

人手作成したフレーズ辞書と自動作成した文パターン辞書を利用したパターンに基づく統計翻訳

桃崎新 村上仁一 坂田純 徳久雅人
鳥取大学 工学部 知能情報工学科

{s122053, murakami, d112004, tokuhisa} @ ike.tottori-u.ac.jp

1 はじめに

パターン翻訳は、入力された原言語文に対し、フレーズ辞書と文パターン辞書を適応させて翻訳文を出力する方法である。しかし、フレーズ辞書と文パターン辞書は、人手で作成するために、開発にコストがかかる。

開発コストを削減するため、江木は、フレーズ辞書と文パターン辞書を自動作成する、Pattern Based SMT[1](以下、全自動手法)を提案した。しかし、翻訳精度はまだ低い。

そのため、坂東は、翻訳精度の向上のために、人手作成されたフレーズ辞書を用いた Pattern Based SMT[2](以下、半自動手法)を提案した。その結果、半自動手法は、翻訳精度の高い文が出力されるようになった。しかし、翻訳精度の低い文も出力されるようになった。そして、翻訳文全体で見ると翻訳精度が低くなった。

この原因は、人手作成したフレーズ辞書は、自動作成したフレーズ辞書よりもフレーズの数が少ないので、文パターン辞書の数が少なくなる。そのため、未知語を含む文が多く出力されてしまったということが考えられる。

本研究では、文パターン辞書を増やすことで、翻訳文全体の翻訳精度の向上を目指す。その方法として、全自動手法で得られた文パターン辞書を用いる。

2 全自動手法 [1]

全自動手法は、フレーズ辞書と、文パターン辞書の両方を自動作成する。全自動手法は大きく分けて5つのステップで翻訳を行う。以下に全自動手法の手順を示す。

2.1 単語辞書の作成

対訳学習文と GIZA++ を利用して単語辞書を作成する。

2.2 単語に基づく文パターン辞書(自動)の作成

対訳学習文と作成した単語辞書を用いて単語に基づく文パターン辞書(自動)を作成する。単語に基づく文パターン辞書(自動)の作成手順と例文を図1に示す。

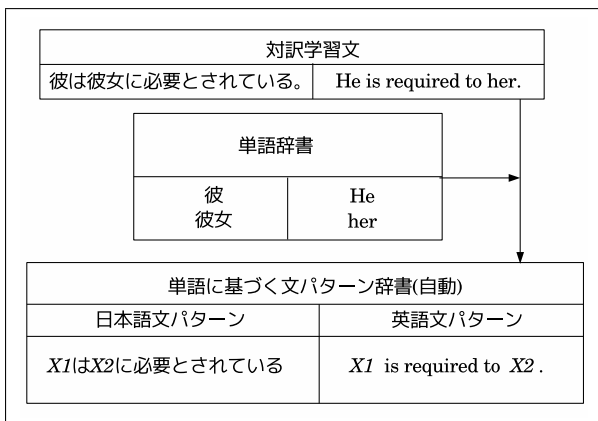


図1 単語に基づく文パターン辞書(自動)の作成手順

2.3 フレーズ辞書(自動)の作成

対訳学習文と単語に基づく文パターン辞書(自動)を用いて、フレーズ辞書(自動)を作成する。

2.4 句に基づく文パターン辞書(全自動)の作成

対訳学習文とフレーズ辞書(自動)を用いて、句に基づく文パターン辞書(全自動)を作成する。

句に基づく文パターン辞書(全自動)を作成する手順と例文を図2に示す。

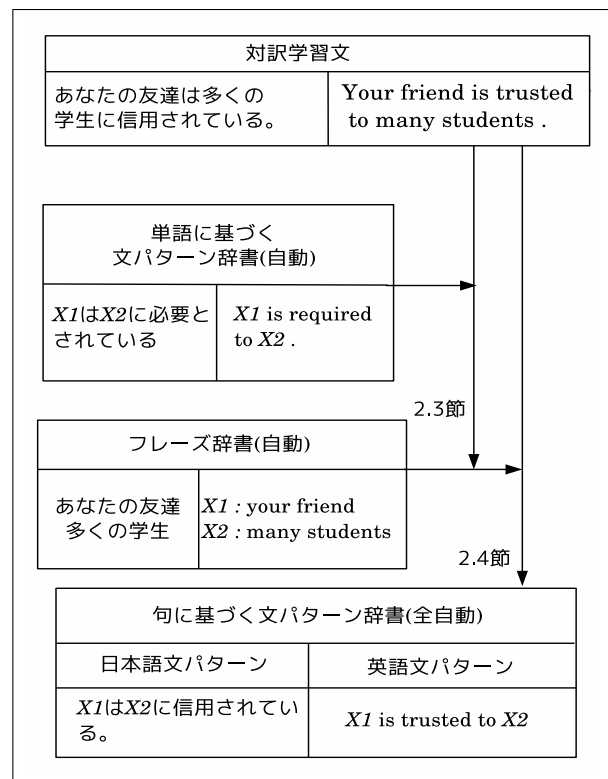


図2 句に基づく文パターン辞書(全自動)の作成手順

2.5 翻訳文の出力

日本語文を入力して、フレーズ辞書(自動)と句に基づく文パターン辞書(全自動)を用いて、翻訳文を出力する。

3 半自動手法 [2]

半自動手法は、人手作成された対訳句を用いてフレーズ辞書を作成することで、文パターン辞書を半自動手法で翻訳を行う。半自動手法の手順を以下に示す。

3.1 フレーズ辞書(手動)の作成

人手作成された対訳句を用いて、フレーズ辞書(手動)を作成する。

3.2 句に基づく文パターン辞書(半自動)の作成

対訳学習文とフレーズ辞書(手動)を用いて、句に基づく文パターン辞書(半自動)を作成する。半自動手法の句に基づく文パターン辞書(半自動)の作成手順を図3に

示す。

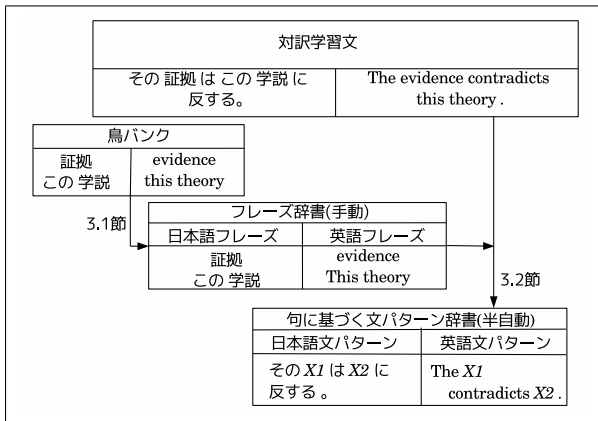


図 3 句に基づく文パターン辞書 (半自動) の作成手順

3.3 翻訳文の出力

日本語文を入力して、フレーズ辞書 (手動) と句に基づく文パターン辞書 (半自動) を用いて、翻訳文を出力する。

4 提案手法

半自動手法は、翻訳精度の高い文が出力されるようになった。しかし、翻訳精度の低い文も出力されるようになった。そして、翻訳文全体で見ると翻訳精度が低くなった。この原因は、以下のことが考えられる。

- ・ 人手作成したフレーズ辞書は、自動作成されたフレーズ辞書よりもフレーズの数が少ない。
- ・ 半自動手法の文パターン辞書は全自動手法よりも文パターンの数が少ない。

これらより、半自動手法で翻訳精度の低い文が多く出力されたと考えられる。

そこで、文パターン辞書の数を増やすことで、翻訳文全体の翻訳精度を向上させる方法を提案する。具体的には、全自動手法で得られた文パターン辞書 (全自動) と、半自動手法で得られたフレーズ辞書 (手動) を用いて翻訳する。

提案手法の手順を以下に示す。

4.1 フレーズ辞書 (手動) の作成

半自動手法より、人手作成された対訳句を用いてフレーズ辞書 (手動) を作成する。

4.2 句に基づく文パターン辞書 (全自動) の作成

全自動手法より、句に基づく文パターン辞書 (全自動) を作成する。

4.3 翻訳文の出力

フレーズ辞書 (手動) と句に基づく文パターン辞書 (全自動) を用いて英語翻訳文を出力する。

提案手法による翻訳手順を図 4 に示す

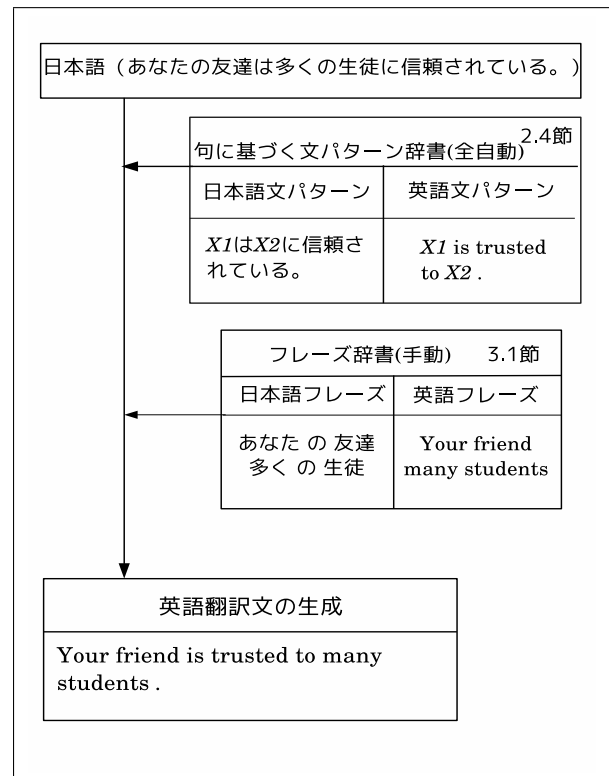


図 4 英語翻訳文出力手順 (提案手法)

5 実験環境

5.1 対訳句

本研究では、対訳句として鳥バンク [4] を用いる。

5.1.1 鳥バンク

鳥バンクは自然言語処理のための言語知識ベースを収録したデータバンクであり、日本語の重文と複文を対象とする「意味類型パターン辞書」が収録されている。本研究では、このパターン辞書から抽出した対訳句を用いる。対訳句の例を表 1 に示す。

表 1 鳥バンクから抽出した対訳句の例

あなたのお父さん	Your father
息子の話	son's story
移民政策	immigration policy

5.2 実験データ

本研究では、日英対訳文として単文コーパスを表 2 の内訳で用いる。

表 2 日英対訳文数

対訳学習文	100,000 文対
テストデータ	100 文

6 実験結果

6.1 提案手法と半自動手法の比較

提案手法と半自動手法で得られた英語翻訳文 100 文に対して、対比較評価を行う。提案手法と半自動手法の対比較評価の結果を表 3 に示す。表 3 中の表記方法について説明する。

- 提案手法 : 提案手法の翻訳精度が半自動手法の翻訳精度より優れている文
- 半自動手法 : 半自動手法の翻訳精度が提案手法の翻訳精度より優れている文

- 差なし：2種類の翻訳精度が同程度である、もしくは両方とも翻訳文として適していない文
- 同一出力：2種類の翻訳文が完全に同一な文

表3 提案手法と半自動手法の対比較評価結果

提案手法	半自動手法	差なし	同一出力
29	29	25	17

表3より、提案手法と半自動手法では差はほとんどない結果となった。

提案手法が半自動手法より優れていた文と、半自動手法が提案手法より優れていた文の出力例を表4と表5にそれぞれ示す。

表4 提案手法が優れている翻訳例

例1	
入力文	正直者がばかを見ることが多い。
参照文	Honesty does not often pay .
提案手法	An honest often watch the fool .
半自動手法	We often see ばかを honesty .
例2	
入力文	黒板のチョークが袖についた。
参照文	The chalk on the blackboard rubbed off on his sleeve .
提案手法	the blackboard chalk was on the sleeve .
半自動手法	The the blackboard chalk lucky sleeve .

表5 半自動手法が優れている翻訳例

例1	
入力文	彼は5日の午後に到着する。
参照文	He will arrive on the afternoon of the 5th .
提案手法	He arrived on the afternoon take five days .
半自動手法	He will arrive in the afternoon of the five days .
例2	
入力文	昼食をたっぷり取った。
参照文	I had a big lunch .
提案手法	I took a lunch .
半自動手法	I took lunch plenty .

6.2 提案手法と全自動手法の比較

提案手法と全自動手法で得られた英語翻訳文100文に対して、対比較評価を行った。表6に提案手法と全自動手法の対比較評価の結果を示す。

表6 提案手法と全自動手法との対比較評価結果

提案手法	全自動手法	差なし	同一出力
37	30	22	11

表6より、提案手法と全自動手法では、明確ではないが、提案手法の方が優れている結果となった。

提案手法が全自動手法より優れていた文と、全自動手法が提案手法より優れていた文の出力例を表7と表8にそれぞれ示す。

表7 提案手法が優れている翻訳例

例1	
入力文	関税は完全に撤廃された。
参照文	Tariffs have been eliminated altogether .
提案手法	Tariffs 撤廃 completely .
全自動手法	関税 It was in full .
例2	
入力文	ウイスキーを1杯もらおう。
参照文	I'll have a whiskey .
提案手法	I'll take a glass of whisky of it .
全自動手法	Let's whiskey gaffle cup .

表8 全自動手法が優れている翻訳例

例1	
入力文	フランス語は父の知り合いのフランス人から教わっています。
参照文	I'm learning French from a French gentleman my father knows .
提案手法	French まず of his father 教わっ empty acquaintances French .
全自動手法	French have lessons in French father acquainted .
例2	
入力文	部屋がぐるぐる回るようだった。
参照文	The room seemed to spin round .
提案手法	He seemed reeled が room .
全自動手法	As room was spinning turns .

7 考察

7.1 全自動手法と半自動手法の比較

全自動手法と半自動手法で得られた英語翻訳文100文に対して、対比較評価を行った。表9に全自動手法と半自動手法の対比較評価の結果を示す。

表9 全自動手法と半自動手法との対比較評価結果

全自動手法	半自動手法	差なし	同一出力
27	28	40	5

表9より、全自動手法と半自動手法では、差はほとんどない結果となった。全自動手法が半自動手法より優れていた文と、半自動手法が全自動手法より優れていた文の出力例を表10と表11に示す。

表10 全自動手法が優れている翻訳例

例1	
入力文	池の表面に薄い氷が張る。
参照文	The pond was frozen over with thin ice .
全自動手法	The pond surface forms stood ice .
半自動手法	On the surface of the pond there weaves a thin ice .
例2	
入力文	部屋がぐるぐる回るようだった。
参照文	The room seemed to spin round .
全自動手法	As room was spinning turns .
半自動手法	That was the turning よう room around .

表 11 半自動手法が優れている翻訳例

例 1	
入力文	彼はたった 1 0 0 0 円しか持っていない。
参照文	He has only 1,000 yen on him .
全自動手法	He is no more than one ten thousand yen .
半自動手法	He has no more than one thousand yen .
例 2	
入力文	自分の健康状態について主治医に相談した。
参照文	He has consulted his physician about his health .
全自動手法	He consulted with his health について doctor .
半自動手法	He consulted doctor health in the condition .

7.2 提案手法と moses[6] の比較

提案手法と moses で得られた英語翻訳文 100 文に対して、対比較評価を行った。表 12 に提案手法と moses の対比較評価の結果を示す。^{*1}

表 12 提案手法と moses との対比較評価結果

提案手法	moses	差なし	同一出力
43	29	23	5

表 12 より、提案手法と moses では提案手法の方が優れている結果になった。提案手法が moses より優れていた文と、moses が提案手法より優れていた文の出力例を表 13 と表 14 に示す。

表 13 提案手法が優れている翻訳例

例 1	
入力文	その講座は 1 月に終わる。
参照文	The course finishes in January .
提案手法	The course will be completed in January .
moses	he Takaoka ends in January .
例 2	
入力文	その計画に不賛成であった。
参照文	He looked upon the project with disfavor .
提案手法	He was disapproval for the plan .
moses	The plan was said .

表 14 moses が優れている翻訳例

例 1	
入力文	そのワインをちびりちびり飲んだ。
参照文	He sipped the wine .
提案手法	He drank an ちびりちびり of it .
moses	He drank in small sips the the wine .
例 2	
入力文	彼は健康を損なった。
参照文	He ruined his health .
提案手法	I hurt his health .
moses	He ruined his health .

7.3 提案手法の有効性

本研究において、人手作成したフレーズ辞書(手動)と自動作成した文パターン辞書(全自動)を用いた翻訳と人手作成したフレーズ辞書と文パターン辞書を用いた翻訳は翻訳精度に大きな差はなかった。この結果から、文パターンの数をこれ以上増やしても、翻訳精度が大きく向上する可能性は低いということが分かった。この原因を、誤り解析をして、今後の研究で考えていきたい。

8 おわりに

本研究では、精度の高いフレーズ辞書を使用しながら、翻訳精度が上がらないという事案に対して、文パターンの数を増やすことにより、翻訳精度の改善を試みた。その方法として、半自動手法で得られたフレーズ辞書(手動)と、全自動手法で得られた文パターン辞書(全自動)を用いて翻訳を行った。そして、半自動手法、全自動手法との対比較実験をそれぞれ行った。

実験結果より、翻訳精度は、提案手法と全自動手法では、明確な差はないが、提案手法が少し優れている結果となった。しかし、提案手法と半自動手法では、ほぼ同等で差がほとんど見られないという結果になってしまった。今後は、文パターン辞書の数を増やす以外の方法を考えて、翻訳精度の向上を検討したい。

参考文献

- [1] 江木孝史：“句に基づく文パターンを用いた英日翻訳”，2014 年修士論文
- [2] 坂東俊樹：“人手作成による対訳句を用いたパターンに基づく統計翻訳”，2014 年卒業論文
- [3] Franz Josef Och, Hermann Ney: “A Systematic Comparison of Various Statistical Alignment Models”, Computational Linguistics, pp.19-51, 2003.
- [4] 鳥バンク：“http://unicorn.ike.tottori-u.ac.jp/toribank/
- [5] 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.
- [6] Moses: “Open Source Toolkit for Statistical Machine Translation”, Proceedings of the ACL 2007 Demo and Poster Sessions, pp.177-180. 2007.

^{*1} ただし、moses の実験は、提案手法と実行条件がそろっていない可能性がある