

統計翻訳における対訳パターン追加の効果

池淵堅斗 *1 村上仁一 *1 坂田純 *1 坂口暎一 *2 徳久雅人 *1

*1 鳥取大学大学院 工学研究科 情報エレクトロニクス専攻

*2 鳥取大学 工学部 知能情報工学科

*1,*2 {s102002, murakami, d112004, s102018, tokuhisa} @ ike.tottori-u.ac.jp

1 はじめに

現在、機械翻訳の分野において、統計翻訳の研究が注目されている。統計翻訳において、対訳文が多ければ多いほど翻訳精度は高くなる。しかし、利用できる対訳文には限りがある。

そこで、Popovic Maja らは、セルビア語英語間の翻訳において、対訳文に対訳句を追加し統計翻訳を行った。その結果、翻訳精度が向上した [1]。また、日野らは同様の手法で、対訳句として鳥バンク [2] と英辞郎 [3] を用いて、日本語英語間の統計翻訳を行った。その結果、日本語英語間においても翻訳精度の向上が確認できた [4]。

ところで、文は句とパターンで構成されている。よって、パターンを追加した場合でも翻訳精度が向上すると考えられる。そこで、本研究では日本語英語間の統計翻訳において対訳文に対訳パターンを追加することによる翻訳精度の変化を調査する。

2 提案手法

本実験の提案手法として、翻訳モデル作成の際の学習データとして、対訳文に対訳パターンを追加する。尚、言語モデル作成の際の学習データは、対訳文のみを用いる。日英統計翻訳の場合の提案手法の手順を以下に示す。

- 手順 1: 日英対訳文の英語文を学習データとして言語モデルを作成する
- 手順 2: 日英対訳文に日英対訳パターンを追加する
- 手順 3: 手順 2 で作成した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 4: 手順 1 と手順 3 で作成したモデルを用いて統計翻訳を行う

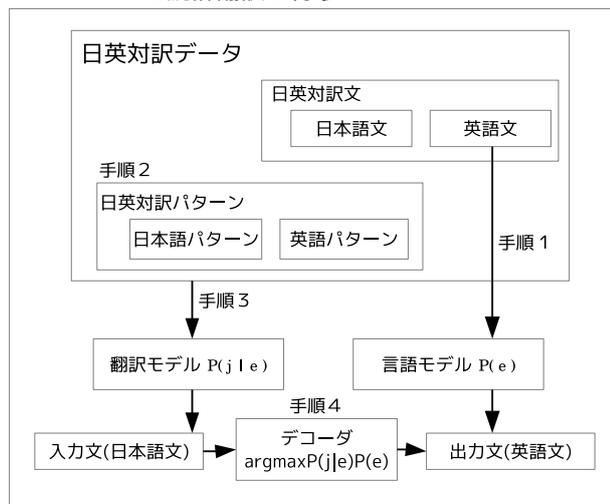


図 1 日英統計翻訳での提案手法の流れ

3 実験環境

3.1 前処理

本研究では、統計翻訳の前処理として、日本語文と日本語パターンに対して “MeCab[5]” を用いて形態素解析を行う。また、英語文と英語パターンに対して “tokenizer.perl[6]” を用いて分かち書きを行う。

3.2 日英対訳文

本研究では、日英対訳文として辞書の例文から抽出した単文コーパス [7] と重文複文コーパス [7] を表 1 の内訳で用いる。また、単文の例を表 2、重文複文の例を表 3 に示す。

表 1 日英対訳文数の内訳

日英対訳文	単文 (文)	重文複文 (文)
学習データ	100,000	100,000
ディベロップメントデータ	1,000	1,000
テストデータ	10,000	10,000

表 2 単文の例

彼の考え方は極端すぎる。
His way of thinking goes too far.

表 3 重文複文の例

その知らせを聞いて彼女の顔は明るくなった。
Her face lit up when she heard the news.

3.3 日英対訳パターン

本研究で用いる日英対訳パターンは人手で作成した対訳句 [2] を用いて自動的に作成する。パターン化を行う手法として、変数化可能な単語を全て変数化し、変数番号を含めた変数を全て N に置き換える。また、変数化可能な単語とは、日英対訳パターン原文に対して、人手で作成した対訳句を照合して、マッチした単語のみを変数化している。また今回の実験では、日英対訳パターン原文 1 文に対してパターンが複数マッチングする可能性があるが、この場合は最初にマッチしたパターンを用いることにする。ゆえに、最適なパターンが得られるとは限らない。尚、日英対訳パターン原文には、それぞれの実験で学習データとして用いる対訳文 10 万文を用いる。

単文の日英対訳パターンの例を表 4、重文複文の日英対訳パターンの例を表 5 に示す。

表 4 単文の日英対訳パターンの例

日本語	英語
ナイロンは原油より作られる。 N は N より N れる。	Nylon is made from crude oil. N is N from N .

表5 重文複文の日英対訳パターンの例

日本語
とうとう彼はそれを買うことを承知した。 とうとうNはNをNことをNた。
英語
He finally agreed to buy it . N finally N to N N .

3.4 デコーダ

本研究ではデコーダとして“Moses[6]”を用いる。

3.5 実験内容

本研究では、以下の条件の組み合わせで実験を行う。

- 日英対訳文として、単文と重文複文を用いる。
- 翻訳実験として、日英翻訳と英日翻訳を行う。

3.6 評価方法

本研究では、翻訳文の評価方法として、自動評価法と人手評価法を用いる。

3.6.1 自動評価法

本研究では、自動評価として、BLEU[8]、METEOR[9]、RIBES[10]、WER[11]を用いる。

3.6.2 人手評価法

本研究では、人手評価として、翻訳結果1万文から100文をランダムに抽出し、対比較評価を行う。対比較評価における表記方法の説明を以下に示す。

- baseline
翻訳精度が比較する手法よりもbaselineが優れていると判断した場合
- 提案手法
翻訳精度がbaselineよりも比較する手法が優れていると判断した場合
- 差なし
同一出力ではないが、翻訳精度に優劣がつけ難い場合
- 一致
一字一句同じ出力の場合

4 パターン追加実験

本章の実験では、翻訳モデル作成の際の学習データとして、日英対訳文に日英対訳パターンを追加した日英対訳データを用いる。尚、言語モデル作成の際の学習データはすべての実験において、10万文の単一言語文のみを使用する。また、baselineとは日英対訳パターンの追加を行っていない場合の実験結果である。

4.1 自動評価結果

自動で作成した日英対訳パターンを用いた各実験における自動評価結果の単文を用いた実験の結果を表6に示す、重文複文を用いた実験の結果を表7に示す。尚、表中の“日英”、“英日”はそれぞれ、日英翻訳、英日翻訳を意味する。

表6 自動評価結果(単文)(%)

	BLEU	METEOR	RIBESE	WER
日英				
baseline	13.6	45.3	70.8	73.7
提案手法	13.8	45.1	71.1	73.1
英日				
baseline	17.8	45.1	76.7	69.0
提案手法	18.1	45.9	76.7	68.0

表7 自動評価結果(重文複文)(%)

	BLEU	METEOR	RIBESE	WER
日英				
baseline	12.4	41.5	67.3	80.9
提案手法	12.7	42.0	67.6	80.4
英日				
baseline	14.3	40.0	70.4	79.4
提案手法	14.3	40.0	70.8	77.6

4.2 人手評価結果

自動で作成した日英対訳パターンを用いた各実験におけるベースラインと提案手法の対比較評価結果の単文を用いた実験の結果を表8に示す、重文複文を用いた実験の結果を表9に示す。

表8 baseline VS 提案手法(単文)(文)

	baseline	提案手法	差なし	一致
日英	4	11	65	15
英日	3	10	70	17

表9 baseline VS 提案手法(重文複文)(文)

	baseline	提案手法	差なし	一致
日英	3	4	83	6
英日	6	8	79	7

4.3 対比較評価例

表8の日英翻訳において、提案手法が良いと判断した例を表10、英日翻訳において、提案手法が良いと判断した例を表11、表9の日英翻訳において、提案手法が良いと判断した例を表12、英日翻訳において、提案手法が良いと判断した例を表13に示す。

表10 baseline VS 提案手法(単文_日英)

提案手法 > baseline	
入力文	そんなことをするなんてなんて彼は 間抜けなんだろう。
正解文	What a stupid man he is to do such a thing !
baseline	He is of a to do such a thing .
提案手法	He stupid of him to have done such a thing .

表 11 baseline VS 提案手法 (単文_英_日)

提案手法 > baseline	
入力文	There is no point in looking for him .
正解文	彼を探しても無駄だ。
baseline	彼を求めてはない。
提案手法	彼を探しても無意味である。

表 12 baseline VS 提案手法 (重文複文_日_英)

提案手法 > baseline	
入力文	仕事が忙しくて、かなわない。
正解文	I am so busy with work I can not stand it .
baseline	I have been too busy at work , and .
提案手法	I have been too busy at work , and will not come true .

表 13 baseline VS 提案手法 (重文複文_英_日)

提案手法 > baseline	
入力文	The vase fell and broke .
正解文	花瓶が倒れて割れてしまった。
baseline	花瓶が落ちてしまった。
提案手法	花瓶が落ちて割れた。

4.4 実験結果のまとめ

- 自動評価, 人手評価ともに, baseline に比べて提案手法の評価値が向上している。
- 自動評価より人手評価の方が大きな効果を示す評価となった。
- 単文と重文複文の場合を比べると, 単文での効果が大きかった。
- 日英翻訳と英日翻訳の場合を比べると, 効果にほとんど差は見られなかった。

5 考察

実験結果より, 統計翻訳における対訳パターンの学習の有効性を示すことができた。その理由としては, パターンは文の重要な構成要素であるため, 本来, 文を学習している統計翻訳において, パターンの学習が効果的であったと考えられる。また, 単文より重文複文の効果が少なかったのは, 重文複文は構造が複雑なため誤ったパターン化を行ってしまうことが原因ではないかと考えられる。

6 追加実験

4章の実験において, パターン追加の有効性が示された。ところで, パターン化には様々な手法がある。そこで本章では, 人手作成パターンを用いた場合の効果の調査や様々なパターン化手法の検討を追加実験として行った。尚, 日英対訳パターン原文には, そのぞれの実験で学習データとして用いる対訳文 10 万文を用いる。

6.1 人手作成パターンの効果の調査

4章の実験において, 自動作成パターンの追加の有効性が示された。そこで本節では, 人手作成パターンの追加の効果の調査を行う。今回は, 有効性を判断するために, 自動作成パターンと人手作成パターンの評価結果の比較を行う。人手作成パターンは鳥バンク [2] のものを用いる。

6.1.1 人手作成パターンのパターン化手法

パターン化手法として以下の 2 手法を用いる。

- 手法 A: 変数化可能な単語を全て変数化し, さらにその変数番号を含めた変数をすべて N に置き換えたパターン。
- 手法 B: 変数化可能な単語を全て変数化し, 変数番号をつけたパターン。

各パターン化手法の例を表 17 に示す。

表 14 各パターン化手法の例

日本語	
原文	とうとう彼はそれを買うことを承知した。
手法 A	$N N$ は N を N ことを N 。
手法 B	$ADV1 N2$ は $N3$ を $V4 \hat{rentai}$ ことを $V5.kako$ 。
英語	
原文	He finally agreed to buy it .
手法 A	$N N N$ to $N N$.
手法 B	$N2 ADV1 V5 \hat{past}$ to $V4 \hat{base} N3$.

6.1.2 人手作成パターンの効果の調査の評価結果

自動作成パターンを用いた実験と手動作成パターンを用いた実験の対比較評価結果を, 表 15 と表 16 に示す。

表 15 自動 VS 手動 (手法 A)(文)

	自動	手動	差なし	一致
日 英	1	2	91	6
英 日	5	3	90	2

表 16 自動 VS 手動 (手法 B)(文)

	自動	手動	差なし	一致
日 英	5	1	88	6
英 日	6	1	87	6

6.1.3 人手作成パターンの効果の調査の考察

実験結果より, 全体的に, 自動作成パターンを用いる場合よりも人手作成パターンを用いる場合の方が翻訳精度が低い値となった。この理由としては, 人手作成パターンは自動作成パターンに比べて, 正確なパターン化が行われている。それ故に, 確実に対応がとれている部分だけパターン化を行っているのだから, 自動作成パターンに比べると変数化される部分が少なくなり, 字面部分が多く残ってしまうことが原因となっていることが考えられる。

6.2 パターン化手法の検討

4章の実験では, パターン化を行う手法として, 変数化可能な単語を全て変数化し, 変数番号を含めた変数を全て N に置き換えたパターン化を行うという手法を用いて, その結果, 提案手法の有効性は示された。しかし, 実際にはパターン化を行う手法としては, 4章で用いた手法以外にも多くの手法がある。そこで, 本節ではさらに, パターン化を行う手法として, 最適な手法について詳しく調べるために, 追加実験を試みる。

6.2.1 パターン化手法

本節の実験では, パターン化手法として, 4章の実験で用いた手法を合わせた計 4 種類の手法を用いて実験を行った。パターン化手法の詳しい説明を以下に示す。

- 手法 A: 変数化可能な単語を全て変数化し, さらにその変数番号を含めた変数をすべて N に置き換えたパターン
- 手法 B: 変数化可能な単語を全て変数化し, 変数番号をつけたパターン
- 手法 C: 変数化可能な動詞以外の単語を変数化し, 変数番号をつけたパターン
- 手法 D: 変数化可能な動詞以外の単語を変数化し, さらにその変数番号を含めた変数を N に置き換えたパターン

各パターン化手法の例を表 17 に示す。

表 17 各パターン化手法の例

日本語	
原文	ナイロンは原油より作られる。
手法 A	N は N より N れる。
手法 B	$N1$ は $N2$ より $VERB3$ れる。
手法 C	$N1$ は $N2$ より作られる。
手法 D	N は N より作られる。
英語	
原文	Nylon is made from crude oil .
手法 A	N is N from N .
手法 B	$N1$ is $VERB3$ ~past from $N2$.
手法 C	$N1$ is made from $N2$.
手法 D	N is made from N .

6.2.2 パターン化手法の検討の評価結果

本節の実験では, パターン化手法として前節で示した, 4 種類のパターン化手法を用いて実験を行う。それ以外の条件は 4 章の実験と全て同様な条件のもとで行う。

各パターン化手法を用いた実験の自動評価結果を表 18 に示す。

表 18 自動評価結果 (単文)(%)

	BLEU	METEOR	RIBES	WER
日 英				
baseline	13.6	45.3	70.8	73.7
手法 A	13.8	45.1	71.1	73.1
手法 B	14.2	46.2	71.1	72.6
手法 C	14.1	45.9	71.6	72.1
手法 D	14.1	46.4	71.3	71.6
英 日				
baseline	17.8	45.1	76.7	69.0
手法 A	18.1	45.9	76.7	68.0
手法 B	18.6	46.2	77.1	67.5
手法 C	18.5	46.1	76.9	68.1
手法 D	18.3	46.1	77.3	67.9

各パターン化手法を用いた実験の人手評価結果を表 19, 表 20, 表 21, 表 22 に示す。

表 19 baseline VS 手法 A(文)

	baseline	手法 A	差なし	一致
日 英	4	11	70	15
英 日	3	10	70	17

表 20 baseline VS 手法 B(文)

	baseline	手法 B	差なし	一致
日 英	4	14	66	16
英 日	2	13	69	16

表 21 baseline VS 手法 C(文)

	baseline	手法 C	差なし	一致
日 英	2	15	69	14
英 日	1	11	72	16

表 22 baseline VS 手法 D(文)

	baseline	手法 D	差なし	一致
日 英	3	14	68	15
英 日	3	13	69	15

6.2.3 パターン化手法の考察

本実験では, すべてのパターン化において, 提案手法が有効であることが示される結果となった。しかし, 各パターン化手法の間には大きな差は見られなかった。故に, パターン化手法の違いにおける有効性の差はあまりないと考えられる。

7 おわりに

日野らは日本語英語間の翻訳において, 小規模のコーパスに対訳句コーパスを追加し翻訳を行った。その結果, 翻訳精度の向上が確認できた。本研究では, 学習データの対訳文に対訳パターンを追加し, 翻訳精度の向上について調べた。実験の結果, 対訳パターン追加の効果が示された。

さらに, 人手作成パターンと自動作成パターンの比較により, 人手作成パターンが有効であるとは限らないと考えられることがわかった。また, 様々なパターン化手法を用いて実験を行った結果, 提案手法の有効性は示すことができたが, 各パターン化手法ごとの精度の差はほとんど見られなかった。

参考文献

- [1] Popovic Maja, and Ney Hermann "Statistical Machine Translation with a small amount of bilingual training data", 5th LRECSALTMIL Workshop on Minority Languages, pp.25-29. 2006.
- [2] 鳥バンク: <http://unicorn.ike.tottori-u.ac.jp/toribank/>
- [3] 英辞郎: <http://www.alc.co.jp/>
- [4] 日野聡子ほか, "統計翻訳における対訳句コーパスの効果" 鳥取大学大学院 平成 24 年度 修士論文.
- [5] MeCab, 日本語形態素解析器 <http://mecab.sourceforge.net/>
- [6] 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.
- [7] 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.
- [8] BLEU, NIST Open MT Scoring <http://www.itl.nist.gov/iad/894.01/tests/mt/>
<http://www.itl.nist.gov/iad/mig/test/mt/>
- [9] METEOR "The METEOR Automatic Machine Translation Evaluation System" <http://www.cs.cmu.edu/~alavie/>
- [10] RIBES: Rank-based Intuitive Bilingual Evaluation Measure <http://www.kecl.ntt.co.jp/icl/lirg/ribes>
- [11] Gregor Leusch, Nicola Ueffing and Hermann Ney. "A Novel String-to-String Distance Measure with Applications to Machine Translation Evaluation." In Proc. of MT Summit IX, 240-247. TRANSLATION ERROR RATE (TER) 7.0 <http://www.cs.umd.edu/~snover/tercom/> (2003)