

概要

統計翻訳は、対訳データから学習する翻訳モデルと、モノリンガルデータから学習する言語モデルを用いて、確率的に翻訳をする。

アラビア語-英語 [1] や、中国語-英語 [2] では数億文もの多量の対訳データが提供されている。一方、日本語-英語で提供されている対訳データの量は少ない。したがって、日英翻訳や英日翻訳において多量の対訳データを用いて翻訳モデルを作成することは困難である。翻訳モデルは対訳データを取得する必要があるため、モデルを学習するためにコストが掛かるが、言語モデルは単一言語だけなので、モデルを学習するコストが低い。

そこで本研究では、データを収集することが容易な、モノリンガルデータを大量に使用して、言語モデルを学習し、翻訳精度の変化を調べた。モノリンガルデータには、電子辞書から抜き出した文 (辞書文) [3]、特許翻訳文、Wikipedia から抜き出した文 (Wikipedia 文) をそれぞれ使用し、テスト文には辞書文を用いた。

その結果、自動評価と人手評価の日英翻訳と英日翻訳の結果で、テスト文と同分野の辞書文の翻訳精度は向上したが、テスト文と別分野の特許翻訳文と Wikipedia 文での向上は見られなかった。よって、テスト文と同分野のモノリンガルデータを増やす手法は有効であることが確認できた。また、追加実験として追加するモノリンガルデータの量を変化させた実験を行い、テスト文と同分野の辞書文では、データの増加量と比例して翻訳精度が向上していることがわかった。

目次

第1章	はじめに	1
第2章	統計翻訳システム	2
2.1	概要	2
2.2	翻訳モデル	3
2.2.1	IBM 翻訳モデル	3
2.2.1.1	Model1	4
2.2.1.2	Model2	5
2.2.1.3	Model3	6
2.2.1.4	Model4	7
2.2.1.5	Model5	8
2.2.2	GIZA++	8
2.2.3	フレーズテーブルの作成方法	9
2.3	言語モデル	12
2.3.1	N-gram モデル	12
2.3.2	スムージング	13
2.4	デコーダー	13
第3章	評価方法	14
3.1	自動評価	14
3.1.1	BLEU	14
3.1.2	NIST	15
3.1.3	METEOR	15
3.1.4	RIBES	16
3.1.5	TER	16
3.2	人手評価	17

第4章	実験方法	18
4.1	実験データ	20
4.2	翻訳モデルの学習	22
4.3	言語モデルの学習	22
4.4	デコーダ	22
4.5	パラメータのチューニング	22
4.5.1	評価方法	23
第5章	実験結果	24
5.1	自動評価	24
5.2	人手評価	25
5.3	実験結果のまとめ	26
5.4	出力例	26
第6章	データ量を変化させた実験	36
6.1	目的	36
6.2	実験内容	36
6.3	実験結果	37
6.3.1	自動評価	37
6.3.2	人手評価	46
6.3.3	実験結果のまとめ	52
6.3.4	出力例	52
第7章	考察	59
7.1	分野の依存性	59
7.2	データ量と翻訳精度の関係	59
7.3	評価の問題点	59
7.4	関連研究との比較	60
7.5	誤り解析	60
第8章	おわりに	62

目次

2.1	統計翻訳の流れ	2
4.1	日英翻訳におけるベースラインの流れ	18
4.2	日英翻訳における提案手法の流れ	19
6.1	日英翻訳の BLEU 値の変化	37
6.2	日英翻訳の NIST 値の変化	38
6.3	日英翻訳の METEOR 値の変化	39
6.4	日英翻訳の RIBES 値の変化	40
6.5	日英翻訳の TER 値の変化	41
6.6	英日翻訳の BLEU 値の変化	42
6.7	英日翻訳の NIST 値の変化	43
6.8	英日翻訳の RIBES 値の変化	44
6.9	英日翻訳の TER 値の変化	45
6.10	辞書文の日英翻訳の人手評価の変化	46
6.11	特許翻訳文の日英翻訳の人手評価の変化	47
6.12	Wikipedia 文の日英翻訳の人手評価の変化	48
6.13	辞書文の英日翻訳の人手評価の変化	49
6.14	特許翻訳文の英日翻訳の人手評価の変化	50
6.15	Wikipedia 文の英日翻訳の人手評価の変化	51

表 目 次

2.1	日英翻訳におけるフレーズテーブルの例	3
2.2	日英方向の単語対応	9
2.3	英日方向の単語対応	9
2.4	intersection の例	10
2.5	union の例	10
2.6	grow の例	10
2.7	grow-diag の例	10
2.8	grow-diag-final の例	11
2.9	grow-diag-final-and の例	11
3.1	対比較の出力例	17
3.2	対比較の評価基準	17
4.1	実験で使用する共通データ	20
4.2	実験で追加するモノリンガルデータ	20
5.1	日英翻訳の自動評価	24
5.2	英日翻訳の自動評価	24
5.3	日英翻訳の人手評価	25
5.4	英日翻訳の人手評価	25
5.5	辞書文を追加した場合の対比較で“+辞書文 ” の出力例 (日英翻訳)	27
5.6	辞書文を追加した場合の対比較で“ベースライン ” の出力例 (日英翻訳)	27
5.7	辞書文を追加した場合の対比較で“差なし” の出力例 (日英翻訳)	28
5.8	特許翻訳文を追加した場合の対比較で“+特許翻訳文 ” の出力例 (日英翻 訳)	28
5.9	特許翻訳文を追加した場合の対比較で“+ベースライン ” の出力例 (日英 翻訳)	29

5.10	特許翻訳文を追加した場合の対比較で“差なし”の出力例(日英翻訳) . . .	29
5.11	Wikipedia 文を追加した場合の対比較で“+Wikipedia 文 ”の出力例(日英翻訳)	30
5.12	Wikipedia 文を追加した場合の対比較で“+ベースライン ”の出力例(日英翻訳)	30
5.13	Wikipedia 文を追加した場合の対比較で“差なし”の出力例(日英翻訳) . .	31
5.14	辞書文を追加した場合の対比較で“+辞書文 ”の出力例(英日翻訳) . . .	31
5.15	辞書文を追加した場合の対比較で“+ベースライン ”の出力例(英日翻訳)	32
5.16	辞書文を追加した場合の対比較で“+差なし”の出力例(英日翻訳)	32
5.17	特許翻訳文を追加した場合の対比較で“+特許翻訳文 ”の出力例(英日翻訳)	33
5.18	特許翻訳文を追加した場合の対比較で“+ベースライン ”の出力例(英日翻訳)	33
5.19	特許翻訳文を追加した場合の対比較で“差なし”の出力例(英日翻訳) . . .	34
5.20	Wikipedia 文を追加した場合の対比較で“+Wikipedia 文 ”の出力例(英日翻訳)	34
5.21	Wikipedia 文を追加した場合の対比較で“+ベースライン ”の出力例(英日翻訳)	35
5.22	Wikipedia 文を追加した場合の対比較で“差なし”の出力例(英日翻訳) . .	35
6.1	追加実験の日英翻訳で使用するモノリンガルコーパス	36
6.2	追加実験の英日翻訳で使用するモノリンガルコーパス	36
6.3	日英翻訳の BLEU 値	37
6.4	日英翻訳の NIST 値	38
6.5	日英翻訳の METEOR 値	39
6.6	日英翻訳の RIBES 値	40
6.7	日英翻訳の TER 値	41
6.8	英日翻訳の BLEU 値	42
6.9	英日翻訳の NIST 値	43
6.10	英日翻訳の RIBES 値	44
6.11	英日翻訳の TER 値	45
6.12	自動評価の特徴	45

6.13	辞書文の日英翻訳の人手評価	46
6.14	特許翻訳文の日英翻訳の人手評価	47
6.15	Wikipedia 文の日英翻訳の人手評価	48
6.16	辞書文の英日翻訳の人手評価	49
6.17	特許翻訳文の英日翻訳の人手評価	50
6.18	Wikipedia 文の英日翻訳の人手評価	51
6.19	人手評価の特徴	51
6.20	データ量を変化させる実験のまとめ	52
6.21	追加する辞書文のデータ量を変化させた場合の出力例 (日英翻訳)	53
6.22	表 6.21 の対比較の判断基準	53
6.23	追加する特許翻訳文のデータ量を変化させた場合の出力例 (日英翻訳)	54
6.24	表 6.23 の対比較の判断基準	54
6.25	追加する Wikipedia 文のデータ量を変化させた場合の出力例 (日英翻訳)	55
6.26	表 6.25 の対比較の判断基準	55
6.27	追加する辞書文のデータ量を変化させた場合の出力例 (英日翻訳)	56
6.28	表 6.27 の対比較の判断基準	56
6.29	追加する特許翻訳文のデータ量を変化させた場合の出力例 (英日翻訳)	57
6.30	表 6.29 の対比較の判断基準	57
6.31	追加する Wikipedia 文のデータ量を変化させた場合の出力例 (英日翻訳)	58
6.32	表 6.31 の対比較の判断基準	58
7.1	英日翻訳における対比較の出力例	60

第1章 はじめに

統計的機械翻訳 (統計翻訳) では, 対訳データから学習する翻訳モデルと, モノリンガルデータから学習する言語モデルを用いて, 確率的に翻訳をする.

統計翻訳において, 対訳データが多ければ多いほど翻訳精度が向上することが知られている. アラビア語-英語 [1] や, 中国語-英語 [2] での統計翻訳は盛んであり, 対訳データとして, 数億文もの多量のデータが提供されている. しかし, 日本語-英語で提供されている対訳データの量は少ない. したがって, 日英翻訳や英日翻訳において多量の対訳データを用いて翻訳モデルを作成することは困難である.

一方言語モデルでは, 多量のモノリンガルデータを学習することにより, 単語の遷移確率の精度を向上することができる. これにより, 出力言語の単語の並び順を正しくすることができる. 特に日本語-英語間では, 日本語の語順が “SOV” であるのに対して, 英語の語順は “SVO” であるので, 統計翻訳の翻訳精度を向上させるためには, 出力言語の語順が重要になる. また, モノリンガルデータは単一言語であるため, データの収集が容易であり, 多量のデータを使用することができる.

そこで本研究では, 多量のモノリンガルデータを使用して, 言語モデルを学習し, 日英翻訳と英日翻訳の翻訳精度の向上を目指す. 実験では, 電子辞書から抜き出した単文 [3] と特許翻訳文, Wikipedia 文から抜き出した文の 3 種類の分野のモノリンガルデータを用いる. 別々の分野のモノリンガルデータをそれぞれ用いることにより, 分野ごとに翻訳結果に与える特性を調べる. その結果, テスト文と同分野の辞書文の翻訳精度は向上したが, テスト文と別分野の特許翻訳文と Wikipedia 文での向上は見られなかった.

本論文の構成は以下の通りである. まず, 2章で統計翻訳システムの概要を示し, 各モデルの学習について述べる. 3章では, 評価方法について述べる. 4章では, 実験に用いるデータやツールといった実験環境について述べる. 5章では, 実験の結果を示す. 5章では, 追加実験について示す. そして, 7章で考察を行い, 最後に8章で結論を述べ, まとめる.

第2章 統計翻訳システム

2.1 概要

統計翻訳とは、翻訳モデルより導き出された言語間の句の出現確率と、言語モデルにより導き出された単語の遷移確率の統計より、最適の翻訳結果を自動的に出力する翻訳方法である。現在の統計翻訳では“句に基づく統計翻訳”が主流である。句に基づく統計翻訳は、文脈における訳語の選択や語順の並び替えの精度において、“単語に基づく統計翻訳”より優れているためである。よって本研究でも、句に基づく統計翻訳を用いる。図 2.1 に日英統計翻訳の例を示す。

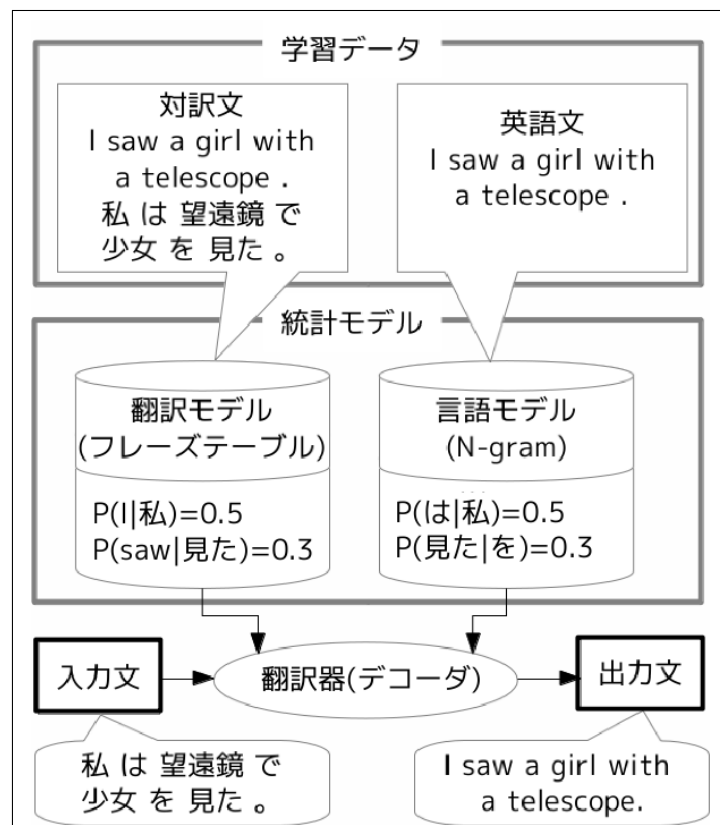


図 2.1: 統計翻訳の流れ

日英統計翻訳では，日本語入力文を J とすると，翻訳モデル $P(J|E)$ と言語モデル $P(E)$ の全ての組み合わせから，確率が最大になる英語文 \hat{E} を出力文とする．また， \hat{E} を探索する翻訳システムをデコーダと呼ぶ．

以下にその基本モデルを示す．

$$\hat{E} = \arg \max_E P(E|J) \quad (2.1)$$

$$\simeq \arg \max_E P(J|E)P(E) \quad (2.2)$$

2.2 翻訳モデル

翻訳モデルは，原言語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．翻訳モデルは一般的に、フレーズテーブルと呼ばれる表で管理されている．以下の表 2.1 にフレーズテーブルの例を示す．

表 2.1: 日英翻訳におけるフレーズテーブルの例

日フレーズ	英フレーズ	$P(J E)$	$\prod P(J E)$	$P(E J)$	$\prod P(E J)$
あの歌	that song	1.2	0.2	1.1	0.4
おもちゃ	toy	0.5	0.6	0.3	0.3
雨の後	after the rain	0.3	1.2	0.1	1.8

左から順に，日本語フレーズ，英語フレーズ，日英方向の翻訳確率 $P(J|E)$ ，日英方向の単語の翻訳確率の積，英日方向の翻訳 $P(E|J)$ ，英日方向の単語の翻訳確率の積である．

2.2.1 IBM 翻訳モデル

翻訳モデルの代表例として，Brown らが提案した IBM 翻訳モデル [4] がある．IBM 翻訳モデルは，Model1 から Model5 まで 5 つのモデルからなり，順に複雑な計算となる．IBM 翻訳モデルは本来，仏英翻訳を想定としているが，本研究では日英翻訳と英日翻訳を行うため，日英翻訳を想定して説明をする．

英語文 E ，日本語文 J の翻訳モデル $P(J|E)$ を計算するためにアラインメント a を用いる．アラインメントとは，日本語単語 J と英単語 E の対応を意味する．以下に IBM モデルの式を示す．

$$P(J | E) = \sum_a P(J, a | E) \quad (2.3)$$

IBM モデルのアライメントでは、日英統計翻訳の場合、日本語単語 j に対する英単語 e は 1 つあり、各英単語 e に対する日本語単語 j は 0 から n 個あると仮定する。また、日本語単語の対応関係として適切な英単語がなかった場合、英語文の文頭に空単語 e_0 があると仮定し、その日本語単語 j と空単語 e_0 を対応させる。

2.2.1.1 Model1

(2.3) 式は次のように置き換えられる。

$$P(J, a|E) = P(m|E) \prod_{k=1}^m P(a_k|a_1^{k-1}, j_1^{k-1}, m, E) P(j_k|a_1^k, j_1^{k-1}, m, E) \quad (2.4)$$

m は日本語の文長であり、 a_1^{k-1} は日本語文の 1 単語目から $k-1$ 単語目までのアライメントである。また、 j_1^{k-1} は日本語文の 1 番目から $k-1$ 番目までの単語を示す。(2.4) 式の右辺は、パラメータが多く複雑なため、計算が困難である。そこで、Model1 では (2.4) 式のパラメータを簡略化する。

- 文の長さの確率 ϵ は、 m と E に依存しない

$$\epsilon \equiv P(m|e)$$

- アライメントの確率は英語文の長さ l に依存する

$$P(a_k|a_1^{k-1}, J_1^{k-1}, m, E) \equiv (l+1)^{-1}$$

- 日本語の翻訳確率 $t(J_k|e_{a_k})$ は、日本語単語 J_k に対応づけられる英単語 e_{a_k} に依存する

$$P(J_k|a_1^k, J_1^{k-1}, m, E) \equiv t(J_k|e_{a_k})$$

これらの定理を用いて、パラメータを簡略化した場合の $P(J, a|E), P(J|E)$ は以下になる。

$$P(J, a|E) = \frac{\epsilon}{(l+1)^m} \prod_{k=1}^m t(j_k|e_{a_k}) \quad (2.5)$$

$$P(J|E) = \frac{\epsilon}{(l+1)^m} \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \prod_{k=1}^m t(j_k|e_{a_k}) \quad (2.6)$$

$$= \frac{\epsilon}{(l+1)^m} \prod_{k=1}^m \sum_{i=0}^l t(j_k|e_i) \quad (2.7)$$

Model1 は、翻訳確率 $t(j|e)$ の初期値が 0 以外の時、EM アルゴリズムを繰り返して得られる唯一の極大値より最適解を推定する。EM アルゴリズムは以下の手順で行われる。

手順 1 翻訳確率 $t(j|e)$ の初期値を設定する

手順 2 日本語と英語の対訳文 $(J^{(s)}, E^{(s)})$, $1 \leq s \leq S$ において、日本語単語 j と英単語 e が対応する回数の期待値を計算する。ここで $\delta(j|j_k)$ は日本語文 J において日本語単語 j が出現する回数を表し、 $\delta(e|e_i)$ は英語文 E において英単語 e が出現する回数を表す。

$$c(j|e; J, E) = \frac{t(j|e)}{t(j|e_0) + \dots + t(j|e_l)} \sum_{k=1}^m \delta(j|j_k) \sum_{i=1}^l \delta(e|e_i) \quad (2.8)$$

手順 3 英語文 $E^{(s)}$ のうち 1 回以上出現する英単語 e に対し、翻訳確率 $t(j|e)$ を計算する。ここで S は日英対訳文の文数を表す。

- 定数 λ_e を以下の式で計算する。

$$\lambda_e = \sum_j \sum_{s=1}^S c(j|e; J^{(s)}, E^{(s)}) \quad (2.9)$$

- 翻訳確率 $t(j|e)$ を以下の式で再計算する。

$$\begin{aligned} t(j|e) &= \lambda_e^{-1} \sum_{s=1}^S c(j|e; J^{(s)}, E^{(s)}) \\ &= \frac{\sum_{s=1}^S c(j|e; J^{(s)}, E^{(s)})}{\sum_k \sum_{s=1}^S c(j|e; J^{(s)}, E^{(s)})} \end{aligned} \quad (2.10)$$

手順 4 翻訳確率 $t(j|e)$ が収束するまで手順 2 と 3 を繰り返す

2.2.1.2 Model2

Model1 において、アライメントの確率は英語文の長さ l のみに依存する。そこで、Model2 では、 k 単語目のアライメント a_k 、日本語文の長さ m にも依存するとし、以下のように示す。

$$a(a_k|k, m, l) \equiv P(a_k|a_1^{k-1}, j_1^{k-1}, m, l) \quad (2.11)$$

これより, (2.6) 式は以下の式ようになる.

$$P(J|E) = \epsilon \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \prod_{k=1}^m t(j_k|e_{a_k}) a(a_k|k, m, l) \quad (2.12)$$

$$= \epsilon \prod_{k=1}^m \sum_{i=0}^l t(j_k|e_{a_k}) \alpha(l|k, m, l) \quad (2.13)$$

Model2 において, 対訳文中の英単語 e と日本語単語 j が対応付けされる回数の期待値 $c(j|e; J^{(s)}, E^{(s)})$ と, 日本語単語の位置 j と英単語の位置 i が対応付けられる回数の期待値 $c(i|k, m, l; J^{(s)}, E^{(s)})$ が存在する. これらは以下の式で求められる.

$$\begin{aligned} c(j|e; J^{(s)}, E^{(s)}) &= \frac{t(j|e)}{t(j|e_0) + \dots + t(j|e_l)} \sum_{k=1}^m \delta(j|j_k) \sum_{i=1}^l \delta(e|e_i) \\ &= \frac{t(j|e) \alpha(i|k, m, l) \delta(f|f_k) \delta(e|e_i)}{t(j_k|e_0) \alpha(0|k, m, l) + \dots + t(j_k|e_l) \alpha(l|k, m, l)} \end{aligned} \quad (2.14)$$

$$\begin{aligned} c(i|k, m, l; J^{(s)}, E^{(s)}) &= \sum_a P(a|e, j) \delta(i, a_k) \\ &= \frac{t(j|e) \alpha(i|k, m, l)}{t(j_k|e_0) \alpha(0|k, m, l) + \dots + t(j_k|e_l) \alpha(l|k, m, l)} \end{aligned} \quad (2.15)$$

なお, Model2 は, EM アルゴリズムで計算した場合, 複数の極大値を持ち, 最適解を獲得できない場合がある. しかし, Model1 は, Model2 において, $a(i|k, m, l) = (l+1)^{-1}$ となる特殊な状態であり, 最適解を求めることができる. このため, Model2 で最適解を求めるとき, Model1 を用いる.

2.2.1.3 Model3

Model3 は, Model1 や Model2 とは異なり, 1つの単語が複数の単語に対応する場合や, 単語の翻訳位置の距離を考慮する. なお, Model3 は以下4つのパラメータを用いる.

- $P(j|e)$
英単語 e が日本語単語 j に翻訳される確率
- $n(\phi|e)$
英単語 e が日本語単語に ϕ 個対応する確率

- $d(k|i, m, l)$
英語文の長さ l , 日本語文の長さ m のときに , i 番目の英単語 e_i が k 番目の日本語単語 j_k に翻訳される確率
- p_0, p_1
英単語に対応しない日本語単語が ϕ_0 個あるときの確率

なお p_0 は以下の式で表される .

$$P(\phi_0|\phi_1^l, e) = \binom{\phi_1 + \dots + \phi_l}{\phi_0} p_0^{\phi_1 + \dots + \phi_l} p_1^{\phi_0} \quad (2.16)$$

よって , Model3 は以下の式で表される .

$$\begin{aligned} P(J|E) &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l P(J, a|E) \\ &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \binom{m - \phi_0}{\phi_0} p_0^{m-2\phi_0} p_1^{\phi_0} \prod_{i=1}^l \phi_i! n(\phi_i|e_i) \\ &\quad \times \prod_{k=1}^m t(j_k|e_{a_k}) d(k|a_k, m, l) \end{aligned} \quad (2.17)$$

Model3 は , すべてのアライメントを計算する量が莫大であるため , EM アルゴリズムによる近似で求める .

2.2.1.4 Model4

Model4 は , Model2 や Model3 と違い , 単語の位置を絶対位置ではなく相対位置で考慮する . また , 各単語ごとの位置を考慮する . なお , Model4 は , 歪み確率 $d(k|i, m, l)$ を以下の 2 通りで考慮する .

- 英単語に対応する日本語単語が 1 つ以上あるとき , その中で最も文頭に近い場合

$$P\left(\prod_{[i]1} = k | \pi_1^{[i]-1}, \gamma_0^l, \phi_0^l, E\right) = d_1(k - \odot_{i-1} | A(e_{[i-1]}), B(j_k)) \quad (2.18)$$

- それ以外の場合

$$P\left(\prod_{[i]x} = k \mid \pi_{[i]1}^{x-1}, \pi_1^{[i]-1}, \gamma_0^l, \phi_0^l, E\right) = d_{>1}(k - \pi_{[i]x-1} \mid B(j_k)) \quad (2.19)$$

ここで A は英単語のクラス, B は日本語単語のクラスを表している.

2.2.1.5 Model5

Model4 は, 単語の位置に関して直前の単語のみを考慮している. そのため, 複数単語が同じ位置に生じたり, 単語が存在しない位置に生成される問題がある. その問題を改善するために Model5 では, 単語を空白部分に配置するように制約が施されている.

2.2.2 GIZA++

GIZA++[5] とは, 統計翻訳に用いるための単語の確率値を計算するツールである. IBM モデルの Model1 ~ 5 を用いて, 翻訳する言語と目的言語における単語の対応関係の確率値を計算する. GIZA++を用いた場合, 以下のファイルが出力される.

1. T TABLE (Translation Table)

T TABLE は, Model1 から Model3 により作成された翻訳確率 $P(f|e)$ のデータである. f は翻訳する言語で, e は目的言語である. T TABLE は各行が, 目的言語の単語 $ID(e_i d)$, 翻訳する言語の単語 $ID(f_i d)$, 翻訳する言語の単語から目的言語の単語へ翻訳する確率 ($P(f_i d|e_i d)$) で構成される.

2. N TABLE (Fertility Table)

N TABLE は, 目的言語の単語における繁殖数を表したデータである. N TABLE は各行が, 目的言語の単語 $ID(e_i d)$, 繁殖数が 0 である確率 (p_0), 繁殖数が 1 である確率 (p_1), ..., 繁殖数が n である確率 (p_n) で構成される.

2.2.3 フレーズテーブルの作成方法

GIZA++よりIBMモデルを推定することにより，最尤な単語アライメントを得る．これを日英・英日方向に対してそれぞれ行う．日本語文“風で松明が消えた”とその対訳英語文“The wind blew out the torch”を学習データとしたときの例として，日英方向の単語対応を表2.2に，英日方向の単語対応を表2.3に示す．また，表中の“●”は対応点である．

表 2.2: 日英方向の単語対応

	The	wind	blew	out	the	torch
風		●				
で			●			
松明						●
が			●			
消え			●			
た			●			

表 2.3: 英日方向の単語対応

	The	wind	blew	out	the	torch
風	●	●				
で					●	
松明			●	●		●
が						
消え						
た	●					

次に，両方向のアライメントから，両方向に1対多の対応を認めた単語アライメントの計算をヒューリスティックなルールにより行う．ヒューリスティックとは，ヒトの日々の意思決定に類似した直感的かつ発見的な思考方法である．基本のヒューリスティックとして，“intersection(積集合)”と，“union(和集合)”，“grow(成長)”，そして“grow-diag”がある．grow, grow-diag は intersection と union の中間である．intersection からスタートし，既に採用した対応点の周りに union の対応点を加えていく．grow では縦と横の方向に，grow-diag では縦と横と対角に union の対応点がある場合に，その対応点を用いる．表2.4に intersection の例を，表2.5に union の例を，表2.6に grow の例を，表2.6に grow-diag の例を示す．なお，表中の ● は，intersection から追加された対応点を示す．

表 2.4: intersection の例

	The	wind	blew	out	the	torch
風		•				
で						
松明						•
が						
消え						
た						

表 2.5: union の例

	The	wind	blew	out	the	torch
風	•	•				
で			•		•	
松明			•	•		•
が			•			
消え			•			
た	•		•			

表 2.6: grow の例

	The	wind	blew	out	the	torch
風		•				
で						
松明						•
が						
消え						
た						

表 2.7: grow-diag の例

	The	wind	blew	out	the	torch
風		•				
で						
松明						•
が						
消え						
た						

最後に、最終処理のヒューリスティックスとして、“final”と、“final-and”を用いる。finalは、少なくとも片方の言語の単語の単語対応がない場合に、unionの単語対応を追加する。また、final-andは、両側言語の単語の単語対応がない場合に、unionの候補対応点を追加する。表 6.10 に grow-diag-final の例を、表 2.9 に grow-diag-final-and の例を示す。ここでの表中の \bullet は、grow-diag から追加された対応点を示す。

表 2.8: grow-diag-final の例

	The	wind	blew	out	the	torch
風	•	•				
で			•		•	
松明						•
が						
消え						
た						

表 2.9: grow-diag-final-and の例

	The	wind	blew	out	the	torch
風	•	•				
で			•		•	
松明						•
が						
消え						
た						

得られた単語アライメントから、全ての矛盾しないフレーズ対を得る。このとき、そのフレーズ対に対して翻訳確率を計算し、その確率値をフレーズ対に付与することで、フレーズテーブルを作成する。

2.3 言語モデル

言語モデルは、モノリンガルデータから単語列の生成確率を付与するモデルである。日英翻訳では、翻訳モデルを用いて生成された翻訳候補から、英語として自然な文を選出するため、英語文の生成確率を用いることで、翻訳結果を出力することができる。

2.3.1 N-gram モデル

統計翻訳では一般的に、 N -gram モデルを用いる。 N -gram モデルは、単語列 $\omega_1, \omega_2, \dots, \omega_i$ が与えられたときに、その後に単語 x がくる確率 $P(x|\omega_1, \omega_2, \dots, \omega_i)$ を、すぐ前の $n-1$ 個の単語を条件とした確率として、計算するモデルである。以下に計算式を示す。

$$P(\omega_1^n) = P(c_1) \times P(c_2|c_1) \times P(c_3|c_2c_1) \cdots \quad (2.20)$$

$$= \prod_{i=1}^n P(\omega_i|c_1c_2c_3 \cdots c_{i-1}) \quad (2.21)$$

$$\simeq \prod_{i=1}^n P(\omega_i|\omega_{i-N+1}, \omega_{i-N+2}, \cdots \omega_{i-1}) \quad (2.22)$$

例えば，“I am Japanese.” という英語文 e に対する 2-gram モデルは以下のようなになる。

$$P(e = \text{“I am Japanese.”}) \simeq P(I) \times P(\text{am}|I) \times P(\text{Japanese|am}) \times P(.|\text{Japanese}) \quad (2.23)$$

3-gram モデルなら $P(\text{Japanese|I am})$ ，4-gram モデルなら $P(.|\text{I am Japanese})$ となる。

2.3.2 スムージング

N -gram モデルにおいて，信頼できる値を算出するためには大規模コーパスを使用しなければいけない．しかし，実際は多くの単語列の出現率は0となるため，信頼できる値を推定できない場合が多い．そこで，それを解決するためのスムージングの手法が提案されている．スムージングとは，生起確率が0となるのを防ぐために，大きい確率値を小さくし，小さい確立値を大きくする手法である．代表的な手法としてバックオフスムージング (back-off smooting) があり，バックオフスムージングでは学習データに出現しない N -gram の値を $(N-1)$ -gram の値から推定を行う．3-gram の場合の例を以下の式に示す．

$$P(w_i|w_{i-1}^{i-2}) = \begin{cases} \lambda(w_{i-2}^i)p(w_i|w_{i-2}^i) & \text{if } N(w_{n-2}^n) \\ (1 - \lambda_0(w_{i-2}^{i-1}))\alpha P(w_i|w_{i-1}) & \text{else } N(w_{n-2}^{n-1}) \\ P(w_i|w_{i-1}) & \text{other} \end{cases} \quad (2.24)$$

λ はディスカウントと呼ばれる係数で，出現しない N -gram に対して確率を付与するために， $(N-1)$ -gram から確率を推定する．また， α は確率の和を1にするための正規化係数である．“-ndiscount” や“-kndiscount” や“-ukndiscount” などの，ディスカウント係数を求める方法が多数存在する．

2.4 デコーダー

デコーダは翻訳モデルと言語モデルの全ての組み合わせから候補を探索して，確率が最大となるものを出力する．代表的なデコーダに mooses[6] がある

また，デコーダには言語モデルや翻訳モデルに対して重みを与えることができる．各モデルに与える重みをパラメータと呼び，このパラメータを最適化するために，パラメータチューニングを行う．

第3章 評価方法

本研究では，出力結果の評価に自動評価と人手評価を用いる．

3.1 自動評価

機械翻訳システムの翻訳精度を自動的に評価する手法として，あらかじめ用意した正解文と，翻訳システムで出力した出力文を比較する手法が一般的である．本研究では，BLEU[7]，NIST[7]，METEOR[8]，RIBES[9]，TER[10]を用いる．

3.1.1 BLEU

BLEU は機械翻訳分野において，一般的に使用される自動評価の手法である．BLEU は語順 (4-gram) が正しい場合に高いスコアを出す．BLEU は以下の式で計算される．

$$BLEU_{score} = BP \times \exp \left(\sum_{n=1}^N \frac{1}{N} \log p_n \right) \quad (3.1)$$

$$p_n = \frac{\sum_i \text{出力文 } i \text{ と正解文 } i \text{ で一致した } N\text{-gram の数}}{\sum_i \text{出力文 } i \text{ の } N\text{-gram の数}} \quad (3.2)$$

ここで， P_n は出力文と正解文の N -gram の一致率を表している．BLEU はこの一致率を 1-gram から 4-gram まで計算し，その幾何平均をとる．また，出力文が正解文より短い場合，“ $\sum_i \text{出力文 } i \text{ の } N\text{-gram の数}$ ” が小さくなり，不当にスコアが高くなる可能性がある．そこで，正解文より短い文に対するペナルティとして， BP を用いる． BP の条件式を以下に示す．

$$BP = \begin{cases} 1 & (c > r) \\ e^{1-\frac{r}{c}} & (c \leq r) \end{cases} \quad (3.3)$$

ここで、 c は出力文の単語数、 r は正解文の単語数を表している。出力文が正解文より長い場合は $BP = 1$ となり、出力文が正解文より短い場合は、その長さに応じて BP が小さくなり、BLEU 値も小さくなる。BLEU は 0 から 1 の値をスコアとして出力し、スコアの値が高いほど翻訳精度が高いことを示す。

3.1.2 NIST

NIST では BLEU と同様に語順 (5-gram) の正しさで比較を行うが、BLEU とは情報量によって重み付けをしている点が異なる。また、ペナルティ関数が異なっており、NIST の方が文長を考慮するペナルティが課せられる。NIST は以下の式で計算される。

$$NIST_{score} = BP \times \sum_{N=1}^N \frac{\sum_i \left(\sum_{\substack{\text{出力文 } i \text{ と正解文 } i \text{ に} \\ \text{共通する } w_1 \dots w_n}} Info_i(w_1 \dots w_n) \right)}{\sum_i \text{出力文 } i \text{ と正解文 } i \text{ で一致した } N\text{-gram の数}} \quad (3.4)$$

$$Info_i(w_1 \dots w_n) = \log_2 \frac{\text{評価コーパス中 } (w_1 \dots w_{n-1}) \text{ の数}}{\text{評価コーパス中の } (w_1 \dots w_n) \text{ の数}} \quad (3.5)$$

NIST は 0 から ∞ の値をスコアとして出力し、スコアの値が高いほど翻訳精度が高いことを示す。

3.1.3 METEOR

METEOR は再現率 R と適合率 P に基づく F 値に対して、単語の非連続性に対するペナルティ関数 Pen を利用した評価基準である。METEOR は以下の式で計算される。

$$METEOR_{score} = F_{mean} \times (1 - Pen) \quad (3.6)$$

$$F_{mean} = \frac{P \times R}{\alpha \times P + (1 - \alpha) \times R} \quad (3.7)$$

$$Pen = \gamma \times \left(\frac{c}{m}\right)^\beta \quad (3.8)$$

ペナルティ関数 Pen において、 m は出力文と正解文の単語の一致率を表す。そして、 c は一致した単語を対象に、正解文と語順が同じものを 1 つのまとまりとして統合した場合の、まとまりの数を表す。そのため、出力文と正解文が同じ文であるとき $c = 1$ となる。また、一致率の計算において、WordNet による類義語を用いて、似た意味を持つ単語は同一であると判断される。METEOR は単語属性 (3 人称単数など) が正しい場合に、高い

スコアを出す特性を持つ。 α, β, γ の値はパラメータである。METEORは0から1の値をスコアとして出力し、スコアの値が高いほど翻訳精度が高いことを示す。METEORは日本語辞書が存在しないため、本研究では出力文が日本語文の英日翻訳では使用しない。

3.1.4 RIBES

RIBESは、正解文と出力文の間で、共通単語の出現順序を順位相関係数を用いて、評価を行う自動評価法である。RIBESは以下の式で計算される。

$$RIBES(S) = NSR \times P^\alpha \quad (3.9)$$

$$RIBES(K) = NKT \times P^\alpha \quad (3.10)$$

$$P = \frac{n}{h} \quad (3.11)$$

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。また P はペナルティであり、 n は出力文と正解文との間で共通な単語の数、 h は出力文の単語数で計算する。 α はペナルティに対する主として使用され、 $0 \leq \alpha \leq 1$ の値である。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。RIBESは0から1の値をスコアとして出力し、スコアの値が高いほど翻訳精度が高いことを示す。

3.1.5 TER

TERはエラーレートであり、WER[10]に句単位での並び替えを加えて語順の制約を緩めた自動評価法である。TERは以下の式で計算される。

$$TER = \frac{\text{挿入語数} + \text{削除語数} + \text{置換語数} + \text{並び替え句数}}{\text{正解文の語数}} \quad (3.12)$$

TERは0から100の値をスコアとして出力し、スコアの値が低いほど翻訳精度が高いことを示す。また、WERについて以下に示す。

WER

WERはエラーレートであり、出力文を正解文に変換するために必要な単語単位の操作(挿入、削除、置換)の回数の割合を表す自動評価法である。WERは以下の式で計算

される。

$$TER = \frac{\text{挿入語数} + \text{削除語数} + \text{置換語数}}{\text{正解文の語数}} \quad (3.13)$$

WER は語順に敏感であり，日本語などの語順の制約の緩い言語では，意味が通る文に対しても，大きなペナルティを課してしまう欠点がある。

3.2 人手評価

本研究では人手評価として，対比較評価を用いる。対比較評価は，二つの文を相対的に比較して，どちらの品質がより優れているかを人手で選択する評価方法である。対比較を行う出力例を表 3.1 に，評価基準を表 3.2 示す。

表 3.1: 対比較の出力例

入力文	I'm always ready to help .
正解文	いつでも喜んでお手伝いします。
ベースライン	私にはいつもました。
提案手法	いつでも助けます。

表 3.2: 対比較の評価基準

ベースライン	ベースラインの出力文が提案手法の出力文より翻訳品質が優れている
提案手法	提案手法の出力文がベースラインの出力文より翻訳品質が優れている
差なし	2つの出力文の翻訳品質に明確な差がない
同出力	2つの出力文が完全に同じ

表 3.1 では，ベースラインの出力文と提案手法の出力文を比較すると，提案手法の出力結果の方が優れている。よって，この例では表 3.2 において“提案手法”を選択する。また，統計翻訳における対比較評価では，翻訳品質を比較するため，必ずしも意味が通る文である必要がない。よって“差なし”や“同出力”では，どちらも意味が通る文である場合と，どちらも意味が通らない文である場合が存在する。

第4章 実験方法

本研究では統計モデルの学習に使用する対訳コーパスの数を固定して、言語モデルの学習に使用するモノリンガルデータの数を増やす実験を日英翻訳・英日翻訳でそれぞれ行う。本手法を提案手法とし、モノリンガルデータを増やす前の手法をベースラインとする。日英翻訳におけるベースラインの流れを図4.1に、提案手法の流れを図4.2に示す。

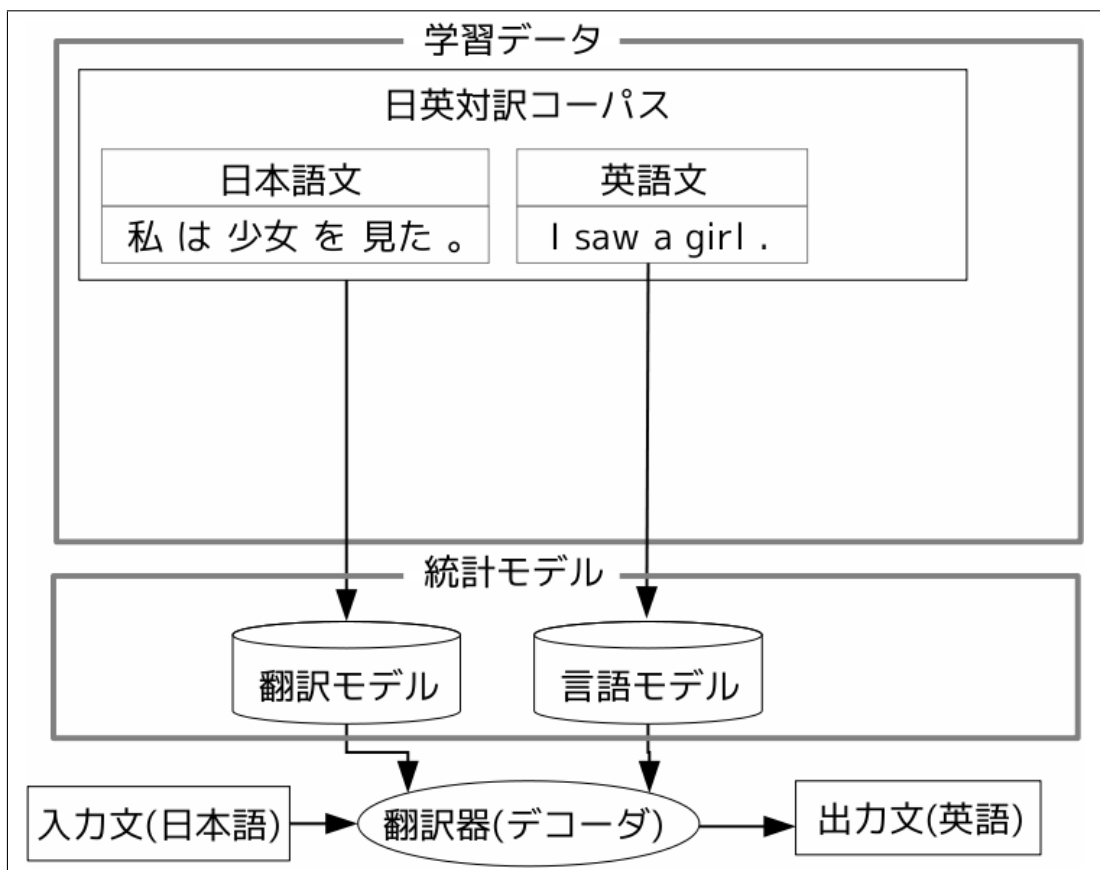


図 4.1: 日英翻訳におけるベースラインの流れ

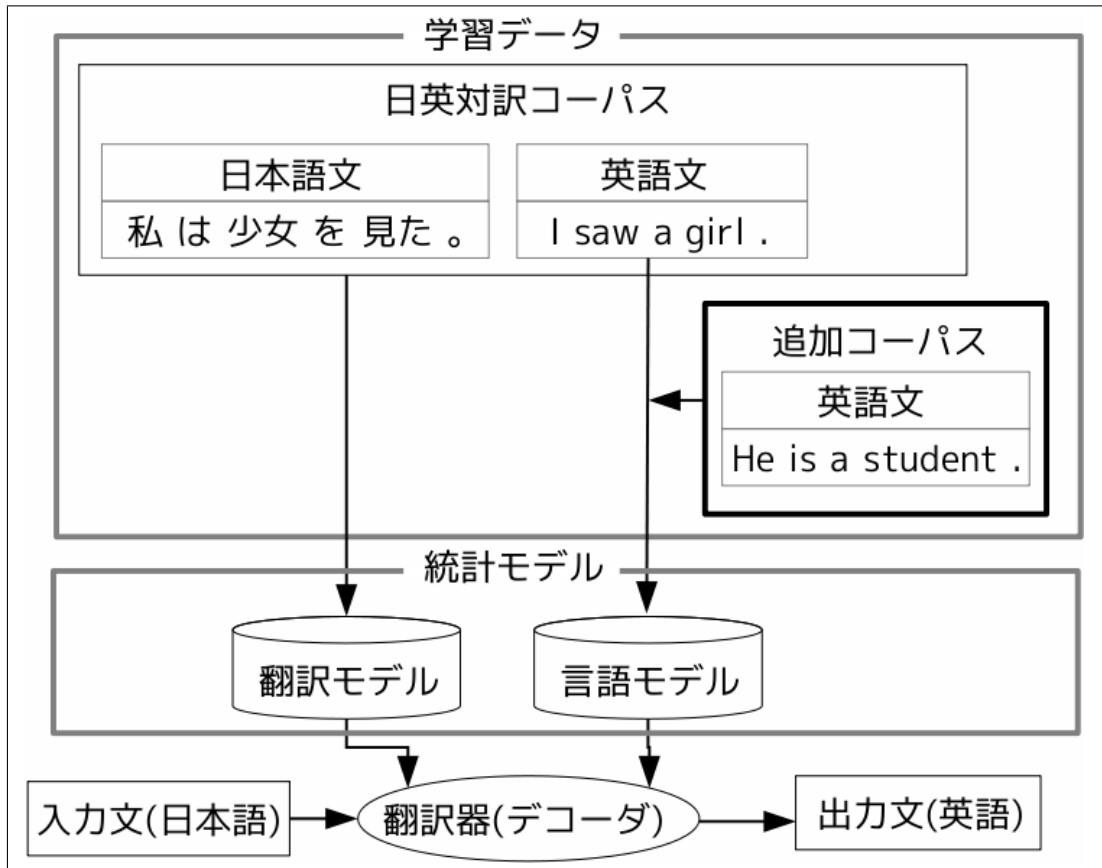


図 4.2: 日英翻訳における提案手法の流れ

日英翻訳における提案手法の手順を以下に示す。

- 手順 1 日英対訳コーパスを学習データとして翻訳モデルを作成する
- 手順 2 日英対訳コーパスの英語文に追加コーパスとして多量のモノリンガルデータを加える
- 手順 3 手順 2 のデータを学習データとして言語モデルを作成する
- 手順 4 手順 1 と手順 3 で作成したモデルを用いて日英統計翻訳を行う

4.1 実験データ

本研究では単文コーパスを用いる．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，MeCab[13]を用いて形態素解析を行う．また，英語文に対して“tokenizer.sed [14]”を用いて正規化を行う．実験で使用する対訳コーパス，テスト文には電子辞書から抜き出した文（辞書文）[3]を使用する．尚，テスト文はオープンデータとする．実験で使用するデータを表 4.1 に示す．

表 4.1: 実験で使用する共通データ

データ名	日英翻訳	英日翻訳
対訳コーパス	100,000 文	100,000 文
テスト文	10,000 文	10,000 文

追加コーパスとして，辞書文，特許翻訳文，Wikipedia から抜き出した文（Wikipedia 文）をそれぞれ使用する．追加コーパスの辞書文の中身は表 4.1 のデータとは別のものである．これらの 3 種類のモノリンガルデータを別々に加え，日英翻訳・英日翻訳で合計 6 つの実験をする．追加するモノリンガルコーパスの内訳を表 4.2 に示す．また，各コーパスの例文を次に示す．

表 4.2: 実験で追加するモノリンガルデータ

モノリンガルデータの種類	日英翻訳	英日翻訳
辞書文	788,433 文	806,324 文
特許翻訳文	3,407,225 文	3,407,231 文
Wikipedia 文	14,579,468 文	12,607,508 文

—— 辞書文 (日本語) ——

石油の発見でその国は裕福になった。
あの人の家はすぐ見つかった。
鉄道株が株式市場で暴落した。

—— 辞書文 (英語) ——

The discovery of oil enriched the country .
I soon found that person's house .
Rail stocks took a dive on the stock market .

—— 特許翻訳文 (日本語) ——

F E T 1 0 1 4 のゲートは信号 A N B に結合している。
この論理 1 はシフトレジスタ出力 1 0 8 へ与えられる。
いったん構成されると、経路を修正するのは困難である。

—— 特許翻訳文 (英語) ——

The gate of fet 1014 is coupled to signal anb .
This logic 1 is provided to the shift register output 108 .
This is especially true for highly polluted areas .

—— Wikipedia 文 (日本語) ——

アンパサンドとは「... と ...」を意味する記号である。
その使用は 1 世紀に遡ることができ、5 世紀中葉から現代に至るまでの変遷がわかる。
プログラミング言語では、C など多数の言語で AND 演算子として用いられる。

—— Wikipedia 文 (英語) ——

The discovery of oil enriched the country .
Hampshire Bus provide a daily service to and from Alton College on school days .
Hampigny is a commune in the Aube department in north-central France .
Hampshire County Council 's Mobile Library Service has served Liss since the 1 9 6 0 s .

4.2 翻訳モデルの学習

翻訳モデルを学習するために，“train-model.perl[6]”を用いる．

4.3 言語モデルの学習

言語モデルを学習するために，“SRILM[11]”の“ngram-count”を用いる． N -gram モデルに 5-gram の言語モデルを用いる．また，SRILM では学習データの量に限界がある．そのため，データ量の多い Wikipedia 文において，“clean-corpus-n.perl”を用いることにより，80 単語以上が出現する文をカットする．統計翻訳における単語量の多い文は不要とされているため，翻訳精度に影響は出ない．カットされた Wikipedia 文の日本語文，英語文のデータ量が表 4.2 である．

4.4 デコーダ

デコーダには，“moses[6]”を用いる．

4.5 パラメータのチューニング

moses のパラメータはいくつかのパラメータを設定することができる．パラメータの最適化として MERT (Minimum Error Rate Training)[12] という手法を用いる．MERT はディベロップメント文と呼ばれるデータを用いて試し翻訳をし，各文に対して上位 100 文程度の翻訳候補のなかで重みを変え，各文について上位にくるようにパラメータを調整する．本研究ではディベロップメント文に 1,000 文使用する．また，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．本研究では 3 種類のコーパスを扱うので，実験条件を揃えるために，各コーパスの実験ごとに同じパラメータを使う．これを各実験のベースラインにも適応するため，それらのベースラインの出力結果は異なることがある．

4.5.1 評価方法

本実験の評価方法として、人手評価と自動評価を用いる。

人手評価

出力文からランダムに100文を取りだし、追加コーパスを加えた実験の結果とベースラインの実験の結果で対比較評価を行う。

自動評価

日英翻訳ではBLEU、NIST、METEOR、RIBES、TER、英日翻訳ではBLEU、NIST、RIBES、TERを用いる。

第5章 実験結果

5.1 自動評価

日英翻訳の自動評価の結果を表 5.1 に，英日翻訳の自動評価の結果を表 5.2 に示す．

表 5.1: 日英翻訳の自動評価

コーパスの内容	BLEU	NIST	METEOR	RIBES	TER
辞書文					
ベースライン	0.1382	4.8771	0.4552	0.7105	69.511
追加コーパスあり	0.1499	4.9611	0.4641	0.7123	70.004
特許翻訳文					
ベースライン	0.1299	4.6154	0.4521	0.7051	73.437
追加コーパスあり	0.1339	4.7449	0.4512	0.7056	70.969
Wikipedia 文					
ベースライン	0.1310	4.6886	0.4540	0.7046	72.719
追加コーパスあり	0.1389	4.8149	0.4536	0.7024	71.893

表 5.2: 英日翻訳の自動評価

コーパスの内容	BLEU	NIST	RIBES	TER
辞書文				
ベースライン	0.1792	4.6187	0.7686	65.958
追加コーパスあり	0.1860	4.6845	0.7690	66.637
特許翻訳文				
ベースライン	0.1738	4.5834	0.7635	68.097
追加コーパスあり	0.1741	4.5835	0.7622	67.415
Wikipedia 文				
ベースライン	0.1672	4.2875	0.7622	71.641
追加コーパスあり	0.1815	4.4749	0.7605	67.957

自動評価の結果から，辞書文を追加した場合は TER 値を除いて，日英翻訳，英日翻訳ともに翻訳精度が向上している．一方特許翻訳文を追加した場合は，日英翻訳では BLEU

値，RIBES 値，TER 値が，英日翻訳では BLEU 値，TER 値の翻訳精度がわずかに向上している．Wikipedia 文を追加した場合は，BLEU 値と TER 値が日英翻訳・英日翻訳ともに翻訳精度がわずかに向上してゐる．

5.2 人手評価

日英翻訳における人手評価の結果を表 6.13 に，英日翻訳における人手評価の結果を表 6.16 に示す．表において，“+辞書文”とは，辞書文をベースラインの対訳コーパスに追加した実験の翻訳結果である．“+特許翻訳文”，“+Wikipedia 文”も同様である．

表 5.3: 日英翻訳の人手評価

ベースライン	+辞書文	差なし	同出力
12	16	42	30
ベースライン	+特許翻訳文	差なし	同出力
9	11	35	45
ベースライン	+Wikipedia 文	差なし	同出力
13	6	51	30

表 5.4: 英日翻訳の人手評価

ベースライン	+辞書文	差なし	同出力
7	14	58	21
ベースライン	+特許翻訳文	差なし	同出力
6	8	54	32
ベースライン	+Wikipedia 文	差なし	同出力
7	9	58	26

人手評価の結果より，辞書文を追加した日英翻訳・英日翻訳では，ベースラインと比べて翻訳精度が向上している．特許翻訳文と Wikipedia 文を追加した日英翻訳・英日翻訳では，翻訳精度に変化がなかった．特に Wikipedia 文を追加した日英翻訳では，ベースラインの方がいい結果となった．

5.3 実験結果のまとめ

テスト文と同分野の辞書文を追加したとき，日英翻訳・英日翻訳の自動評価では BLEU 値以外に変化がなかった．しかし，人手評価では日英翻訳・英日翻訳ともに向上が見られたため，人手評価を優先し，翻訳精度が向上したと言える．一方，テスト文と別分野の特許翻訳文を追加したとき，日英翻訳・英日翻訳の自動評価に変化がなかった．人手評価では日英翻訳・英日翻訳ともに変化が見られなかったため，翻訳精度にほとんど変化がなかったと言える．テスト文と別分野の Wikipedia 文を追加したときの日英翻訳では，自動評価で METEOR 値，RIBES 値が下がり，人手評価の結果が下がったことから，Wikipedia 文を追加する実験は有効性が示せなかった．特許翻訳文と Wikipedia 文を追加したときの翻訳精度は，日英翻訳・英日翻訳ともにあまり向上しなかった．

5.4 出力例

日英翻訳における辞書文を追加した場合のベースラインとの対比較の出力例を表 5.5，表 5.6，表 5.7 に，特許翻訳文を追加した場合のベースラインとの対比較の出力例を表 5.8，表 5.9，表 5.10 に，Wikipedia 文を追加した場合のベースラインとの比較の出力例を表 5.11，表 5.12，表 5.13 に示す．英日翻訳における辞書文を追加した場合のベースラインとの対比較の出力例を表 5.14，表 5.15，表 5.16 に，特許翻訳文を追加した場合のベースラインとの対比較の出力例を表 5.17，表 5.18，表 5.19 に，Wikipedia 文を追加した場合のベースラインとの比較の出力例を表 5.20，表 5.21，表 5.22 に示す．

また，各表において，判断基準となった箇所を太字で示す．

表 5.5: 辞書文を追加した場合の対比較で “+辞書文 ” の出力例 (日英翻訳)

入力文	彼は彼女に写真を見せた。
正解文	He showed her the photographs .
ベースライン	He gave a photograph in her .
+辞書文	He showed his photos to her .
入力文	彼はわがチームの主将としてふさわしい。
正解文	He is working to be captain of our team .
ベースライン	He is suitable for a captain of our team .
+辞書文	He is a suitable as captain of our team .
入力文	その試合はテレビで見ました。
正解文	I watched the game on television .
ベースライン	I saw a TV in the game .
+辞書文	I saw the match on television .

表 5.6: 辞書文を追加した場合の対比較で “ベースライン ” の出力例 (日英翻訳)

入力文	彼は鋭い勘をもっています。
正解文	He has a keen intuition .
ベースライン	He has a keen great instinct and intuition .
+辞書文	He has a great instinct and intuition .
入力文	その新しい奨学金は若い技術者に刺激となるだろう。
正解文	The new scholarship will provide a stimulus for young engineers .
ベースライン	The new scholarship will respond to young and engineers .
+辞書文	The new young engineers and will respond to the scholarship .
入力文	あの学生たちを例に挙げよう。
正解文	Take those students for example .
ベースライン	That students in the 挙げよ example .
+辞書文	He 挙げよ with the students .

表 5.7: 辞書文を追加した場合の対比較で“差なし”の出力例 (日英翻訳)

入力文	このおかげで私はいっそう思いどおり選択できる。
正解文	This will afford me wider freedom of choice .
ベースライン	This saved me with can choose confirmed .
+辞書文	This saved me with can choose confirmed .
入力文	12歳に達しない子供は入場無料。
正解文	Children under 12 are admitted free .
ベースライン	The child at twelve years old no admission free of charge .
+辞書文	Admission is free for children at twelve years old .
入力文	鬼ごっこ遊びに飽きてしまった。
正解文	We were tired of playing tag .
ベースライン	I am to play tag .
+辞書文	I am going to play tag .

表 5.8: 特許翻訳文を追加した場合の対比較で“+特許翻訳文”の出力例 (日英翻訳)

入力文	何物も彼を満足させなかった。
正解文	Nothing contented him .
ベースライン	I did not satisfy him as many things .
+特許翻訳文	Several things can also did not satisfy him .
入力文	ぼくは純粋な動機からそう言っているのだ。
正解文	I am speaking from a disinterested motive .
ベースライン	I have to say that from the neat motives .
+特許翻訳文	I say that from a neat motives .
入力文	その新しい奨学金は若い技術者に刺激となるだろう。
正解文	The new scholarship will provide a stimulus for young engineers .
ベースライン	The new scholarship to will be a young engineers .
+特許翻訳文	The new scholarship will respond to a young technology .

表 5.9: 特許翻訳文を追加した場合の対比較で “+ベースライン ” の出力例 (日英翻訳)

入力文	我田引水 は よせ 。
正解文	Don't be so self-seeking .
ベースライン	You shouldn't find 我田引水 .
+特許翻訳文	我田引水 .
入力文	長い 登り の あと で その 休息 は ありがたい 息抜き と なっ た 。
正解文	The rest gave welcome relief after the long climb .
ベースライン	The 休息 became ありがたい relaxation in the after a long ascends .
+特許翻訳文	The 休息 and ありがたい relaxation in the after a long ascends .
入力文	その後 の 彼 の 足取り は わから ない 。
正解文	His movements after that are unknown .
ベースライン	The don't understand his steps .
+特許翻訳文	The following his steps .

表 5.10: 特許翻訳文を追加した場合の対比較で “差なし” の出力例 (日英翻訳)

入力文	その 赤ん坊 は 母親 の ネックレス を いじっ て いた 。
正解文	The baby was fingering her mother's necklace .
ベースライン	The baby was nervously necklace on her mother .
+特許翻訳文	The baby was his mother's necklace nervously .
入力文	この おかげ で 私 は いっそう 思い どり 選択 できる 。
正解文	This will afford me wider freedom of choice .
ベースライン	This saved me with a confirmed can be choice .
+特許翻訳文	This help I can be confirmed with choice .
入力文	いつ でも 喜んで お手伝い します 。
正解文	I'm always ready to help .
ベースライン	Come whenever be happy to help you .
+特許翻訳文	be happy to help you any time .

表 5.11: Wikipedia 文を追加した場合の対比較で “+Wikipedia 文 ” の出力例 (日英翻訳)

入力文	羊がめえと鳴いていた。
正解文	The sheep were bleating .
ベースライン	The sheep えとめ were lowing .
+Wikipedia 文	A sheep were lowing えとめ .
入力文	住宅投資にも先細りの兆しが見える。
正解文	There are also signs that housing investments are tapering off .
ベースライン	I can see no signs of 先細り housing investments in .
+Wikipedia 文	There are signs of 先細り housing investment , can be seen .
入力文	言語は思想伝達の道具である。
正解文	Language is the vehicle of thought .
ベースライン	Language communicates tools .
+Wikipedia 文	Language is the transmission of the tool .

表 5.12: Wikipedia 文を追加した場合の対比較で “+ベースライン ” の出力例 (日英翻訳)

入力文	あの大学に入るには高い学力が必要だ。
正解文	Entering that university requires high scholastic ability .
ベースライン	That university will enter high in point of scholarship is necessary .
+Wikipedia 文	That university in the high point of scholarship is necessary .
入力文	その新しい奨学金は若い技術者に刺激となるだろう。
正解文	The new scholarship will provide a stimulus for young engineers .
ベースライン	The new scholarship will be stimulus to a young technology .
+Wikipedia 文	, and the new scholarship will respond to a young engineers .
入力文	この発言は群衆のかっさいを受けた。
正解文	This remark drew an ovation from the crowd .
ベースライン	This remark was received the applause of the crowd .
+Wikipedia 文	This remark the applause of the crowd .

表 5.13: Wikipedia 文を追加した場合の対比較で“差なし”の出力例 (日英翻訳)

入力文	彼女の余暇は読書でふさがっている。
正解文	Reading engages her spare time .
ベースライン	I am engaged in her spare time reading .
+Wikipedia 文	I am engaged by reading her leisure .
入力文	His suspension from tournament play will begin immediately .
正解文	彼のトーナメント 試合の 出場 停止 処分 は 即刻 始まる だろう。
ベースライン	He will be part of the match to stop the of the tournament will start without delay .
+Wikipedia 文	He will be part of the game to start the stories of the tournament .
入力文	He dropped away in a marathon race .
正解文	彼はマラソンで脱落した。
ベースライン	He made a hiatus in the marathon .
+Wikipedia 文	He took a hiatus in the marathon .

表 5.14: 辞書文を追加した場合の対比較で“+辞書文 ”の出力例 (英日翻訳)

入力文	That person can speak a little Japanese .
正解文	あの人は日本語がちょっとできます。
ベースライン	あの人は少しも話せる。
+辞書文	あの人は片言の日本語を話します。
入力文	This was a time of retreat for the Third World .
正解文	それは第三世界にとっては退却の時期だった。
ベースライン	これは時間をの 後退した。
+辞書文	これは第三世界の 後退した。
入力文	We can hardly settle things by theory alone .
正解文	物事はそう理論どおりにはいかない。
ベースライン	私たちはだけでは解決できない。
+辞書文	物事は理論だけでは解決できない。

表 5.15: 辞書文を追加した場合の対比較で“+ベースライン”の出力例 (英日翻訳)

入力文	Something must be done immediately to deal with this problem .
正解文	その問題には早急に手をうつ必要がある。
ベースライン	この問題はすぐにはどこかに処理しなくてはならない。
+辞書文	この問題の処理をしたがすぐにはならない。
入力文	This remark drew an ovation from the crowd .
正解文	この発言は群衆のっさいを受けた。
ベースライン	この発言は群衆から歓迎を引いた。
+辞書文	群衆からっさいがこの発言をした。
入力文	The car will not start .
正解文	自動車が動かない。
ベースライン	その車は出発しないだろう。
+辞書文	その車はしないだろう。

表 5.16: 辞書文を追加した場合の対比較で“+差なし”の出力例 (英日翻訳)

入力文	I am responsible for it .
正解文	その責めは私にある。
ベースライン	私はそれに責任がある。
+辞書文	それは私の責任だ。
入力文	I watched the game on television .
正解文	その試合はテレビで見ました。
ベースライン	テレビでその試合を見た。
+辞書文	私はその試合をテレビを見た。
入力文	Children under 12 are admitted free .
正解文	12歳に達しない子供は入場無料。
ベースライン	子供は無料で入学を12している。
+辞書文	子供は12人の入場を許可された。

表 5.17: 特許翻訳文を追加した場合の対比較で“+特許翻訳文”の出力例 (英日翻訳)

入力文	I deftly averted the blow .
正解文	その一撃を巧みにかわした。
ベースライン	私はその一発で乗り切ったの。
+特許翻訳文	私の一撃を背けた。
入力文	He has a keen intuition .
正解文	彼は鋭い勘をもっています。
ベースライン	彼は鋭い intuition ている。
+特許翻訳文	彼は鋭い intuition を持っている。
入力文	She made a good try of it .
正解文	それを思う存分やってみた。
ベースライン	彼女はよいに努めたのではない。
+特許翻訳文	彼女はいいトライした。

表 5.18: 特許翻訳文を追加した場合の対比較で“+ベースライン”の出力例 (英日翻訳)

入力文	We were totally crushed by the terrible news .
正解文	そのひどい知らせにすっかり打ちひしがれた。
ベースライン	私たちはその恐ろしいニュースですっかり打ちひしがれた。
+特許翻訳文	私たちは恐ろしいニュースでいた。
入力文	This cold medicine is effective .
正解文	この風邪薬はよく効く。
ベースライン	この風邪薬有効である。
+特許翻訳文	この冷たい薬は有効である。
入力文	The new scholarship will provide a stimulus for young engineers .
正解文	その新しい奨学金は若い技術者に刺激となるだろう。
ベースライン	若い技術者の新しい奨学金は刺激を与えてくれるだろう。
+特許翻訳文	技術者は新しい奨学金は刺激になるだろう。

表 5.19: 特許翻訳文を追加した場合の対比較で“差なし”の出力例 (英日翻訳)

入力文	The match went down to the wire .
正解文	試合は後半もつれた。
ベースライン	その試合は針金をに下がった。
+特許翻訳文	その試合は電線に下がった。
入力文	The appreciation of yen increased imports more than the growth of exports and a drop in foreign demand pushed down the economic growth .
正解文	円高で、輸出の伸び以上に輸入が増え、外需の落ち込みもマイナスに作用した。
ベースライン	需要は経済成長以上の輸入は一滴も、円で輸出の成長を外国鑑賞押し倒さた。
+特許翻訳文	経済成長は、増加した需要の生長を以上の輸入は鑑賞に押し倒さ円海外輸出している。
入力文	I am responsible for it .
正解文	その責めは私にある。
ベースライン	私はそれをの責任者となっている。
+特許翻訳文	私の責任者となっている。

表 5.20: Wikipedia 文を追加した場合の対比較で“+Wikipedia 文”の出力例 (英日翻訳)

+Wikipedia 文 の出力例	
入力文	The new scholarship will provide a stimulus for young engineers .
正解文	その新しい奨学金は若い技術者に刺激となるだろう。
ベースライン	その新しい奨学金の技術者に刺激を与えてくれるだろう。
+Wikipedia 文	その新しい奨学金は若い技術者の刺激を与えた。
入力文	This cold medicine is effective .
正解文	この風邪薬はよく効く。
ベースライン	この寒さには薬有効である。
+Wikipedia 文	この風邪薬効果がある。
入力文	This remark drew an ovation from the crowd .
正解文	この発言は群衆のかっさいを受けた。
ベースライン	この発言は人ごみの中から歓迎を引いた。
+Wikipedia 文	この発言は群衆から歓迎を引いた。

表 5.21: Wikipedia 文を追加した場合の対比較で “+ベースライン ” の出力例 (英日翻訳)

入力文	The car will not start .
正解文	自動車 が 動か ない。
ベースライン	その 車 は 出発 し ない だろ う。
+Wikipedia 文	その 車 は ない だろ う。
入力文	He chauffeured for a time right after the war .
正解文	彼 は 戦 争 直 後 し ば ら く の 間 お 抱 え 運 転 手 を し て い た 。
ベースライン	彼 は 戦 争 後 の 時 間 に は chauffeured た 。
+Wikipedia 文	彼 は 時 間 を chauffeured 。
入力文	She cut the apple in half .
正解文	彼 女 は リンゴ を 半 分 に 切 っ た 。
ベースライン	彼 女 は 半 分 に は リンゴ を 切 っ た 。
+Wikipedia 文	彼 女 に は リンゴ を 切 っ た 。

表 5.22: Wikipedia 文を追加した場合の対比較で “差なし” の出力例 (英日翻訳)

入力文	All my efforts came to nothing .
正解文	努 力 は 水 の あ わ と な っ た 。
ベースライン	せ っ か く の 私 の 努 力 は 何 に も な ら な か っ た 。
+Wikipedia 文	せ っ か く の 私 の 努 力 も ふ い に な っ た 。
入力文	Take part in a plan .
正解文	計 略 に 乗 る 。
ベースライン	計 画 を に し な さ い 。
+Wikipedia 文	出 場 を 計 画 し た 。
入力文	I gulped down a quick breakfast .
正解文	急 い で 朝 食 を 飲 み こ ん だ 。
ベースライン	私 は 朝 食 の 力 は を 一 気 に 飲 み 干 し た 。
+Wikipedia 文	私 は 朝 食 が 下 を 一 気 に 飲 み 干 し た 。

第6章 データ量を変化させた実験

6.1 目的

追加実験として、辞書文と Wikipedia 文を使用し、追加する文を最大 (FULL) のデータから $1/2$, $1/4$ と減らしていき、モノリンガルデータの量と翻訳精度の関係を調べる。

6.2 実験内容

実験環境は第4章と同じである。今回新たな実験データとして、追加するコーパスの $1/2$, $1/4$, $1/8$ のデータを使用し、辞書文と Wikipedia 文で日英翻訳と英日翻訳を行う。追加実験で使用するモノリンガルコーパスの内訳を表 6.1 と表 6.2 に示す。

表 6.1: 追加実験の日英翻訳で使用するモノリンガルコーパス

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
追加コーパス $1/8$	98,555 文	425,903 文	1,822,433 文
追加コーパス $1/4$	197,110 文	851,806 文	3,644,867 文
追加コーパス $1/2$	394,221 文	1,703,612 文	7,289,734 文
追加コーパス FULL	788,443 文	3,407,225 文	14,579,468 文

表 6.2: 追加実験の英日翻訳で使用するモノリンガルコーパス

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
追加コーパス $1/8$	100,790 文	425,903 文	1,575,938 文
追加コーパス $1/4$	201,581 文	851,807 文	3,151,877 文
追加コーパス $1/2$	403,162 文	1,703,615 文	6,303,754 文
追加コーパス FULL	806,324 文	3,407,231 文	12,607,508 文

6.3 実験結果

6.3.1 自動評価

日英翻訳の自動評価 BLEU の結果を表 6.3 と図 6.1 に，NIST の結果を表 6.4 と図 6.2 に，METEOR の結果を表 6.5 と図 6.3 に，RIBES の結果を表 6.6 と図 6.4 に，TER の結果を表 6.7 と図 6.5 に，英日翻訳の自動評価 BLEU の結果を表 6.8 と図 6.6 に，NIST の結果を表 6.9 と図??に，RIBES の結果を表 6.10 と図 6.8 に，TER の結果を表 6.11 と図 6.9 に示す．また，自動評価の特徴をまとめたものを表 6.12 に示す．

表 6.3: 日英翻訳の BLEU 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	0.1382	0.1299	0.1310
追加コーパス 1/8	0.1382	0.1358	0.1359
追加コーパス 1/4	0.1402	0.1353	0.1374
追加コーパス 1/2	0.1460	0.1350	0.1378
追加コーパス FULL	0.1499	0.1339	0.1389

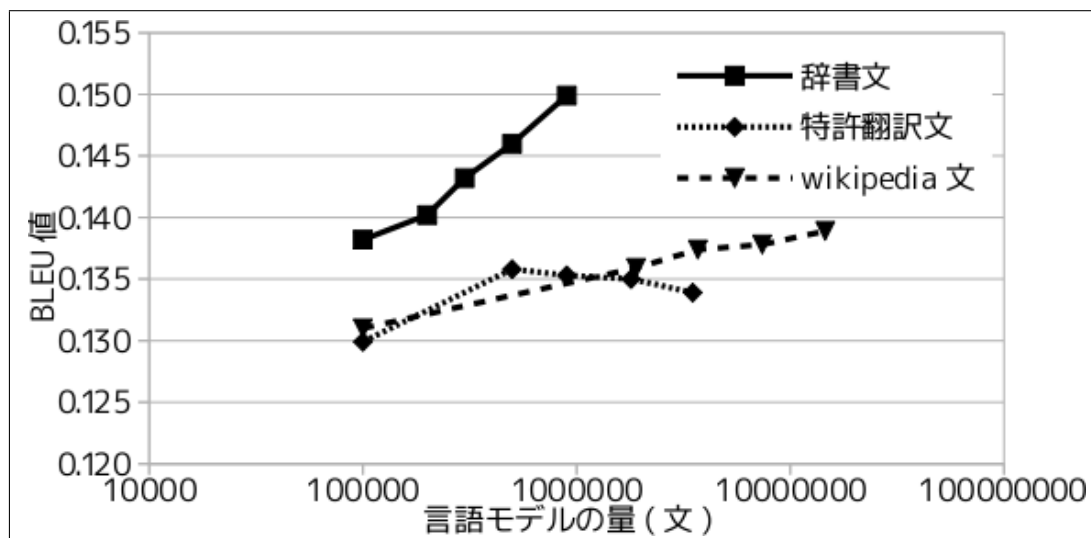


図 6.1: 日英翻訳の BLEU 値の変化

表 6.4: 日英翻訳の NIST 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	4.8771	4.6154	4.6886
追加コーパス 1/8	4.8893	4.7254	4.7639
追加コーパス 1/4	4.9144	4.7391	4.7922
追加コーパス 1/2	4.9281	4.7455	4.7959
追加コーパス FULL	4.9611	4.7449	4.8149

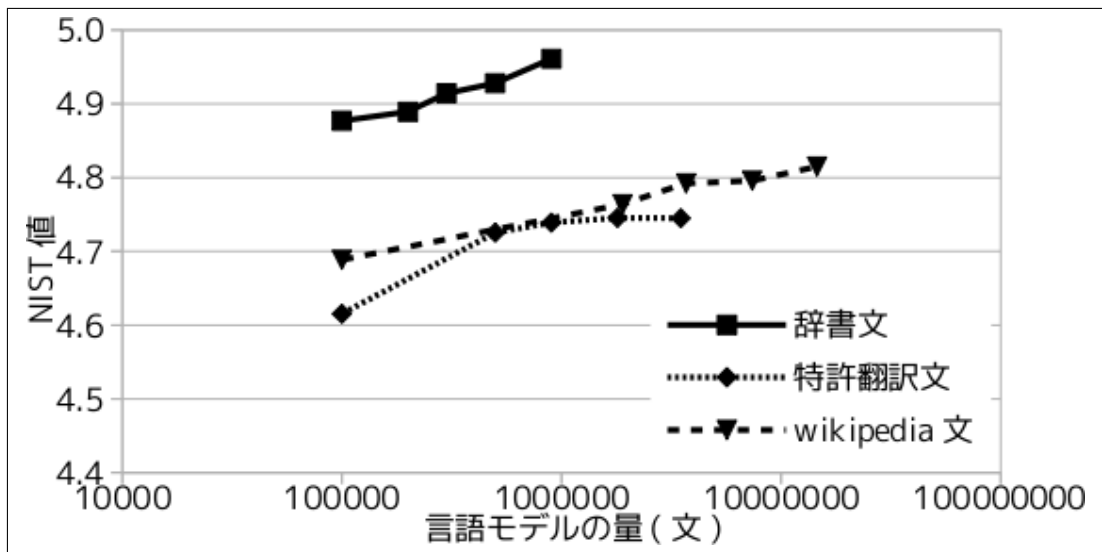


図 6.2: 日英翻訳の NIST 値の変化

表 6.5: 日英翻訳の METEOR 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	0.4552	0.4521	0.4540
追加コーパス 1/8	0.4566	0.4548	0.4532
追加コーパス 1/4	0.4591	0.4544	0.4541
追加コーパス 1/2	0.4609	0.4525	0.4539
追加コーパス FULL	0.4641	0.4512	0.4536

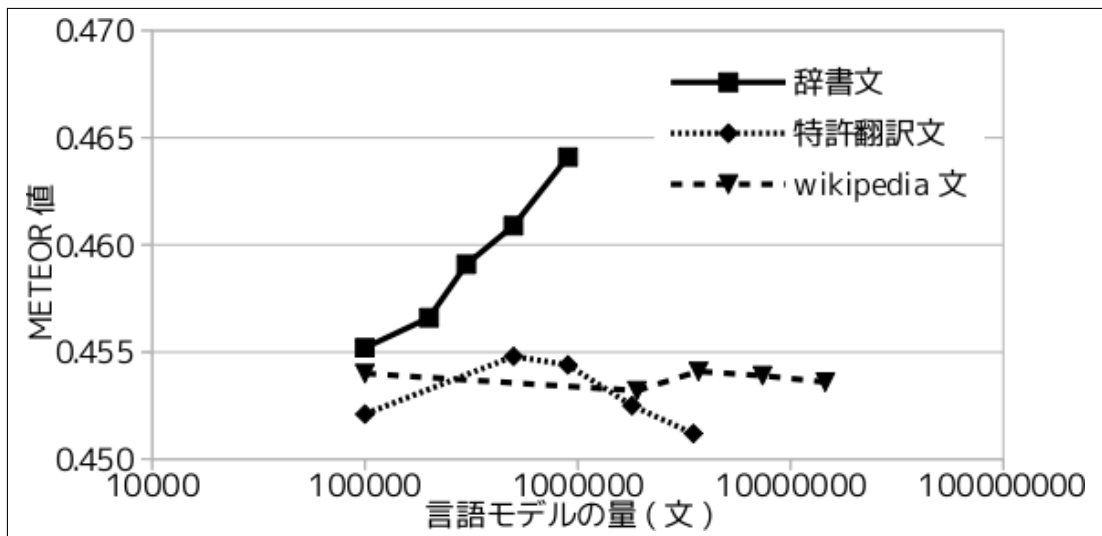


図 6.3: 日英翻訳の METEOR 値の変化

表 6.6: 日英翻訳の RIBES 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	0.7105	0.7051	0.7046
追加コーパス 1/8	0.7090	0.7075	0.7043
追加コーパス 1/4	0.7104	0.7064	0.7033
追加コーパス 1/2	0.7104	0.7065	0.7024
追加コーパス FULL	0.7123	0.7056	0.7024

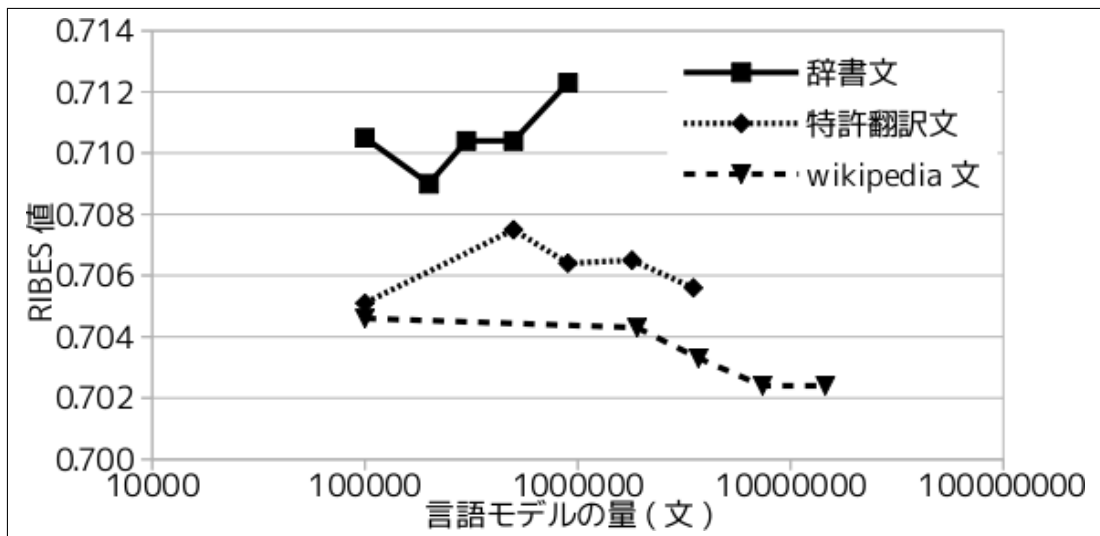


図 6.4: 日英翻訳の RIBES 値の変化

表 6.7: 日英翻訳の TER 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	69.511	73.437	72.719
追加コーパス 1/8	69.875	71.445	72.133
追加コーパス 1/4	69.973	71.189	71.991
追加コーパス 1/2	70.007	71.128	71.997
追加コーパス FULL	70.004	70.969	71.893

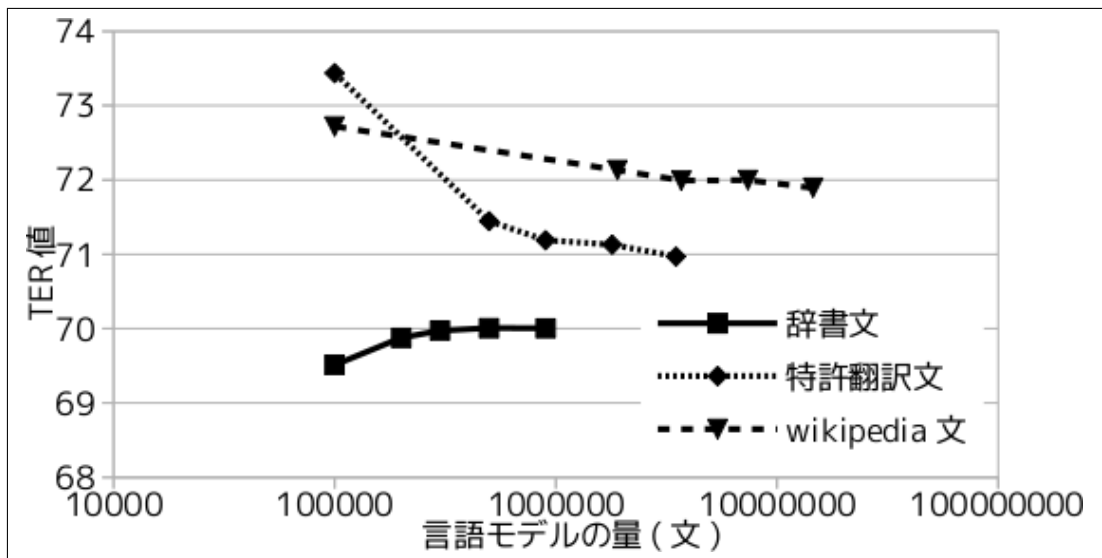


図 6.5: 日英翻訳の TER 値の変化

表 6.8: 英日翻訳の BLEU 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	0.1792	0.1738	0.1672
追加コーパス 1/8	0.1821	0.1748	0.1731
追加コーパス 1/4	0.1829	0.1753	0.1747
追加コーパス 1/2	0.1849	0.1747	0.1751
追加コーパス FULL	0.1860	0.1741	0.1753

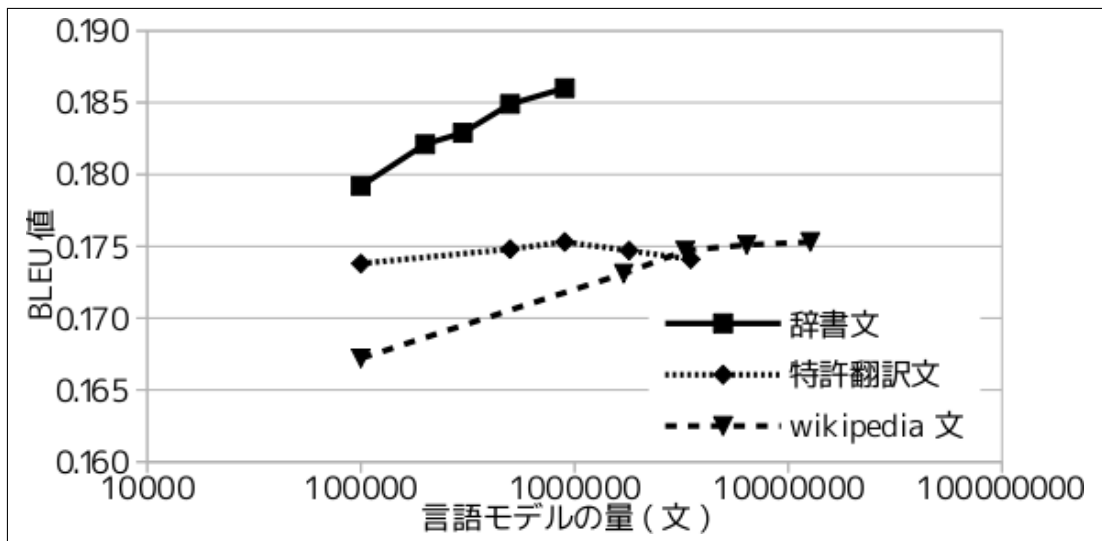


図 6.6: 英日翻訳の BLEU 値の変化

表 6.9: 英日翻訳の NIST 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	4.6187	4.5412	4.2875
追加コーパス 1/8	4.6410	4.5730	4.4056
追加コーパス 1/4	4.6493	4.5834	4.4370
追加コーパス 1/2	4.6733	4.5779	4.4531
追加コーパス FULL	4.6845	4.5835	4.4749

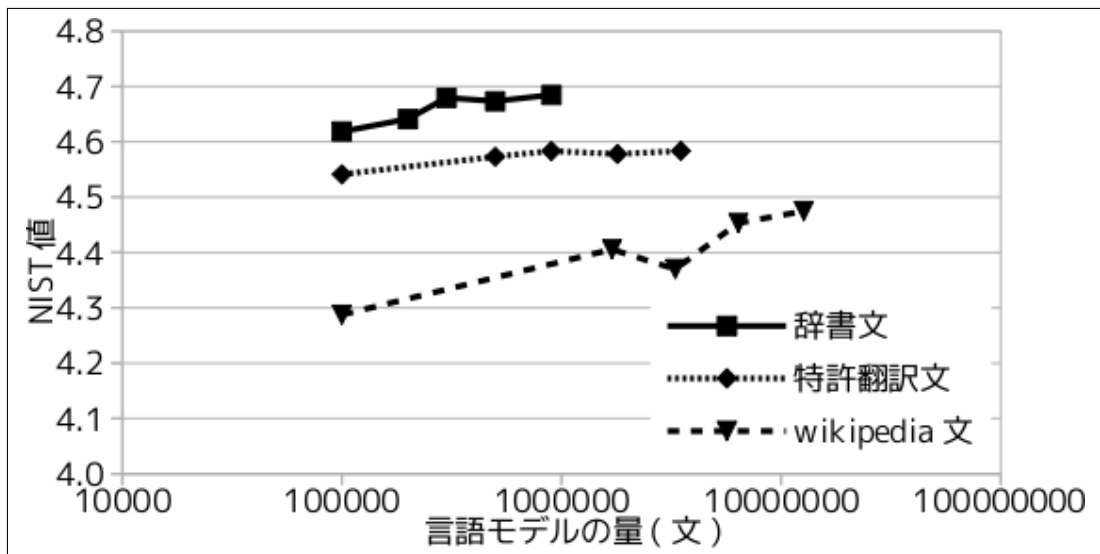


図 6.7: 英日翻訳の NIST 値の変化

表 6.10: 英日翻訳の RIBES 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	0.7686	0.7635	0.7628
追加コーパス 1/8	0.7684	0.7626	0.7624
追加コーパス 1/4	0.7681	0.7627	0.7625
追加コーパス 1/2	0.7674	0.7627	0.7623
追加コーパス FULL	0.7690	0.7622	0.7605

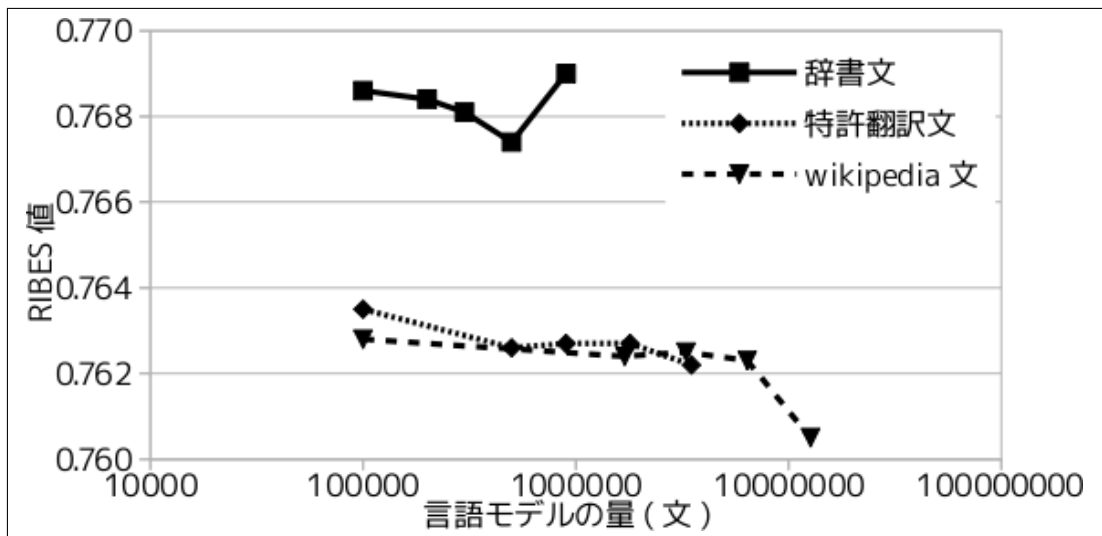


図 6.8: 英日翻訳の RIBES 値の変化

表 6.11: 英日翻訳の TER 値

コーパスの内容	辞書文	特許翻訳文	Wikipedia 文
ベースライン	65.958	68.097	71.641
追加コーパス 1/8	66.332	67.667	69.223
追加コーパス 1/4	66.471	67.544	68.793
追加コーパス 1/2	66.459	67.549	68.252
追加コーパス FULL	66.637	67.415	67.957

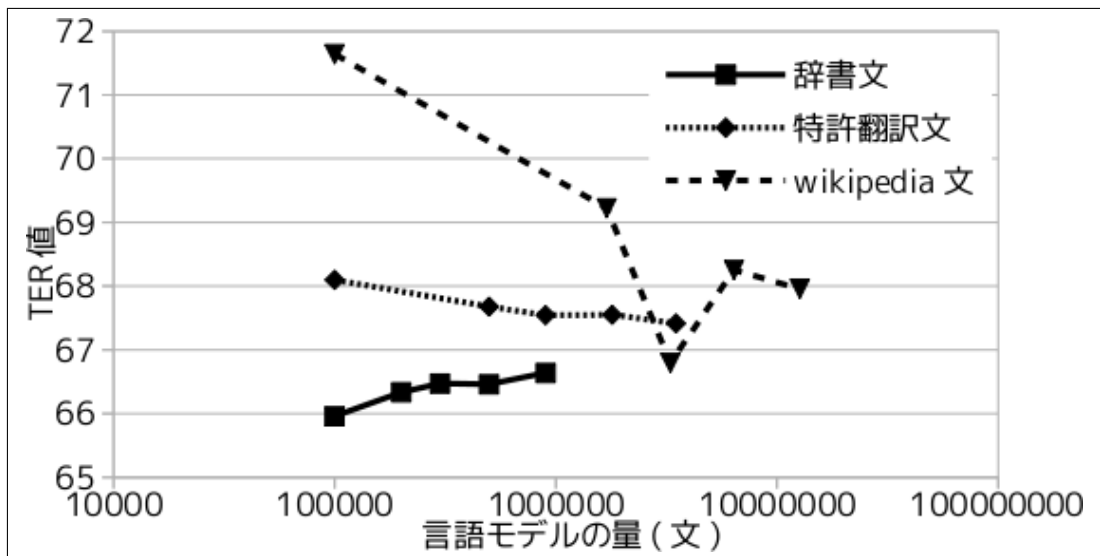


図 6.9: 英日翻訳の TER 値の変化

表 6.12: 自動評価の特徴

追加コーパスの種類	評価	特徴
辞書文	日英	BLEU 値と NIST 値と METEOR 値がデータ量と比例して向上
	英日	BLEU 値がデータ量と比例して向上
特許翻訳文	日英	BLEU 値, METEOR 値, RIBES 値が山なりの形
	英日	ほとんど変化なし
Wikipedia 文	日英	ほとんど変化なし
	英日	BLEU 値がデータ量と比例して向上

6.3.2 人手評価

人手評価における，辞書文の日英翻訳の結果を表 6.13 と図 6.10 に，特許翻訳文の日英翻訳の結果を表 6.14 と図 6.11 に，Wikipedia 文の日英翻訳の結果を表 6.15 と図 6.12 に，辞書文の英日翻訳の結果を表 6.16 と図 6.13 に，特許翻訳文の英日翻訳の結果を表 6.17 と図 6.14 に，Wikipedia 文の英日翻訳の結果を表 6.18 と図 6.15 に示す．また，人手評価の特徴をまとめたものを表 6.19 に示す．

表 6.13: 辞書文の日英翻訳の人手評価

ベースライン	追加コーパス 1/8	差なし	同出力
7	10	36	47
ベースライン	追加コーパス 1/4	差なし	同出力
7	12	46	35
ベースライン	追加コーパス 1/2	差なし	同出力
12	15	46	27
ベースライン	追加コーパス FULL	差なし	同出力
12	16	42	30

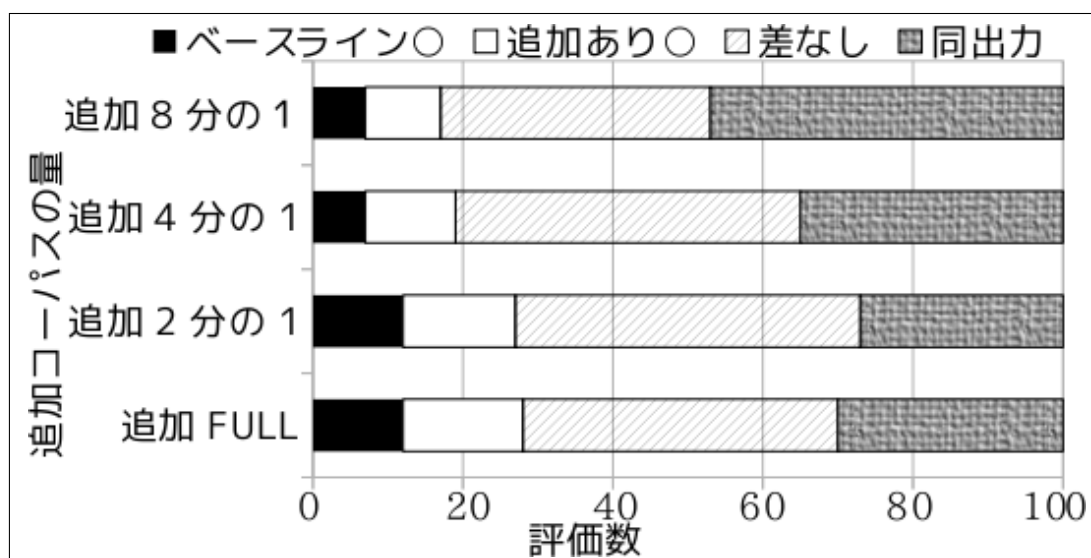


図 6.10: 辞書文の日英翻訳の人手評価の変化

表 6.14: 特許翻訳文の日英翻訳の人手評価

ベースライン	追加コーパス 1/8	差なし	同出力
11	8	25	56
ベースライン	追加コーパス 1/4	差なし	同出力
10	9	33	48
ベースライン	追加コーパス 1/2	差なし	同出力
10	9	36	45
ベースライン	追加コーパス FULL	差なし	同出力
9	11	35	45

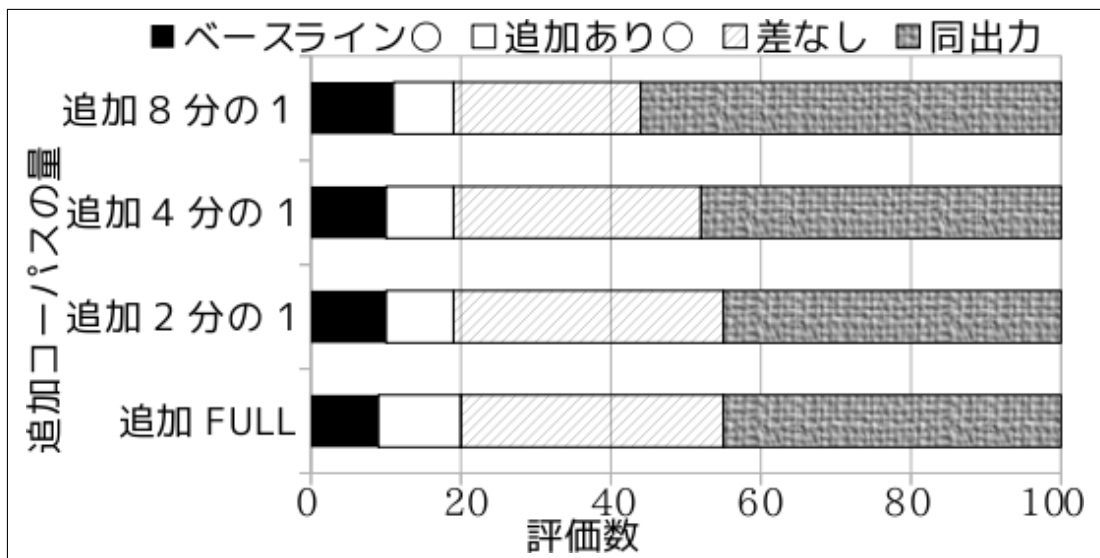


図 6.11: 特許翻訳文の日英翻訳の人手評価の変化

表 6.15: Wikipedia 文の日英翻訳の人手評価

ベースライン	追加コーパス 1/8	差なし	同出力
7	5	49	39
ベースライン	追加コーパス 1/4	差なし	同出力
8	7	49	36
ベースライン	追加コーパス 1/2	差なし	同出力
9	7	51	33
ベースライン	追加コーパス FULL	差なし	同出力
13	6	51	30

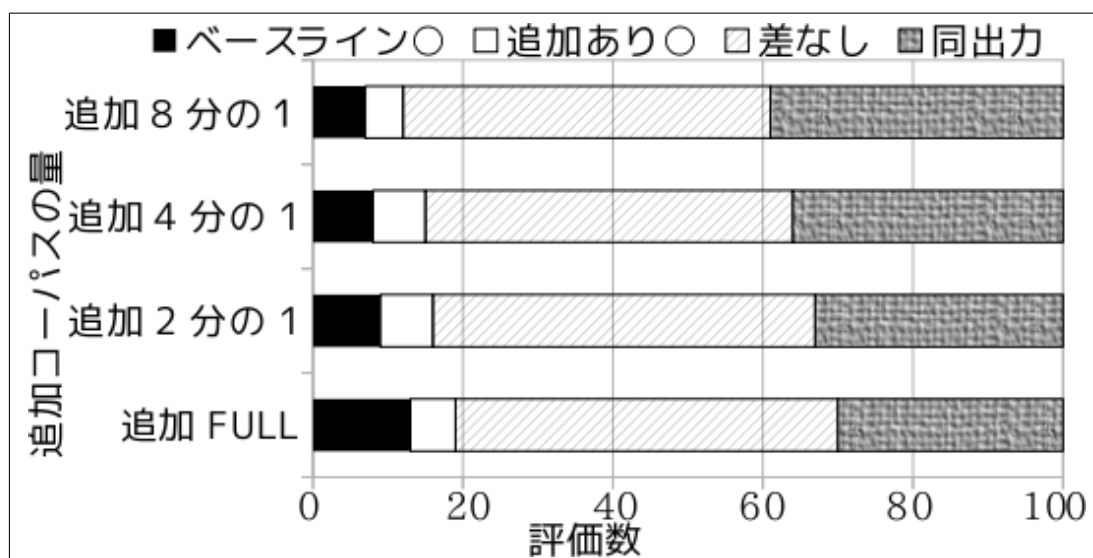


図 6.12: Wikipedia 文の日英翻訳の人手評価の変化

表 6.16: 辞書文の英日翻訳の人手評価

ベースライン	追加コーパス 1/8	差なし	同出力
5	5	54	36
ベースライン	追加コーパス 1/4	差なし	同出力
6	9	56	29
ベースライン	追加コーパス 1/2	差なし	同出力
5	11	58	26
ベースライン	追加コーパス FULL	差なし	同出力
7	14	58	21

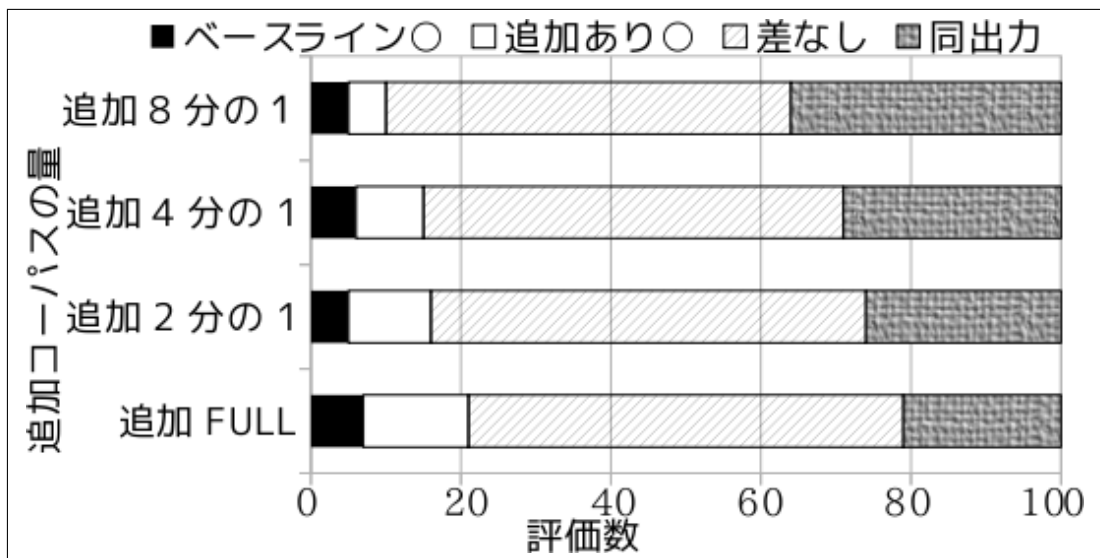


図 6.13: 辞書文の英日翻訳の人手評価の変化

表 6.17: 特許翻訳文の英日翻訳の人手評価

ベースライン	追加コーパス 1/8	差なし	同出力
3	4	51	42
ベースライン	追加コーパス 1/4	差なし	同出力
3	5	52	40
ベースライン	追加コーパス 1/2	差なし	同出力
4	8	54	34
ベースライン	追加コーパス FULL	差なし	同出力
6	8	54	32

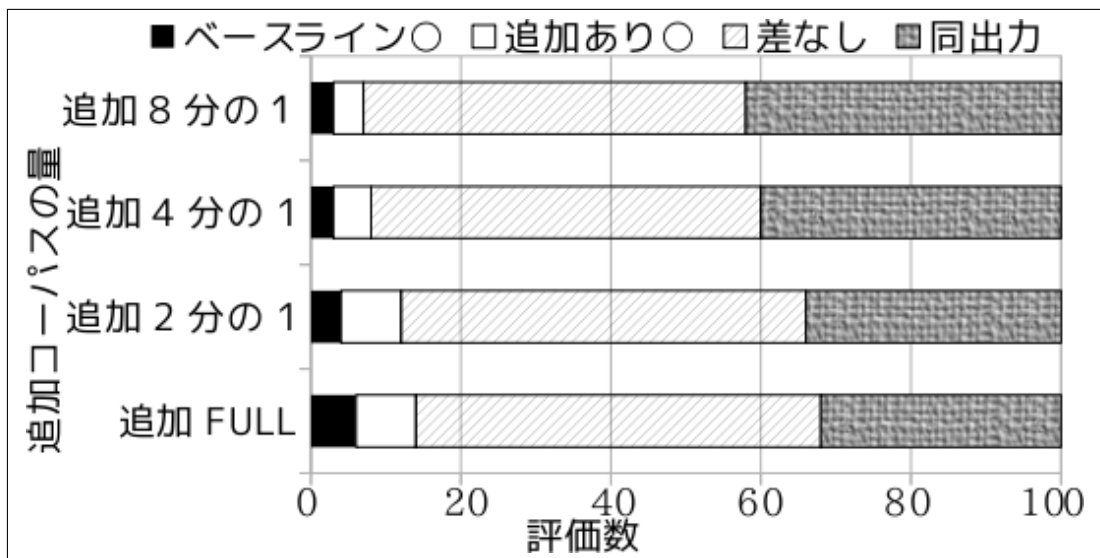


図 6.14: 特許翻訳文の英日翻訳の人手評価の変化

表 6.18: Wikipedia 文の英日翻訳の人手評価

ベースライン	追加コーパス 1/8	差なし	同出力
3	7	50	40
ベースライン	追加コーパス 1/4	差なし	同出力
6	9	52	33
ベースライン	追加コーパス 1/2	差なし	同出力
6	9	58	27
ベースライン	追加コーパス FULL	差なし	同出力
7	9	58	26

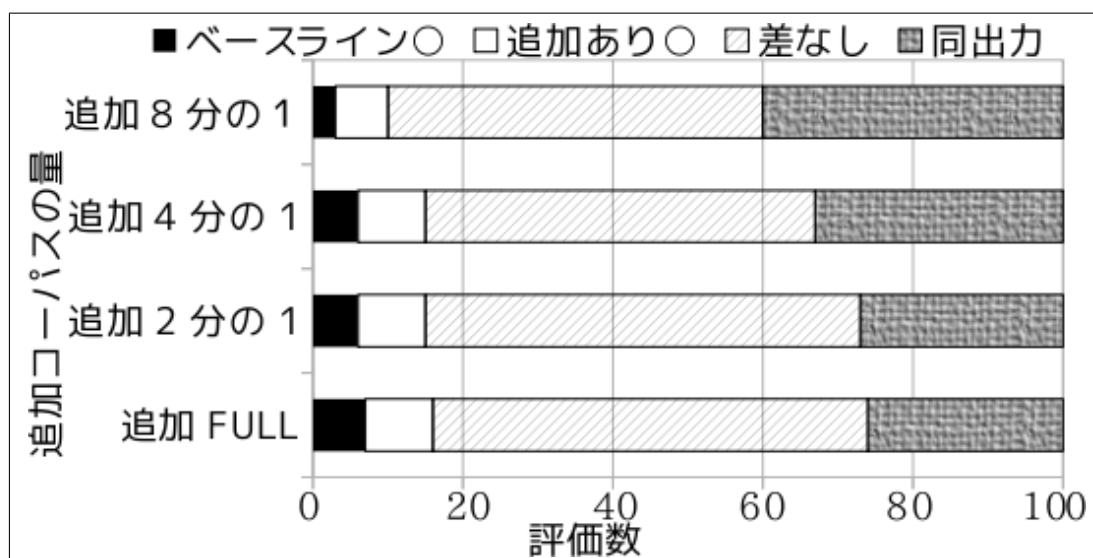


図 6.15: Wikipedia 文の英日翻訳の人手評価の変化

表 6.19: 人手評価の特徴

追加コーパスの種類	評価	特徴
辞書文	日英	追加 FULL \approx 追加 1/2 > 追加 1/4 \approx 追加 1/8
	英日	データ量と比例して向上
特許翻訳文	日英	変化なし
	英日	ほとんど変化なし
Wikipedia 文	日英	追加 FULL は悪く, それ以外は変化なし
	英日	ほとんど変化なし

6.3.3 実験結果のまとめ

自動評価の結果をまとめた表 6.12 と人手評価の結果をまとめた表 6.19 のそれぞれの結果を 1 つにまとめたものを，表 6.20 に示す．

表 6.20: データ量を変化させる実験のまとめ

追加コーパスの種類	評価	総評
辞書文	日英	データ量と比例して向上
	英日	データ量と比例して向上
特許翻訳文	日英	変化なし
	英日	変化なし
Wikipedia 文	日英	変化なし
	英日	変化なし

表 6.20 より，テスト文と同分野の辞書文を追加する量を変化させると，日英翻訳・英日翻訳ともに翻訳精度がデータ量と比例して向上していることがわかる．一方，テスト文と別分野の特許翻訳文と Wikipedia 文を追加する量を変化させると，日英翻訳・英日翻訳ともに翻訳精度に変化がないことがわかる．この結果から，テスト文と同分野のモノリンガルデータ増加させると，データ量が増えるごとに，翻訳精度が向上することが推測できる．

6.3.4 出力例

日英翻訳における辞書文を追加した場合の出力例を表 6.21，対比較の基準を表 6.22 に，辞書文を追加した場合の出力例を表 6.23，対比較の基準を表 6.24 に，Wikipedia 文を追加した場合の出力例を表 6.25，対比較の基準を表 6.26 に示す．英日翻訳における辞書文を追加した場合の出力例を表 6.27，対比較の基準を表 6.28 に，辞書文を追加した場合の出力例を表 6.29，対比較の基準を表 6.30 に，Wikipedia 文を追加した場合の出力例を表 6.31，対比較の基準を表 6.32 に示す．

また，対比較の判断基準の表において，“追加コーパス FULL ”を“追加 FULL ”のように省略する．“追加コーパス 1/2 ”，“追加コーパス 1/4 ”，“追加コーパス 1/8 ”も同様に，“追加 1/2 ”，“追加 1/4 ”，“追加 1/8 ”と省略する．

表 6.21: 追加する辞書文のデータ量を変化させた場合の出力例 (日英翻訳)

出力例 1	
入力文	彼は頭が天井に着くほど背が高い。
正解文	He is so tall that his head touches the ceiling .
ベースライン	His brain as tall reaches the ceiling .
追加コーパス 1/8	He is so tall will reach the ceiling .
追加コーパス 1/4	He is so tall will reach the ceiling .
追加コーパス 1/2	He is so tall will reach the ceiling .
追加コーパス FULL	He is so tall that his head will reach the ceiling .
出力例 2	
入力文	全員 この知らせにかなりがっかりしてしまった。
正解文	Everyone grew rather depressed at these tidings .
ベースライン	Don't be discouraged all in this news .
追加コーパス 1/8	Don't be discouraged all in this news .
追加コーパス 1/4	Don't be discouraged all in this news .
追加コーパス 1/2	Don't be discouraged us all in this news .
追加コーパス FULL	Don't be discouraged all in this news .
出力例 3	
入力文	彼はわがチームの主将としてふさわしい。
正解文	He is working to be captain of our team .
ベースライン	He is suitable for a captain of our team .
追加コーパス 1/8	He befitting as captain of our team .
追加コーパス 1/4	He befitting as captain of our team .
追加コーパス 1/2	He is a suitable for a captain of our team .
追加コーパス FULL	He is a suitable as captain of our team .

表 6.22: 表 6.21 の対比較の判断基準

追加コーパス量	対比較判断基準		
	出力例 1	出力例 2	出力例 3
追加コーパス 1/8	追加 1/8	同出力	追加 1/8
追加コーパス 1/4	追加 1/4	同出力	追加 1/4
追加コーパス 1/2	追加 1/2	差なし	差なし
追加コーパス FULL	追加 FULL	同出力	追加 FULL

表 6.23: 追加する特許翻訳文のデータ量を変化させた場合の出力例 (日英翻訳)

出力例 1	
入力文	このおかげで私はいっそう思いどおり選択できる。
正解文	This will afford me wider freedom of choice .
ベースライン	This saved me with a confirmed can be choice .
追加コーパス 1/8	This help I can be any with confirmed .
追加コーパス 1/4	This help I can be any with confirmed .
追加コーパス 1/2	This help I can be any confirmed as desired .
追加コーパス FULL	This help I can be confirmed with choice .
出力例 2	
入力文	何物も彼を満足させなかった。
正解文	Nothing contented him .
ベースライン	I did not satisfy him as many things .
追加コーパス 1/8	He did not satisfy the several drawings .
追加コーパス 1/4	He did not satisfy the several drawings .
追加コーパス 1/2	He did not satisfy the several drawings .
追加コーパス FULL	Several things can also did not satisfy him .
出力例 3	
入力文	正しい郵便料金を貼ってください。
正解文	Please affix the correct postage .
ベースライン	Please put the fares correct mail .
追加コーパス 1/8	Please put the fares correct mail .
追加コーパス 1/4	Please put a proper postal rates .
追加コーパス 1/2	Please put a proper postal rates .
追加コーパス FULL	Please put a proper postal rates .

表 6.24: 表 6.23 の対比較の判断基準

追加コーパス量	対比較判断基準		
	出力例 1	出力例 2	出力例 3
追加コーパス 1/8	ベースライン	差なし	同出力
追加コーパス 1/4	ベースライン	差なし	追加 1/4
追加コーパス 1/2	差なし	差なし	追加 1/2
追加コーパス FULL	差なし	追加 FULL	追加 FULL

表 6.25: 追加する Wikipedia 文のデータ量を変化させた場合の出力例 (日英翻訳)

出力例 1	
入力文	言語は思想伝達の道具である。
正解文	Language is the vehicle of thought .
ベースライン	Language communicates tools .
追加コーパス 1/8	Language is of distributing tools .
追加コーパス 1/4	Language is the transmission of the tool .
追加コーパス 1/2	Language is the transmission of the tool .
追加コーパス FULL	Language is the transmission of the tool .
出力例 2	
入力文	それを思う存分やってみた。
正解文	She made a good try of it .
ベースライン	I tried to my heart's content it .
追加コーパス 1/8	I tried to my heart's content it .
追加コーパス 1/4	I tried to my heart's content it .
追加コーパス 1/2	I tried every minute of it .
追加コーパス FULL	I tried every minute of it .
出力例 3	
入力文	言語は思想伝達の道具である。
正解文	Language is the vehicle of thought .
ベースライン	Language communicates tools .
追加コーパス 1/8	Language is of distributing tools .
追加コーパス 1/4	Language is the transmission of the tool .
追加コーパス 1/2	Language is the transmission of the tool .
追加コーパス FULL	Language is the transmission of the tool .

表 6.26: 表 6.25 の対比較の判断基準

追加コーパス量	対比較判断基準		
	出力例 1	出力例 2	出力例 3
追加コーパス 1/8	差なし	同出力	差なし
追加コーパス 1/4	追加コーパス 1/4	同出力	追加 1/4
追加コーパス 1/2	追加コーパス 1/2	ベースライン	追加 1/2
追加コーパス FULL	追加コーパス FULL	ベースライン	追加 FULL

表 6.27: 追加する辞書文のデータ量を変化させた場合の出力例 (英日翻訳)

出力例 1	
入力文	We can hardly settle things by theory alone .
正解文	物事はそう理論どおりにはいかない。
ベースライン	私たちはだけでは解決できない。
追加コーパス 1/8	私たちだけでは解決できない。
追加コーパス 1/4	だけでは解決できない。
追加コーパス 1/2	私たちの理論だけで解決できない。
追加コーパス FULL	物事は理論だけでは解決できない。
出力例 2	
入力文	He chauffeured for a time right after the war .
正解文	彼は戦争直後しばらくの間お抱え運転手をして いた。
ベースライン	彼は chauffeured 時間後にした。
追加コーパス 1/8	彼はその戦争後の chauffeured だ。
追加コーパス 1/4	彼は戦争後の一時 chauffeured 。
追加コーパス 1/2	彼は戦争直後 chauffeured だ。
追加コーパス FULL	彼は chauffeured 時間後にした。
出力例 3	
入力文	That person can speak a little Japanese .
正解文	あの人は日本語がちょっとできます。
ベースライン	His brain as tall reaches the ceiling .
追加コーパス 1/8	あの人は片言の日本語を話せる。
追加コーパス 1/4	あの人は片言の日本語を話せる。
追加コーパス 1/2	あの人は少し日本語が話せる。
追加コーパス FULL	あの人は片言の日本語を話します。

表 6.28: 表 6.27 の対比較の判断基準

追加コーパス量	対比較判断基準		
	出力例 1	出力例 2	出力例 3
追加コーパス 1/8	差なし	差なし	追加 1/8
追加コーパス 1/4	差なし	追加 1/4	追加 1/4
追加コーパス 1/2	追加 1/2	追加 1/2	追加 1/2
追加コーパス FULL	追加 FULL	同出力	追加 FULL

表 6.29: 追加する特許翻訳文のデータ量を変化させた場合の出力例 (英日翻訳)

出力例 1	
入力文	This cold medicine is effective .
正解文	この風邪薬はよく効く。
ベースライン	この風邪薬有効である。
追加コーパス 1/8	この風邪薬は有効である。
追加コーパス 1/4	この風邪薬は有効である。
追加コーパス 1/2	この風邪薬は有効である。
追加コーパス FULL	この冷たい薬は有効である。
出力例 2	
入力文	The baby was fingering her mother's necklace .
正解文	その赤ん坊は母親のネックレスをいじっていた。
ベースライン	赤ん坊が彼女の母ネックレス fingering れた。
追加コーパス 1/8	赤ん坊が彼女の母ネックレス fingering た。
追加コーパス 1/4	赤ん坊が彼女の母ネックレス fingering た。
追加コーパス 1/2	赤ん坊が彼女の母ネックレス fingering た。
追加コーパス FULL	赤ん坊は母親のネックレス fingering た。
出力例 3	
入力文	We were totally crushed by the terrible news .
正解文	そのひどい知らせにすっかり打ちひしがれた。
ベースライン	私たちはその恐ろしいニュースですっかり打ちひしがれた。
追加コーパス 1/8	私たちはその恐ろしいニュースですっかり打ちひしがれた。
追加コーパス 1/4	私たちはその恐ろしいニュースですっかり打ちひしがれた。
追加コーパス 1/2	私たちは恐ろしいニュースでいた。
追加コーパス FULL	私たちは恐ろしいニュースでいた。

表 6.30: 表 6.29 の対比較の判断基準

追加コーパス量	対比較判断基準		
	出力例 1	出力例 2	出力例 3
追加コーパス 1/8	追加 1/8	差なし	同出力
追加コーパス 1/4	追加 1/4	差なし	同出力
追加コーパス 1/2	追加 1/2	差なし	ベースライン
追加コーパス FULL	ベースライン	追加 FULL	ベースライン

表 6.31: 追加する Wikipedia 文のデータ量を変化させた場合の出力例 (英日翻訳)

出力例 1	
入力文	She is always at the top of her class .
正解文	彼女はいつもクラスの首席を占めている。
ベースライン	彼女はクラスの上ではいつもている。
追加コーパス 1/8	彼女はいつも彼女の上にあります。
追加コーパス 1/4	彼女はクラスのトップである。
追加コーパス 1/2	彼女はクラスのトップである。
追加コーパス FULL	彼女はクラスのトップである。
出力例 2	
入力文	Entering that university requires high scholastic ability .
正解文	あの大学に入るには高い学力が必要だ。
ベースライン	その大学 Entering scholastic 能力を必要とする。
追加コーパス 1/8	その大学 Entering scholastic 能力を必要とする。
追加コーパス 1/4	その大学 Entering scholastic 能力が高い。
追加コーパス 1/2	その大学 Entering scholastic 能力が高い。
追加コーパス FULL	その大学 Entering scholastic 能力が高い。
出力例 3	
入力文	I gulped down a quick breakfast .
正解文	急いで朝食を飲みこんだ。
ベースライン	私は朝食の力はを一気に飲み干した。
追加コーパス 1/8	私はすぐ朝食を一気に飲み干した。
追加コーパス 1/4	私は朝食が下を一気に飲み干した。
追加コーパス 1/2	私は朝食が下を一気に飲み干した。
追加コーパス FULL	私は朝食が下を一気に飲み干した。

表 6.32: 表 6.31 の対比較の判断基準

追加コーパス量	対比較判断基準		
	出力例 1	出力例 2	出力例 3
追加コーパス 1/8	差なし	同出力	追加 1/8
追加コーパス 1/4	追加 1/4	ベースライン	差なし
追加コーパス 1/2	追加 1/2	ベースライン	差なし
追加コーパス FULL	追加 FULL	ベースライン	差なし

第7章 考察

7.1 分野の依存性

今回の実験では、同分野のモノリンガルデータを増やすと日英・英日翻訳で翻訳精度は向上するが、別分野のモノリンガルデータを追加すると、日英翻訳では翻訳精度が上がらないことがわかった。よって、言語モデルが統計翻訳に与える分野依存性は高く、別分野による影響がないことがわかる。

7.2 データ量と翻訳精度の関係

第5章の結果より、辞書文のデータを増やすと人手・自動評価において日英・英日翻訳の翻訳精度が、増加量と比例して向上しているので、テスト文と同分野のデータ量を増やす実験は有用であると考えられる。また、この結果が分野の依存性と関連していることが予測できる。

7.3 評価の問題点

今回の実験での人手評価は、出力文100文に対して対比較を行ったが、翻訳品質に差のある文は20文ほどしかなかった。20文で翻訳精度の差を評価するには信頼性が低い。よって評価する文数を増やす必要がある。具体的には差のある文章が100文程度必要なので、人手評価に用いるデータ量は500文が望ましい。

統計翻訳における自動評価では、様々な問題が報告されている。松本は、単文を用いて自動評価と人手評価を行い、評価結果に差異が生じたことを報告している[15]。本実験でも自動評価の結果と人手評価の結果に差があった。今後、自動評価に対する方法を調べる必要がある。

7.4 関連研究との比較

Brantsらはアラビア語英語間の統計翻訳において、1300万~2兆トークンもの多量のモノリンガルデータを使用し、統計翻訳を行った[16]。4種類のモノリンガルデータ、2つの翻訳手法を用いて、BLEU値が分野に関係なく向上することを報告した。また、Schwenkは仏英翻訳で、3億単語の多量のモノリンガルデータを使用し、統計翻訳を行い、BLEU値が上昇したことを報告している[17]。

本実験でも日英翻訳・英日翻訳でのBLEU値が同分野と別分野で向上している。しかし、他の自動評価の値はばらつきがある。上記の関連研究は、BLEU値が向上したことのみを報告しており、他の自動評価の値は示されない。これは、BLEUの評価の問題点の1つであると考えている。

7.5 誤り解析

分野ごとの属性を調べるために、各出力文に対して誤り解析を行った。解析の結果、分野ごとの特徴が出た出力例を表7.1に示す。

表 7.1: 英日翻訳における対比較の出力例

入力文	The noise wakened him .
正解文	その音で彼は目を覚ました。
	+辞書文の出力例
ベースライン	彼はその音目覚めた。
+辞書文	その物音で彼は目覚めた。
	+特許翻訳文の出力例
ベースライン	彼はその音目覚めた。
+特許翻訳文	彼はその音目覚めた。
	+Wikipedia文の出力例
ベースライン	彼はその音目覚めた。
+Wikipedia文	騒音が目覚めた。

テスト文と同分野の辞書文を追加した場合、ベースラインと比較して文法が良くなっている。この結果より、翻訳精度が向上したことがわかる。また、テスト文と別分野の特許翻訳文を追加した場合、ベースラインと同じ出力結果となっている。この結果より、翻訳精度に変化がなかったことがわかる。一方、テスト文と別分野のWikipedia文を追

加した場合、ベースラインと比較して、出力文に“彼”が抜けている。この結果より、翻訳精度が悪くなったことがわかる。

この結果より、辞書文を追加する実験では、同分野の学習データが増えたため、翻訳精度が向上したと考えられる。翻訳精度が悪くなってしまった Wikipedia 文では、原データの特性として“彼”のような人称代名詞をあまり用いないため、多量のモノリンガルデータにより、“彼”の翻訳確率が下がってしまったと考えられる。特許翻訳文は同様に、人称代名詞をあまり用いない特性を持つが、Wikipedia 文と比較して、モノリンガルデータの量が少なかったため、“彼”の翻訳確率にあまり影響が出なかったと考えられる。

この問題を解決するために、分野が言語モデルに与える影響を調整するため、各モノリンガルデータに言語アダプテーションを用いる手法が考えられる。

第8章 おわりに

本研究では，統計翻訳において，データの収集が安易なモノリンガルデータを多量に用いて，言語モデルを学習し，翻訳精度を調査した．使用するモノリンガルデータとして，辞書文，特許翻訳文，Wikipedia 文を扱い，日英翻訳，英日翻訳をそれぞれ行った．テスト文には辞書文をそれぞれ用いて，合計6種類の実験を行った．

実験の結果，テスト文と同分野のデータを増やすと，日英翻訳と英日翻訳の翻訳精度が向上した．一方，テスト文と別分野のデータを増やすと，日英翻訳と英日翻訳の翻訳精度はあまり変化しなかった．これにより，分野依存性が高いことを示すことができた．

また，追加するモノリンガルデータの量を変化させて，翻訳精度を調査する実験を行った．結果として，テスト文と同分野の辞書文を追加する実験では，データ量が増加すると，それに比例して翻訳精度が向上することがわかった．

今後は重文複文コーパスや，統計翻訳の別手法での実験を行いたい．

謝辞

最後に，1年間に渡り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工学講座C研究室の村上仁一准教授，徳久雅人講師，村田真樹教授そして計算機工学講座C研究室の方々に厚く御礼申し上げます．

また，参考にさせていただいた論文の著者の方々に対して深く感謝申し上げます．

参考文献

- [1] Li et al.: “Parallel Aligned Treebanks at LDC: New Challenges Interfacing Existing Infrastructures”, LREC 2012, pp.1848-1855, 2012.
- [2] Ruiqiang Zhang and Eiichiro Sumita: “Boosting Statistical Machine Translation by Lemmatization and Linear Interpolation”, Proceedings of the ACL 2007, pp.181-184, 2007.
- [3] 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. 2012.
- [4] 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.
- [5] GIZA++ : <http://www.fjoch.com/GIZA++>
- [6] Moses: Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondrej Bojar, 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.
- [7] BLEU: Papineni Kishore, Salim Roukos, Todd Ward, Wei-Jing Zhu, “BLEU: a method for automatic evaluation of machine translation”, 40th Annual meeting of the Association for Computational Linguistics, pp.311-318, 2002.
- [8] METEOR: Banerjee Satanjeev, Lavie Alon, “METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments”, Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005), pp.65-72, 2005.

- [9] RIBES: 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年次大会発表論文集, pp.1111-
1114, 2011.
- [10] Richard Schwartz, Linnea Micciulla, John Makhoul. “A Study of Translation Edit
Rate with Targeted Human Annotation”, AMTA, pp.223-231, 2006.
- [11] SRILM: Andreas Stolcke, “SRILM - an Extensible Language Modeling Toolkit”,
7th International Conference on Spoken Language Processing, pp.901-904, 2002.
- [12] Mert: Franz Josef Och: “Minimum Error Rate Training in Statistical Machine
Translation”, In Proceeding of the 41st Annual Meeting of the Association for Com-
putational Linguistics, pp.160-167, 2003.
- [13] MeCab: Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto, “Applying Conditional
Random Fields to Japanese Morphological Analysis”, Proceedings of the 2004 Con-
ference on Empirical Methods in Natural Language Processing (EMNLP-2004),
pp.230-237, 2004.
- [14] tokenizer.sed
<http://www.cis.upenn.edu/treebank/tokenizer.sed>
- [15] 松本拓也, 村上仁一, 徳久雅人 “機械翻訳における人手評価と自動評価の考察”, NLP-
2012 , pp.505-508, 2012
- [16] Thorsten Brants, Ashok C. Popat, Peng Xu, Franz J. Och, and Jeffrey Dean, “Large
Language Models in Machine Translation”, EMNLP-2007, pp.858-867, 2007.
- [17] Holger Schwenk, “Investigations on Large-Scale Lightly-Supervised Training for Sta-
tistical Machine Translation”, Proceedings of IWSLT 2008, pp.182-189, 2008.