 自動評価

単文コーパスを用いた実験における，ベースライン (RMT+PSMT) と提案手法 (RMT+HSMT) の自動評価結果を表 6.1 に示す．重文複文コーパスを用いた実験における，ベースライン (RMT+PSMT) と提案手法 (RMT+HSMT) の自動評価結果を表 6.2 に示す．

表 6.1 単文における自動評価

	BLEU	METEOR	RIBES
RMT+PSMT	0.1727	0.4979	0.7455
RMT+HSMT	0.1698	0.4979	0.7479

表 6.2 重文複文における自動評価

	BLEU	METEOR	RIBES
RMT+PSMT	0.1440	0.4458	0.7012
RMT+HSMT	0.1435	0.4444	0.7059

単文コーパスを用いた実験では，いずれの自動評価法においても，ベースラインと提案手法の評価にほとんど差が見られなかった．重文複文コーパスを用いた実験でも，いずれの自動評価法においても，ベースラインと提案手法の評価にほとんど差が見られなかった．

6.2 人手評価

単文コーパスを用いた実験における，ベースライン (RMT+PSMT) と提案手法 (RMT+HSMT) の対比較評価結果を表 6.3 に示す．重文複文コーパスを用いた実験における，ベースライン (RMT+PSMT) と提案手法 (RMT+HSMT) の対比較評価結果を表 6.4 に示す．

表 6.3 単文における対比較評価

RMT+PSMT	RMT+HSMT	差なし	同一出力
11	8	49	32

表 6.4 重文複文における対比較評価

RMT+PSMT	RMT+HSMT	差なし	同一出力
10	8	61	21

単文コーパスを用いた実験では，ベースラインと提案手法の評価にほとんど差が見られなかった．重文複文コーパスを用いた実験でも，ベースラインと提案手法の評価にほとんど差が見られなかった．

6.3 翻訳例

本節では，実験によって得られたベースライン (RMT+PSMT) の出力文，提案手法 (RMT+HSMT) の出力文の例を示す．例は対比較評価の結果別に表にまとめる．また，中間言語である RMT の出力も記載する．

6.3.1 ベースライン (RMT+PSMT) の例

単文での実験における RMT+PSMT の例を表 6.5 に示す。

表 6.5 単文における RMT+PSMT の例

例 1	
入力文	これはベティーがかいた絵です。
正解文	This is a picture that Betty painted .
RMT	This is the picture which Betty drew .
RMT+PSMT	This is a picture that Betty drew .
RMT+HSMT	This is a picture which Betty .
例 2	
入力文	風が一日中吹き荒れた。
正解文	It blew hard all day .
RMT	The wind blew violently all day long .
RMT+PSMT	The wind blew hard all day long .
RMT+HSMT	The wind blew and all day long .
例 3	
入力文	彼は努力して社会的に高い地位に就いた。
正解文	He worked himself into a high position in society .
RMT	He tried hard and took the high post socially .
RMT+PSMT	He tried hard and took a high social position .
RMT+HSMT	He tried hard and took a high social .

表 6.5 の例 1 は、RMT+HSMT は動詞の欠落があるため、RMT+PSMT と判断した。
 例 2 は、RMT+HSMT は誤った単語対応をしているため、RMT+PSMT と判断した。
 また、例 3 は、RMT+HSMT は名詞の欠落があるため、RMT+PSMT と判断した。

重文複文での実験における RMT+PSMT の例を表 6.6 に示す .

表 6.6 重文複文における RMT+PSMT の例

例 1	
入力文	彼はあまりにもひどいそつきなので、もはやだれも彼を信用していない。
正解文	He is such a liar that nobody trusts him anymore .
RMT	Since he is a too much severe liar , nobody already trusts him .
RMT+PSMT	He is a terrible liar, nobody already believes him .
RMT+HSMT	He is too poor liar, Nobody believes him .
例 2	
入力文	なんとんでも家族は人間の共同生活の基礎である。
正解文	The family is still the primary human association .
RMT	Whatever it may call it , a family is the foundation of man's community life .
RMT+PSMT	No matter what you say , the family is a foundation of the human community life .
RMT+HSMT	No matter what you say , the family is the foundation of human life .
例 3	
入力文	彼らには能力に応じた中等教育を施すべきである。
正解文	They should be provided with a secondary education suited to their abilities .
RMT	You should give them secondary education according to capability .
RMT+PSMT	You should be given secondary education in accordance with their abilities .
RMT+HSMT	We should give them merely collateral in accordance with their abilities .

表 6.6 の例 1 は , RMT+HSMT は副詞の欠落があるため , RMT+PSMT と判断した .
 例 2 は , RMT+HSMT は誤った句対応をしているため , RMT+PSMT と判断した .
 また , 例 3 は , RMT+HSMT は誤った句対応をしているため , RMT+PSMT と判断した .

6.3.2 提案手法 (RMT+HSMT) の例

単文での実験における RMT+HSMT の例を表 6.7 に示す。

表 6.7 単文における RMT+HSMT の例

例 1	
入力文	陪審の評決が出た。
正解文	The jury has reached a verdict .
RMT	A jury's verdict came out .
RMT+PSMT	The jury's was guilty verdict .
RMT+HSMT	A jury's guilty verdict came out .
例 2	
入力文	飛行機は滑走路を行きすぎた。
正解文	The plane overran the runway .
RMT	The airplane went the runway too much .
RMT+PSMT	The plane went the runway .
RMT+HSMT	The plane went too much the runway .
例 3	
入力文	私のかばんがなくなってしまった。
正解文	My bag has gone .
RMT	My bag has been lost .
RMT+PSMT	My bag .
RMT+HSMT	My bag has been lost .

表 6.7 の例 1 は、RMT+HSMT は、RMT+PSMT よりも意味の正しい動詞を出力しているため、RMT+HSMT と判断した。例 2 は、RMT+PSMT は副詞の欠落があるため、RMT+HSMT と判断した。また、例 3 は、RMT+PSMT は動詞の欠落があるため、RMT+HSMT と判断した。

重文複文での実験における RMT+HSMT の例を表 6.8 に示す .

表 6.8 重文複文における RMT+HSMT の例

例 1	
入力文	K 商業 は 手ごわい 相手 だ 。
正解文	K Commercial High School is a formidable opponent .
RMT	K commerce is a strong partner .
RMT+PSMT	The K commerce is a .
RMT+HSMT	The K commercial is a strong rival .
例 2	
入力文	子ども の とき に 覚えた ことは 一生 忘れない もの だ 。
正解文	What is learned in the cradle is carried to the grave .
RMT	He never forgets what was memorized at the time of a child .
RMT+PSMT	I will never forget what I felt when she was a child .
RMT+HSMT	I will never forget what I learned at the time of a child .
例 3	
入力文	友人 に 宛て た 手紙 が 舞い 戻っ て きた 。
正解文	The letter to my friend returned to me .
RMT	The letter addressed to the friend has returned .
RMT+PSMT	A letter addressed to his friend .
RMT+HSMT	A letter addressed to his friend came back .

表 6.8 の例 1 は , RMT+PSMT は名詞の欠落があるため , RMT+HSMT と判断した .
 例 2 は , RMT+PSMT は誤った単語対応をしているため , RMT+HSMT と判断した .
 また , 例 3 は , RMT+PSMT は動詞の欠落があるため , RMT+HSMT と判断した .

6.3.3 差なしの例

単文での実験における差なしの例を表 6.9 に示す。

表 6.9 単文における差なしの例

例 1	
入力文	彼女はまだ口紅をつける年ではない。
正解文	She is not old enough to use lipstick .
RMT	She is not a year which still attaches a lipstick .
RMT+PSMT	She is not a year is still on in on lipstick .
RMT+HSMT	She is not a year still on lipstick is out .
例 2	
入力文	コンパスは常に北を指す。
正解文	The compass always points to the north .
RMT	Compass always points out north .
RMT+PSMT	Compass always points out the north .
RMT+HSMT	Compass always points north .
例 3	
入力文	悪い風邪が全国ではやっている。
正解文	Bad colds are prevailing all over the country .
RMT	Bad cold is doing in the whole country .
RMT+PSMT	The bad cold has a all over the country .
RMT+HSMT	Bad colds are all over the country .

表 6.9 の例 1 は、RMT+PSMT も RMT+HSMT も、不自然に述語が重複しており、意味が通らないため差なしと判断した。例 2 は、RMT+PSMT は不適切な前置詞を出力し、RMT+HSMT は前置詞の欠落があり、意味が通らないため差なしと判断した。また、例 3 は、どちらも、不適切な動詞を出力しており、意味が通らないため差なしと判断した。

重文複文での実験における差なしの例を表 6.10 に示す。

表 6.10 重文複文における差なしの例

例 1	
入力文	彼が協力してくれると心強い。
正解文	We would feel reassured if he collaborated .
RMT	I am secure when he cooperates .
RMT+PSMT	His cooperation would reassure us .
RMT+HSMT	We would feel emboldened if he collaborated .
例 2	
入力文	彼の生きていたのを見て全く驚いた。
正解文	I was amazed to see him alive .
RMT	He was completely surprised to see him being alive .
RMT+PSMT	He was really surprised to see him being alive .
RMT+HSMT	He was really surprised to see him alive .
例 3	
入力文	あとになってから彼の事故を知った。
正解文	I learned later of his accident .
RMT	His accident was got to know later .
RMT+PSMT	His accident to later .
RMT+HSMT	His background accident was later .

表 6.10 の例 1 は、RMT+PSMT も RMT+HSMT も、意味の通る正しい翻訳をしているため差なしと判断した。例 2 は、どちらも不適切な主語を出力しており、意味が通らないため差なしと判断した。また、例 3 は、どちらも意味が通らない文を出力しているため、差なしと判断した。

6.3.4 同一出力の例

単文での実験における同一出力の例を表 6.11 に示す .

表 6.11 単文における同一出力の例

例 1	
入力文	手続きは個人でして下さい。
正解文	Carry out the procedure by yourselves , please .
RMT	Procedure is an individual and please give it to me .
RMT+PSMT	There is an individual Please give it to me .
RMT+HSMT	There is an individual Please give it to me .
例 2	
入力文	風が帆をはらませた。
正解文	Wind swelled the sails .
RMT	The wind impregnated the sail .
RMT+PSMT	The wind impregnated the sails .
RMT+HSMT	The wind impregnated the sails .
例 3	
入力文	残金は来年度の予算に繰り入れる。
正解文	The balance is to be transferred to next year 's budget .
RMT	The balance is transferred to the budget for next year .
RMT+PSMT	The balance is transferred to the budget for the next year .
RMT+HSMT	The balance is transferred to the budget for the next year .

重文複文での実験における同一出力の例を表 6.12 に示す。

表 6.12 重文複文における同一出力の例

例 1	
入力文	裁判にかけられ有罪と分かった。
正解文	He was tried and found guilty .
RMT	It was applied to the trial and turned out to be guilt .
RMT+PSMT	It was tried and turned out to be guilty .
RMT+HSMT	It was tried and turned out to be guilty .
例 2	
入力文	彼女の演技はかなりひどい失敗だった。
正解文	Her performance was rather a failure .
RMT	Her performance was quite severe failure .
RMT+PSMT	Her performance was rather a failure .
RMT+HSMT	Her performance was rather a failure .
例 3	
入力文	町は死んだように静まり返っていた。
正解文	The town had become so silent that it seemed dead .
RMT	The town was as silent as the grave , as it died .
RMT+PSMT	The town was the grave, dead .
RMT+HSMT	The town was the grave, dead .

第7章 考察

7.1 ベースラインと提案手法に差のない原因

6章より，自動評価，人手評価ともにベースライン (RMT+PSMT) と提案手法 (RMT+HSMT) に翻訳精度の差が確認できなかった．この原因を調査するために，前処理である RMT の出力に着目する．

HSMT は文の構造を考慮する．したがって，提案手法において，RMT が文の構造の誤った文を出力したときに，HSMT によって文の構造の正しい文に翻訳されれば，翻訳精度の向上が期待できる．よって，以下の表 7.1 に示す判断条件で調査を行う．なお，本研究での文法的な正しさとは，SVO などの文型の正確さを指し，単語選択の正確さは考慮していない．

表 7.1 判断条件

RMT 文法 ×	ルールベース翻訳の出力文のうち文の構造を誤っている文
RMT+HSMT 文法	“RMT 文法 ×” かつ RMT+HSMT で文の構造を正しく翻訳している文

表 7.1 の判断条件で，RMT の出力文から 100 文抽出し調査を行った．結果を表 7.2 に示す．

表 7.2 判断条件で調査した結果

	単文	重文複文
RMT 文法 ×	3	5
RMT+HSMT 文法	0	0

表 7.2 より，RMT によって誤った文法が出力された文は単文 100 文中 3 文，重文複文 100 文中 5 文と非常に少なかった．RMT による前処理で，ほとんどの文が文法的に正しく翻訳されていたことが確認できる．また，文法的に誤った出力文のうち，HSMT によって正しい文法へと翻訳された文は単文 100 文中 0 文，重文複文 100 文中 0 文となった．

RMT 文法×のときの出力を以下に示す．表 7.3 が単文実験，表 7.4 が重文複文実験の出力である．

表 7.3 単文実験における RMT 文法×

単文 1	
入力文	この国々には今も男女間に社会的な差が存在する。
RMT 文法×	A difference still social among men and women exists in these countries .
RMT+HSMT	The difference still exists between men and women in these countries of society .
単文 2	
入力文	彼は成功にほくそえんている。
RMT 文法×	A success requires him by ほくそえん .
RMT+HSMT	Success requires him by ほくそえん .
単文 3	
入力文	二十二日の会合で決定、日本医師会雑誌に公表する。
RMT 文法×	It will announce to determination and the Japan Medical Association magazine officially by the meeting on the 22nd .
RMT+HSMT	It will release to Japan has advanced medical Association , the magazine by the meeting on Monday .

表 7.4 重文複文実験における RMT 文法×

重文複文 1	
入力文	生きるべきか死ぬべきか、それが問題だ。
RMT 文法×	Whether it should live should die or it is a problem .
RMT+HSMT	Whether or not to live my life to die or is the question .
重文複文 2	
入力文	この方がいつかお話しした松田さんです。
RMT 文法×	This gentleman is when or Mr . Matsuda that talked .
RMT+HSMT	The gentleman is and ” Matsuda that .

重文複文 3	
入力文	もし草木に花が咲かないものとしたら自然はどんなにさびしいものであろう。
RMT 文法 ×	If a flower shall not bloom on plants , nature rusts however , will be carried out and will be that of a potato .
RMT+HSMT	When the flowers are not bloom on plants , how rusts , and will be the potatoes .
重文複文 4	
入力文	彼の写真はその原生林の中に遺跡があることを証明した。
RMT 文法 ×	It proved that his photograph has ruins in the primeval forest .
RMT+HSMT	It proved his photographs have site in a moist jungle .
重文複文 5	
入力文	被告は社長と共謀して彼に海外旅行保険を掛けて殺害した。
RMT 文法 ×	The defendant conspired with the president , and hung and murdered overseas travel insurance to him .
RMT+HSMT	The defendant They conspired with the president , and and overseas trip insurance to him .

表 7.3 の単文 1 は、RMT が副詞 “still” の位置を誤って出力しているため、文法的に正しくないと判断した。RMT+HSMT では “still” が動詞の前に移動しており、この点は改善されている。しかし、“of society” は “difference” の後ろに来るべきである。また、単文 2 において、RMT は未知語の影響によって、文法的に正しい翻訳をしていないと考えられる。HSMT による後処理を行った後も、この文法構造がそのまま残っている。表 7.4 の重文複文 1 においても、“or” の位置において同様のことが確認できる。以上から、RMT を前処理に用いた提案手法は、前処理に大きく依存すると考えられる。

HSMT は文の構造を考慮する。しかし、RMT で文法が崩れた文は、修復することが困難であることが分かった。したがって、RMT+PSMT の PSMT の代わりに HSMT を用いるのは、有効ではない。以上が、RMT+HSMT が RMT+PSMT の実験結果とほとんど差がない原因だと考えている。

第8章 各手法の比較

本章では、本実験で用いた各翻訳手法(RMT, PSMT, HSMT, RMT+PSMT, RMT+HSMT)を自動評価と人手評価で比較した結果を示す。

8.1 自動評価

単文コーパスを用いた実験における、各手法の自動評価結果を表 8.1, 重文複文コーパスを用いた実験における、各手法の自動評価結果を表 8.2 に示す。

表 8.1 単文における自動評価比較

	BLEU	METEOR	RIBES
RMT	0.1296	0.4695	0.7253
PSMT	0.1345	0.4520	0.7115
HSMT	0.1417	0.4573	0.7198
RMT+PSMT	0.1727	0.4979	0.7455
RMT+HSMT	0.1698	0.4979	0.7479

表 8.2 重文複文における自動評価比較

	BLEU	METEOR	RIBES
RMT	0.0909	0.4030	0.6699
PSMT	0.1190	0.4152	0.6742
HSMT	0.1238	0.4233	0.6779
RMT+PSMT	0.1440	0.4458	0.7012
RMT+HSMT	0.1435	0.4444	0.7059

8.2 人手評価

第7章において，RMT(ルールベース翻訳)の出力を調査した結果，文法構造の正しい文を多く出力していた．そこで，本章では，RMT+PSMT 対 RMT，RMT+HSMT 対 RMT の対比較評価を行う．また，前処理を行っていない単体の HSMT と，RMT+HSMT の対比較評価を行う．

8.2.1 RMT+PSMT 対 RMT

単文での，RMT+PSMT と RMT の対比較評価結果を表 8.3 に示す．重文複文での，RMT+PSMT と RMT の対比較評価結果を表 8.4 に示す．

表 8.3 単文における対比較評価 (RMT+PSMT 対 RMT)

RMT+PSMT	RMT	差なし	同一出力
3	16	77	4

表 8.4 重文複文における対比較評価 (RMT+PSMT 対 RMT)

RMT+PSMT	RMT	差なし	同一出力
1	36	60	3

8.2.2 RMT+HSMT 対 RMT

単文での，RMT+HSMT と RMT の対比較評価結果を表 8.5 に示す．重文複文での，RMT+HSMT と RMT の対比較評価結果を表 8.6 に示す．

表 8.5 単文における対比較評価 (RMT+HSMT 対 RMT)

RMT+HSMT	RMT	差なし	同一出力
2	17	76	4

表 8.6 重文複文における対比較評価 (RMT+HSMT 対 RMT)

RMT+HSMT	RMT	差なし	同一出力
1	50	59	2

8.2.3 HSMT 対 RMT+HSMT

単文での、HSMT と RMT+HSMT の対比較評価結果を表 8.7 に示す。重文複文での、HSMT と RMT+HSMT の対比較評価結果を表 8.8 に示す。

表 8.7 単文における対比較評価 (HSMT 対 RMT+HSMT)

HSMT	RMT+HSMT	差なし	同一出力
3	33	60	4

表 8.8 重文複文における対比較評価 (HSMT 対 RMT+HSMT)

HSMT	RMT+HSMT	差なし	同一出力
4	24	59	2

表 8.3～8.6 より、RMT+HSMT や RMT+PSMT よりも、RMT のみの翻訳が優れていることが確認できる。また、表 8.7、8.8 より、RMT+HSMT は、HSMT のみの翻訳よりも優れていることが分かった。したがって、統計翻訳手法を用いて翻訳する際には、前処理として RMT を用いることは有効である。しかし、翻訳の精度は RMT 単体で用いた方が高くなると考えられる。

句に基づく統計翻訳や階層型統計翻訳は句の対応で翻訳を行っている。後処理の句の対応付けを誤ることが原因で、翻訳精度を悪くしている例が多くあった。翻訳例は、付録の表 A.1～A.24 に示す。そのため、より正しい句を選出する統計翻訳手法が必要であると考えている。

また、RMT+PSMT 対 RMT の対比較評価と RMT+HSMT 対 RMT の対比較評価結果は、自動評価と逆転する結果となった。この矛盾は自動評価の問題点を示していると考えている。より人手評価の結果に近い評価を行う自動評価法の検討が必要である。

第9章 おわりに

本研究では，従来のハイブリッド翻訳に用いられる句に基づく統計翻訳の代わりに階層型統計翻訳を用いることで，翻訳精度の向上を試みた．その結果，提案手法の BLEU スコアは単文において 0.1698，重文複文において 0.1435 となった．このスコアは従来のハイブリッド翻訳と比較して差のない結果であった．また，提案手法と従来のハイブリッド翻訳を対比較評価した結果においても，ほとんど差のない結果であった．

ルールベース翻訳の出力文を調査したところ，単文において 100 文中 97 文，重文複文において 100 文中 95 文が文法的に正しい文であった．よって提案手法では，ルールベース翻訳による前処理で多くの文が，英語の文法構造を出力している．つまり，後の統計翻訳において，文法構造を考慮した翻訳をする必要がなかった．したがって，ベースラインの句に基づく統計翻訳の代わりに，階層型統計翻訳を用いても同様な翻訳結果が得られたのだと考えている．

また，ルールベース翻訳のみの出力とハイブリッド手法(提案手法，ベースライン)の出力を比較した結果，ルールベース翻訳が優れていた．原因として，ルールベース翻訳において，正確な文法構造を得ることができたにも関わらず，後の統計翻訳において，誤った句対応が取られたことが挙げられる．今後は，正しい句対応を取る統計翻訳を検討することで，ハイブリッド翻訳の翻訳精度の向上を目指す．

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] 福田智大: “ルールベース翻訳を前処理に用いた統計翻訳”, 鳥取大学 平成 21 年度 卒業論文, 2010.
- [2] 久保田裕介, 村上仁一, 徳久雅人, 村田真樹: “階層型統計翻訳の調査”, 言語処理学会 第 18 回年次大会, pp.259-262, 2012.
- [3] H.Schwenk, Marta R.Costa-jussa, and Jose A.R.Fonollosa: “Continuous space language models for the IWSLT 2006 Task”, in IWSLT, pp.166-173, 2006.
- [4] Terumasa Ehara: “Rule Based Machine Translation Combined with Statistical Post Editor for Japanese to English Patent Translation”, Proceedings of Machine Translation Summit XI, Workshop on Patent Translation, pp.13-18, 2007.
- [5] Chiang David: “A Hierarchical Phrase-Based Model for Statistical Machine Translation.”, Proceedings of ACL-2005, pp.263-270, 2005.
- [6] Franz Josef Och: “Minimum Error Rate Training for Statistical Machine Translation”, Proceedings of ACL, pp.160-167, 2003.
- [7] Peter F.Brown, Stephen A.Della Pietra, Vincent J.Della Pietra, Robert L.Mercer: “The mathematics of statistical machine translation:Parameter Estimation”, Computational Linguistics, 1993.
- [8] Franz Josef Och, Hermann Ney: “A Systematic Comparison of Various Statistical Alignment Models”, Computational Linguistics, pp.19-51, 2003.
- [9] Papineni Kishore, Salim Roukos, Todd Ward, Wei-Jing Zhu: “BLEU: a Method for Automatic Evaluation of Machine Translation”, Proceeding of ACL, pp.311-318, 2002.

- [10] Banerjee Satanjeev, Lavie Alon: “METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments”, Proceeding of ACL-2005, pp.65-72, 2005.
- [11] 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.
- [12] WordNet: <http://wordnet.princeton.edu/>
- [13] Shinya Amano, Hideki Hirakawa, Yoshinao Tsutsumi: “TAURAS: The Toshiba machine translation system”, Manusr Program MT Summit, pp.15-23, 1987.
- [14] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Alexandra Constantin, Evan Herbst: “Moses: Open Source Toolkit for Statistical Machine Translation”, Proceedings of ACL-2007, pp.177-180, 2007.
- [15] Andreas Stolcke: “SRILM - am Extensible Language Modeling Toolkit”, 7th International Conference on Spoken Language Processing, pp.901-904, 2002.
- [16] 村上仁一, 藤波進: “日本語と英語の対訳文対の収集と著作権の考察”, 第一回コーパス日本語学ワークショップ, pp.119-130, Mar. 2012.
- [17] ChaSen: <http://chasen-legacy.sourceforge.jp/>

付録A 各手法の対比較評価例

A.1 RMT+PSMT 対 RMT (翻訳例)

A.1.1 RMT の例

単文での RMT は 100 文中 316 文であった。例を表 A.1 に示す。

表 A.1 単文における RMT の例 (RMT+PSMT 対 RMT)

例 1	
入力文	景気が回復した。
正解文	The economy has recovered .
RMT+PSMT	Business .
RMT	Business recovered .
例 2	
入力文	銀行から預金を引き出した。
正解文	I drew my money from the bank .
RMT+PSMT	I painted You from the bank .
RMT	Deposit was drawn out from the bank .
例 3	
入力文	彼は狭いすき間を無理やり通り抜けた。
正解文	He wedged himself through the narrow opening .
RMT+PSMT	He passed through the narrow space .
RMT	He passed through the narrow crevice forcibly .

表 A.1 の例 1 は、RMT+PSMT は動詞を出力していないため、RMT と判断した。例 2 は、RMT+PSMT は誤った句対応をしているため、RMT と判断した。例 3 は、RMT+PSMT は副詞の欠落があるため、RMT と判断した。

重文複文での RMT は 100 文中 36 文であった。例を表 A.2 に示す。

表 A.2 重文複文における RMT の例 (RMT+PSMT 対 RMT)

例 1	
入力文	ゆっくりとペダルを漕いで出発した。
正解文	He set out , pedaling slowly .
RMT+PSMT	The pedal was and left .
RMT	The pedal was rowed slowly and it left .
例 2	
入力文	ケーブル 自体 を 持っ て 引 っ 張 ら な い で く だ さ い 。
正解文	Do not pull on the cable itself .
RMT+PSMT	Do not with the cable .
RMT	Please do not pull with the cable itself .
例 3	
入力文	レフェリー は 彼ら に 不 利 な 判 定 を 下 し た 。
正解文	The referee decided against them .
RMT+PSMT	The referee made a judgment for them .
RMT	The referee made the disadvantageous judgment for them .

表 A.2 の例 1 は , RMT+PSMT は動詞や副詞の欠落があるため , RMT と判断した。例 2 は , RMT+PSMT は動詞の欠落があるため , RMT と判断した。例 3 は , RMT+PSMT は形容詞の欠落があるため , RMT と判断した。

A.1.2 RMT+PSMT の例

単文での RMT+PSMT は 100 文中 3 文であった。例を表 A.3 に示す。

表 A.3 単文における RMT+PSMT の例 (RMT+PSMT 対 RMT)

例 1	
入力文	その理論は広く応用がきく。
正解文	The theory has wide application .
RMT+PSMT	The theory has wide application .
RMT	application hears widely the theory .
例 2	
入力文	やかんを火にかけなさい。
正解文	Put the kettle on the fire .
RMT+PSMT	Put the kettle on the fire .
RMT	Cover a kettle over fire .
例 3	
入力文	彼は在米中に死んだそうだ。
正解文	He is said to have died while he stayed in America .
RMT+PSMT	It is said that he died in the United States .
RMT	It is said that he died in in the U .S .

表 A.3 の例 1 は、RMT は主語の選択と単語の選択を誤っているため、RMT+PSMT と判断した。例 2 は、RMT は動詞を誤っているため、RMT+PSMT と判断した。例 3 は、前置詞が重複しているため、RMT+PSMT と判断した。

重文複文での RMT+PSMT は 100 文中 1 文であった。例を表 A.4 に示す。

表 A.4 重文複文における RMT+PSMT の例 (RMT+PSMT 対 RMT)

例 1	
入力文	彼の行動から見ると彼は仕事がきらいらしい。
正解文	Judging from his actions , it appears that he dislikes his work .
RMT+PSMT	Judging from his actions , it seems that he dislikes his work .
RMT	Seemingly for him, work is a tendency when it sees from his action .

表 A.4 の例 1 は、RMT は主語の選択の誤りがあり意味の通らない文になっているため、RMT+PSMT と判断した。

A.1.3 差なしの例

単文での差なしの例を表 A.5 に示す .

表 A.5 単文における差なしの例 (RMT+PSMT 対 RMT)

例 1	
入力文	彼女は鍋で大根を煮た。
正解文	She cooked radish in a pot .
RMT+PSMT	She cooked a radish in the pot .
RMT	She boiled the Japanese radish in a pan .
例 2	
入力文	彼は私にチェスで勝った。
正解文	He was victorious over me in chess .
RMT+PSMT	He won me at chess .
RMT	He won me in chess .
例 3	
入力文	この病気の治療は長びく。
正解文	The disease requires lengthy treatment .
RMT+PSMT	This ill treatment is postponed .
RMT	This sick medical treatment is prolonged .

表 A.5 の例 1 は , RMT も RMT+PSMT も , 意味の通る正しい翻訳をしているため差なしと判断した . 例 2 は , RMT も RMT+PSMT も , 同様な意味であるため差なしと判断した . 例 3 は , RMT も RMT+PSMT も , 意味の通る正しい翻訳をしているため差なしと判断した .

重文複文での差なしの例を表 A.6 に示す。

表 A.6 重文複文における差なしの例 (RMT+PSMT 対 RMT)

例 1	
入力文	両親は彼の希望に添った学校を探した。
正解文	His parents looked for a school in line with his desires .
RMT+PSMT	My parents looked for the school meeting his desire .
RMT	Parents looked for the school meeting his hope .
例 2	
入力文	私たちのほとんどは旅をするだけの十分な余暇がない。
正解文	Most of us do not have sufficient leisure to travel .
RMT+PSMT	Most of us do not have sufficient intervals of travel .
RMT	Most of us do not have sufficient leisure which travels .
例 3	
入力文	彼女にご馳走してもらったので食費が浮いた。
正解文	She treated me so I saved money on food .
RMT+PSMT	Because I got her to treat, food expenses .
RMT	Since I got her to treat, food expenses floated .

表 A.6 の例 1 は、RMT も RMT+PSMT も、意味の通る正しい翻訳をしているため差なしと判断した。例 2 は、RMT も RMT+PSMT も、意味の通る正しい翻訳をしているため差なしと判断した。例 3 は、RMT は動詞の誤りがあり、RMT+PSMT は動詞の欠落があり、どちらも正しくないため差なしと判断した。

A.1.4 同一出力の例

単文での同一出力の例を表 A.7 に示す．重文複文での同一出力の例を表 A.8 に示す．

表 A.7 単文における同一出力の例 (RMT+PSMT 対 RMT)

例 1	
入力文	そのハンターはあらゆる音に気を配っていた。
正解文	That hunter is alert to every sound .
RMT+PSMT	The hunter was careful about all sound .
RMT	The hunter was careful about all sound .
例 2	
入力文	彼らの一人息子は死んでしまった。
正解文	Their only son was lost to them .
RMT+PSMT	Their only son has died .
RMT	Their only son has died .
例 3	
入力文	幾多の失敗の先例がある。
正解文	There are a number of unsuccessful examples .
RMT+PSMT	There is a precedent of many failures .
RMT	There is a precedent of many failures .

表 A.8 重文複文における同一出力の例 (RMT+PSMT 対 RMT)

例 1	
入力文	そのニュースを聞いて驚いた。
正解文	I was surprised at the news .
RMT+PSMT	He was surprised to hear the news .
RMT	He was surprised to hear the news .
例 2	
入力文	独創的思想家がないことは嘆かわしい。
正解文	There is a sad dearth of original thinkers .
RMT+PSMT	It is deplorable that there is no original thinker .
RMT	It is deplorable that there is no original thinker .

A.2 RMT+HSMT 対 RMT (翻訳例)

A.2.1 RMT の例

単文での RMT は 100 文中 17 文であった。例を表 A.9 に示す。

表 A.9 単文における RMT の例 (RMT+HSMT 対 RMT)

例 1	
入力文	海が私たちの目の前に広がった。
正解文	The sea spread out before us .
RMT+HSMT	The sea in front of us .
RMT	The sea spread in the face of us .
例 2	
入力文	加山君はテニスクラブの主将をしている。
正解文	Kayama is the captain of our tennis club .
RMT+HSMT	Miss Kayama holds the captain of the tennis club .
RMT	Mr . Kayama is acting as a captain of the tennis club .
例 3	
入力文	毒が回ってきた。
正解文	The poison was taking effect .
RMT+HSMT	The poison is .
RMT	Poison has taken effect .

表 A.9 の例 1 は，RMT+HSMT は動詞を出力しておらず文になっていないため，RMT と判断した。例 2 は，RMT+HSMT は名詞の誤りがあるため，RMT と判断した。例 2 は，RMT+HSMT は動詞の欠落があるため，RMT と判断した。

重文複文での RMT は 100 文中 50 文であった。例を表 A.10 に示す。

表 A.10 重文複文における RMT の例 (RMT+HSMT 対 RMT)

例 1	
入力文	彼女はよく食べるのに、少しも太らない。
正解文	Even though she eats a lot , she does not gain any weight at all .
RMT+HSMT	She often but he does not grow fat at all .
RMT	She does not grow fat at all, although often eaten .
例 2	
入力文	いちばんいい衣裳を着ていた。
正解文	She was in her best clothes .
RMT+HSMT	The best clothes were worn thin .
RMT	The best clothes were worn .
例 3	
入力文	労働は不名誉ではない。
正解文	Labour is honourable .
RMT+HSMT	Labor is not a .
RMT	Labor is not dishonorable .

表 A.10 の例 1 は、RMT+HSMT は主語の選択の誤りと、単語の欠落があるため、RMT と判断した。例 2 は、RMT+HSMT は誤った形容詞が出現しているため、RMT と判断した。例 3 は、RMT+HSMT は形容詞の欠落があるため、RMT と判断した。

A.2.2 RMT+HSMT の例

単文での RMT+HSMT は 100 文中 2 文であった。例を表 A.11 に示す。

表 A.11 単文における RMT+HSMT の例 (RMT+HSMT 対 RMT)

例 1	
入力文	彼はその木に自分のイニシャルを刻んだ。
正解文	He carved his initials on the tree .
RMT+HSMT	He carved his initials on the tree .
RMT	He minced his initial to the tree .
例 2	
入力文	町の人口は 1 2、0 0 0 である。
正解文	The town numbers 12,000 inhabitants .
RMT+HSMT	The population of a town is 12 and 000 .
RMT	The population of the town is twelve thousand .

表 A.11 の例 1 は、RMT は日本語の“刻ん”に対応する英単語の選択を誤っているため、RMT+HSMT と判断した。例 2 は、RMT は数字の表現に誤りがあるため、RMT+HSMT と判断した。

重文複文での RMT+HSMT は 100 文中 1 文であった。例を表 A.12 に示す。

表 A.12 重文複文における RMT+HSMT の例 (RMT+HSMT 対 RMT)

例	
入力文	もう 1 本 やって 切り上げよう。
正解文	Let 's have another bottle to finish with .
RMT+HSMT	I will make another and then finish up .
RMT	I will do one more and will revalue .

表 A.12 の例は、RMT は動詞の選択を誤っているため、RMT+HSMT と判断した。

A.2.3 差なしの例

単文での差なしの例を表 A.13 に示す。

表 A.13 単文における差なしの例 (RMT+HSMT 対 RMT)

例 1	
入力文	それでは君の無知を告白するようなものだ。
正解文	That amounts to a confession of your own ignorance .
RMT+HSMT	It is like that time confessing your ignorance .
RMT	It is like then, confessing your ignorance .
例 2	
入力文	彼には義理がある。
正解文	I have an obligation to him .
RMT+HSMT	I must repay my obligation to him .
RMT	I owe him a favor .
例 3	
入力文	倒木が道路をふさいでいた。
正解文	A fallen tree had blocked the road .
RMT+HSMT	He bore up the fallen tree the road .
RMT	倒木 had taken up the road .

表 A.13 の例 1 は、RMT も RMT+HSMT も、意味が通るため差なしと判断した。例 2 は、RMT も RMT+HSMT も、意味に差がないため差なしと判断した。例 3 は、RMT は未知語があり、RMT+HSMT は未知語はなくなっているが、主語の選択を誤っているため差なしと判断した。

重文複文での差なしの例を表 A.14 に示す。

表 A.14 単文における差なしの例 (RMT+HSMT 対 RMT)

例 1	
入力文	警察は八方手を尽くして犯人を追っている。
正解文	The police are trying all possible means to trace the criminal .
RMT+HSMT	The police tried in all directions running after the criminal .
RMT	The police tried the means in all directions and has run after the criminal .
例 2	
入力文	彼女が校長室に入るのを見た。
正解文	I saw her go into the principal 's room .
RMT+HSMT	She observed him enter the principal office .
RMT	She was seen go into the principal's office .
例 3	
入力文	父はおなかがすくと不機嫌になる。
正解文	My father gets grouchy when he is hungry .
RMT+HSMT	My father will be sore on you are hungry .
RMT	A father will become ill-humored if you are hungry .

表 A.14 の例 1 は，RMT も RMT+HSMT も，意味の通る正しい翻訳をしているため差なしと判断した．例 2 は，RMT も RMT+HSMT も，意味の通る正しい翻訳をしているため差なしと判断した．例 3 は，RMT も RMT+HSMT も，主語の選択を誤って正しくないため，差なしと判断した．

A.2.4 同一出力の例

単文での同一出力の例を表 A.15 に示す．重文複文での同一出力の例を表 A.16 に示す．

表 A.15 単文における同一出力の例 (RMT+HSMT 対 RMT)

例 1	
入力文	全 国 民 が 王 の 死 去 を 悼 ん だ 。
正解文	The whole land mourned the passing of the king .
RMT+HSMT	The whole nation mourned for a king 's death .
RMT	The whole nation mourned for a king 's death .
例 2	
入力文	彼 は 次 々 と 本 を 出 版 す る 。
正解文	He publishes book after book .
RMT+HSMT	He publishes one book after another .
RMT	He publishes one book after another .
例 3	
入力文	そ の ハ ン タ ー は あ ら ゆ る 音 に 気 を 配 っ て い た 。
正解文	That hunter is alert to every sound .
RMT+HSMT	The hunter was careful about all sound .
RMT	The hunter was careful about all sound .

表 A.16 単文における同一出力の例 (RMT+HSMT 対 RMT)

例 1	
入力文	あ そ こ に は 幽 霊 が 出 る と い う う わ さ だ 。
正解文	They say that place is haunted by ghosts .
RMT+HSMT	There is a rumor that a ghost appears there .
RMT	There is a rumor that a ghost appears there .

A.3 HSMT 対 RMT+HSMT (翻訳例)

A.3.1 HSMT の例

単文での HSMT は 100 文中 33 文であった。例を表 A.17 に示す。

表 A.17 単文における HSMT の例 (HSMT 対 RMT+HSMT)

例 1	
入力文	寒いので彼女は風邪をひいた。
正解文	Because it was cold , she caught a cold .
RMT	Since it was cold , she caught cold .
HSMT	She was caught a cold from the cold .
RMT+HSMT	It was cold are She caught a cold .
例 2	
入力文	彼がテニスクラブに入る。
正解文	He will join a tennis club .
RMT	The tablecloth slid down from the table .
HSMT	He will enter the tennis club .
RMT+HSMT	He is in the tennis club .
例 3	
入力文	カセットを正しく入れる。
正解文	Insert the cassette correctly .
RMT	A cassette is put in correctly .
HSMT	Place cassette properly .
RMT+HSMT	The cassette is properly .

表 A.17 の例 1 は、RMT+HSMT は意味の通らない文を出力しているため、HSMT と判断した。例 2 は、HSMT がより正しい動詞を出力しているため、HSMT と判断した。例 3 は、RMT+HSMT は動詞が誤っているため、HSMT と判断した。

重文複文での HSMT は 100 文中 24 文であった。例を表 A.18 に示す。

表 A.18 重文複文における HSMT の例 (HSMT 対 RMT+HSMT)

例 1	
入力文	昨夜そこを通ったときお祭りが行なわれていた。
正解文	A festival was going on there when I passed through last night .
RMT	The festival was performed when it passed along that last night .
HSMT	There was a festival , when I went there last night .
RMT+HSMT	The festival drew when I passed through that last night .
例 2	
入力文	人から補助を受けたくない。
正解文	I want to be independent .
RMT	I do not want to receive assistance from people .
HSMT	I do not want to aid from the people .
RMT+HSMT	I do not want to help from .
例 3	
入力文	彼は頭を打たれて目をまわした。
正解文	He was stunned by a blow on the head .
RMT	He had the head struck and turned eyes .
HSMT	A blow on the head stunned him .
RMT+HSMT	He had a head stunned him .

表 A.18 の例 1 は，RMT+HSMT は単語選択を誤っているため，HSMT と判断した。例 2 は，RMT+HSMT は名詞の欠落があるため，HSMT と判断した。例 3 は，RMT+HSMT は文法的に正しくないため，HSMT と判断した。

A.3.2 RMT+HSMT の例

単文での RMT+HSMT は 100 文中 3 文であった。例を表 A.19 に示す。

表 A.19 単文における RMT+HSMT の例 (HSMT 対 RMT+HSMT)

例 1	
入力文	彼は音楽で緊張をほぐした。
正解文	He eased his tension with music .
RMT	He loosened strain by music .
HSMT	He ほぐし tension in the music .
RMT+HSMT	He loosened tension with music .
例 2	
入力文	彼は国務次官に任官した。
正解文	He was appointed Undersecretary of State .
RMT	He was appointed to the Under Secretary of State .
HSMT	He of 次官 to 任官 .
RMT+HSMT	He was appointed to the Secretary of State .
例 3	
入力文	彼女は鍋で大根を煮た。
正解文	She cooked radish in a pot .
RMT	She boiled the Japanese radish in a pan .
HSMT	I boiled some She The in the pot .
RMT+HSMT	She heated the radish in the pot .

表 A.19 の例 1 は、HSMT は未知語を出力しているため、RMT+HSMT と判断した。例 2 は、HSMT は未知語を出力しているため、RMT+HSMT と判断した。例 3 は、HSMT は主語を誤り、文法も誤っているため、RMT+HSMT と判断した。

重文複文での RMT+HSMT は 100 文中 4 文であった。例を表 A.20 に示す。

表 A.20 重文複文における RMT+HSMT の例 (HSMT 対 RMT+HSMT)

例 1	
入力文	彼 が 来 れ る か ど う か 疑 わ し い 気 が す る 。
正解文	I am doubtful whether he can come .
RMT	I feel that it is doubtful for whether he comes or not .
HSMT	He is doubtful whether 来 れ .
RMT+HSMT	I feel I doubt whether he will come or not .
例 2	
入力文	彼 は 東 京 を 食 い 詰 め て 郷 里 へ も ど っ た 。
正解文	Down and out in Tokyo , he went back to his native place .
RMT	He became penniless Tokyo and returned to his home town .
HSMT	He ate up to Tokyo and return to 郷 里 .
RMT+HSMT	He ate up Tokyo and returned to his hometown .
例 3	
入力文	そ れ が 彼 が そ の 木 を 切 る の に 使 っ た ナ イ フ で す 。
正解文	It is the knife he cut the tree with .
RMT	It is the knife used for him to cut the tree .
HSMT	That is that he was spent a knife to cut the tree .
RMT+HSMT	It is a knife used for him to cut the tree .

表 A.20 の例 1 は，HSMT は未知語を出力しているため，RMT+HSMT と判断した。例 2 は，HSMT は未知語を出力しているため，RMT+HSMT と判断した。例 3 は，HSMT は名詞の欠落があるため，RMT+HSMT と判断した。

A.3.3 差なしの例

単文での差なしの例を表 A.21 に示す。

表 A.21 単文における差なしの例 (HSMT 対 RMT+HSMT)

例 1	
入力文	口を堅く閉じた。
正解文	He clamped his mouth shut .
RMT	The mouth was closed firmly .
HSMT	He closed his mouth .
RMT+HSMT	The mouth closed tightly .
例 2	
入力文	銀行から預金を引き出した。
正解文	I drew my money from the bank .
RMT	Deposit was drawn out from the bank .
HSMT	He drew a from the bank .
RMT+HSMT	Deposit I painted from the bank .
例 3	
入力文	彼は狭いすき間を無理やり通り抜けた。
正解文	He wedged himself through the narrow opening .
RMT	He passed through the narrow crevice forcibly .
HSMT	He passed a narrow by force .
RMT+HSMT	He went through a narrow space .

表 A.21 の例 1 は、HSMT も RMT+HSMT も意味の通る正しい翻訳をしているため、差なしと判断した。例 2 は、HSMT は目的語の欠落があり、RMT+HSMT は動詞を誤っているため差なしと判断した。例 3 は、HSMT は名詞の欠落があり、RMT+HSMT は副詞が欠落しているため差なしと判断した。

重文複文での差なしの例を表 A.22 に示す。

表 A.22 重文複文における差なしの例 (HSMT 対 RMT+HSMT)

例 1	
入力文	カードを伏せたまま配る。
正解文	Deal cards face down .
RMT	
HSMT	She serves card .
RMT+HSMT	My cards with it .
例 2	
入力文	今頃は店をしまう時分だ。
正解文	It is closing time now .
RMT	It distributes with a card turned down .
HSMT	It is time to get the store by .
RMT+HSMT	It is time to keep the store now .
例 3	
入力文	彼は家族を扶養する能力がない。
正解文	He cannot support his family .
RMT	He does not have the capability to support a family .
HSMT	He has no ability to support his family .
RMT+HSMT	He does not have the ability to support his family .

表 A.22 の例 1 は、HSMT も RMT+HSMT も単語が欠落しているため、差なしと判断した。例 2 は、HSMT も RMT+HSMT も動詞を誤っているため、差なしと判断した。例 3 は、HSMT も RMT+HSMT も同様の意味であるため、差なしと判断した。

A.3.4 同一出力の例

単文での同一出力の例を表 A.23 に示す．重文複文での同一出力は 100 文中には存在しなかったため，他の文から抽出した．例を表 A.24 に示す．

表 A.23 単文における同一出力の例 (HSMT 対 RMT+HSMT)

例 1	
入力文	その理論は広く応用がきく。
正解文	The theory has wide application .
RMT	application hears widely the theory .
HSMT	The theory has wide application .
RMT+HSMT	The theory has wide application .
例 2	
入力文	時計は 11 時を打った。
正解文	The clock struck eleven .
RMT	The clock struck at 11 .
HSMT	The clock struck at eleven .
RMT+HSMT	The clock struck at eleven .
例 3	
入力文	やかんを火にかけなさい。
正解文	Put the kettle on the fire .
RMT	Cover a kettle over fire .
HSMT	Put the kettle on the fire .
RMT+HSMT	Put the kettle on the fire .

表 A.24 重文複文における同一出力の例 (HSMT 対 RMT+HSMT)

例 1	
入力文	どうしていいかわからない。
正解文	I am at a loss what to do .
RMT	There is no telling what I may do .
HSMT	I do not know what to do .
RMT+HSMT	I do not know what to do .