

多量の対訳句と少量の対訳文を用いた統計翻訳

池淵堅斗 村上仁一 坂田純 徳久雅人
鳥取大学大学院 工学研究科 情報エレクトロニクス専攻
{s102002, murakami, d112004, tokuhisa} @ ike.tottori-u.ac.jp

1 はじめに

近年、機械翻訳の分野で原言語から目的言語に翻訳する統計翻訳が注目されている。統計翻訳においては、対訳データの量が多ければ多いほど翻訳精度が高くなる。しかし、利用できる対訳データには限りがある。

この課題に対して、Maja Popovićらは、セルビア語英語間の翻訳において、5万句程度の対訳句と少量の対訳文を組み合わせた対訳データを用いて統計翻訳を行った。その結果、自動評価において翻訳精度が向上した[1]。

また日野らは、日本語英語間において、対訳文と対訳句を組み合わせた対訳データを用いた統計翻訳を行った。その結果、翻訳精度が向上した[2]。

現在、対訳文は一般的に多く存在しているが、10万句以上の大規模な対訳句のコーパスはほとんど存在しない。よって、多量の対訳句を用いた統計翻訳の実験はほとんど行われていない。

そこで、本研究では、多量の日英対訳句と少量の日英対訳文を組み合わせた対訳データを学習データとして用いることによる翻訳精度の調査を行う。

2 本研究における統計翻訳の流れ

本研究では、日英対訳句と日英対訳文を組み合わせた日英対訳データを翻訳モデル作成の学習データとして用いる。図1に日英方向の統計翻訳の流れを示す。

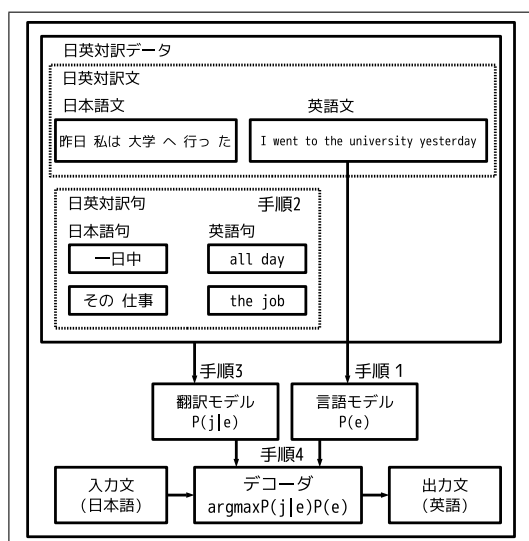


図1 日英方向の統計翻訳の流れ

提案手法の手順を以下に示す。

- 手順1 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句を追加する
- 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

3 実験環境

3.1 前処理

統計翻訳の前処理として、日本語句と日本語文に対して、MeCab[3]を用いて形態素解析を行う。また、英語句と英語文に対して“tokenizer.sed [4]”を用いて分かち書きを行う。

3.2 日英対訳句

本研究では、日英対訳句の種類として、鳥バンク [5] と英辞郎 [6] を用いる。日英対訳句の内訳を表1に示す。

表1 日英対訳句数

日英対訳句	学習データ (句)
鳥バンク	548,694
英辞郎	1,366,458

3.2.1 鳥バンク [5]

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日英対訳句を用いる。鳥バンクの日英対訳句の例を表2に示す。

表2 鳥バンクから抽出した日英対訳句の例

彼のお母さん	his mother
息子の話	son's story
社会 通念	commonly accepted ideas

3.2.2 英辞郎 [6]

英辞郎は、EDP(Electronic Dictionary Project)がアップデートし続けている英和・和英辞書である。英辞郎のデータには対訳句の他に翻訳例や注釈、本来の文に出てこない“~”等の記号が含まれる。英辞郎の日英対訳句の例を表3に示す。

表3 英辞郎から抽出した日英対訳句の例

に急落する	plummet to
から出てくる	come out from
の結果として生じる	come out from

3.3 日英対訳文

本研究では、日英対訳文として単文 [7] を用いる。本研究で用いる日英対訳文の内訳を表4に示す。日英対訳文の例を表5に示す。

表4 日英対訳文の内訳

日英対訳単文	文数 (文)
学習データ	100,000
テストデータ	10,000
デベロップメントデータ	1,000

表5 日英対訳文の例

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

3.4 デコーダ

本研究ではデコーダとして“Moses[8]”を用いる。

3.5 パラメータチューニング

本研究ではパラメータの最適化のために“mert-moses.pl”を用いる。パラメータの値は、全ての対訳文と全ての対訳句を学習データとした場合に得られた値を全ての実験に用いる。

3.6 評価方法

本研究では、翻訳結果の評価として自動評価と人手評価を行う。自動評価は自動評価法“BLEU[9]”, “METEOR[10]”, “RIBES[11]”, “TER[12]”, “WER[12]”を用いる。人手評価は翻訳結果からランダムに100文抽出し、対比較評価を行う。

4 鳥バンクの対訳句を用いた実験

本章の実験では、鳥バンクの日英対訳句(以降、対訳句)と日英対訳文(以降、対訳文)を組み合わせた日英対訳データ(以降、対訳データ)を翻訳モデル作成の学習データとして用いる。また、翻訳精度の変化の詳細を調査するために、対訳句数は一定数(548,694句)に保ち、対訳文数の変化に伴う翻訳精度の調査を行う。

4.1 自動評価結果

各実験の翻訳結果に対する各自動評価値を図2に示す。

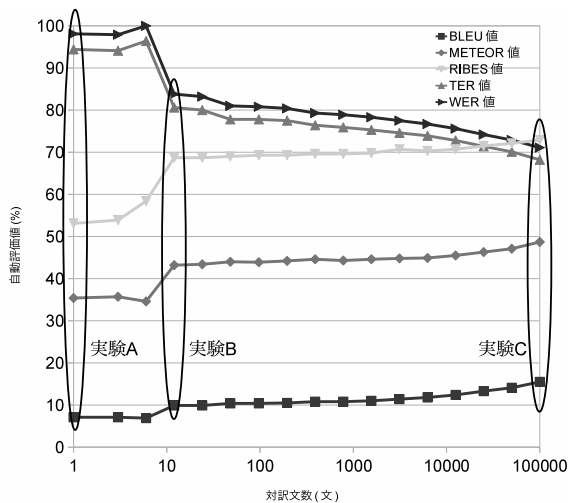


図2 対訳文数の変化に伴う自動評価値の変化

図2より、学習データとして用いる対訳文数が増加するにつれて自動評価値も増加している。しかし特に、対訳文数が6文の場合から12文へと増加する場合において、すべての自動評価値が急激に増加していることがわかる。

対訳文数が0文の場合を“実験A”, 12文の場合を“実験B”, 100,000文の場合を“実験C”として各自動評価値を表6に示す。

表6 対訳文数の変化に伴う自動評価値の変化

実験の種類	実験A	実験B	実験C
学習データ(文)	0	12	100,000
BLEU(%)	7.0	9.9	15.5
METEOR(%)	35.2	43.2	48.7
RIBES(%)	52.8	68.7	72.9
TER(%)	94.9	80.6	68.2
WER(%)	98.6	83.8	71.1

4.2 人手評価結果

人手評価手法として出力結果からランダムに100文抽出して対比較評価を用いる。対比較評価結果の表記の説明を以下に示す。

- 対比較評価において、比較する両実験の翻訳結果の翻訳精度に明確な差がないと判断した場合に“差なし”とする。

- 対比較評価において、比較する両実験の翻訳結果が完全に同一出力の場合に“一致”とする。

実験Aと実験Bの対比較評価結果を表7, 実験Bと実験Cの対比較評価結果を表8に示す。

表7 実験A VS 実験B

実験A > 実験B	実験A < 実験B	差なし	一致
1文	13文	82文	4文

表8 実験B VS 実験C

実験B > 実験C	実験B < 実験C	差なし	一致
5文	17文	75文	3文

表7より、自動評価結果同様、対訳文0文から12文に増加させることにより、翻訳精度が急激に向上していることがわかる。

4.3 対比較評価例

実験Aと実験Bの対比較評価例を表9, 実験Bと実験Cの対比較評価例を表10に示す。

表9 実験A VS 実験B

実験A > 実験B	
入力文	空気は黒い煙でもうもうとしていた。
正解文	The air was thick with black smoke .
実験A	The air has been thick in black smoke .
実験B	I have been thick in black smoke 。 air .

実験A < 実験B

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 .
実験A	I can of later his legs .
実験B	I do not know his legs 。 since then .

表10 実験B VS 実験C

実験B > 実験C

入力文	彼の声は落ち着き払っていた。
正解文	His voice was even .
実験B	His voice was with serenity .
実験C	He was very in his voice .

実験B < 実験C

入力文	彼はビール一杯で酔った。
正解文	He got drunk on one beer .
実験B	He is beer full drunk .
実験C	He was intoxicated with a glass of beer .

4.4 実験結果のまとめ

- 鳥バンクの多量の対訳句(約55万句)と少量の対訳文(12文)を学習データとして用いた場合には、鳥バンク(約55万句)と対訳文(0文)を用いる場合と比較して、急激に翻訳精度が向上する。
- 対訳文として12文以上用いる場合においては、対訳文の文数に比例して翻訳精度が向上する。

5 英辞郎の対訳句を用いた実験

本章の実験では、英辞郎の対訳句と対訳文を組み合わせた対訳データを翻訳モデル作成の学習データとして用いる。また、翻訳精度の変化の詳細を調査するために、対訳句数は一定数(1,366,458句)に保ち、対訳文数の変化に伴う翻訳精度の調査を行う。

5.1 自動評価結果

各実験の翻訳結果に対する各自動評価値を図3に示す。

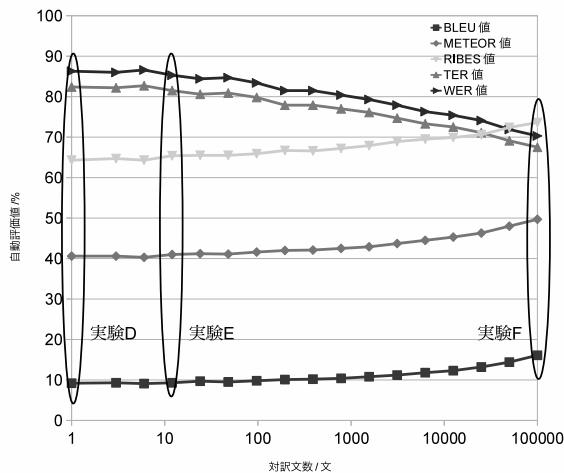


図3 対訳文数と自動評価値の関係

図3より、学習データとして用いる対訳文数が増加するにつれて自動評価値も増加している。しかし、特徴的な変化は見られなかった。

対訳文数が0文の場合を“実験D”，12文の場合を“実験E”，100,000文の場合を“実験F”として各自動評価値を表11に示す。

表11 対訳文数の変化に伴う自動評価値の変化

実験の種類	実験D	実験E	実験F
学習データ(文)	0	12	100,000
BLEU(%)	9.4	9.3	16.1
METEOR(%)	40.1	41	49.7
RIBES(%)	64.8	65.4	73.6
TER(%)	81.7	81.5	67.5
WER(%)	85.7	85.3	70.3

5.2 人手評価結果

実験Dと実験Eの対比較評価結果を表12、実験Eと実験Fの対比較評価結果を表13に示す。

表12 実験D VS 実験E

実験D > 実験E	実験D < 実験E	差なし	一致
2文	1文	60文	37文

表13 実験E VS 実験F

実験E > 実験F	実験E < 実験F	差なし	一致
1文	15文	82文	2文

表12より、自動評価結果同様、対訳文0文から12文に増加させることによる翻訳精度の差は確認できなかった。

5.3 対比較評価例

実験Dと実験Eの対比較評価例を表14、実験Eと実験Fの対比較評価例を表15に示す。

表14 実験D VS 実験E

実験D > 実験E	
入力文	彼女はいつもクラスの首席を占めている。
正解文	She is always at the top of her class .
実験D	She is always at the top of his class .
実験E	, she always at the top of his class .
実験D < 実験E	
入力文	彼はマラソンで脱落した。
正解文	He dropped away in a marathon race .
実験D	He is in the marathon race has dropped .
実験E	he has dropped in the marathon .

表15 実験E VS 実験F

実験E > 実験F	
入力文	空気は黒い煙でもうもうとしていた。
正解文	The air was thick with black smoke .
実験E	Black smoke have been cloud is in the air .
実験F	The air in the black smoke .
実験E < 実験F	
入力文	彼女はリンゴを半分にした。
正解文	She cut the apple in half .
実験E	She is half an apple .
実験F	She cut the apple in half .

5.4 実験結果まとめ

- 英辞郎の多量の対訳句(約137万句)と少量の対訳文(12文)を学習データとして用いた場合には、英辞郎(約137万句)と対訳文(0文)を用いる場合と比較して、翻訳精度にほとんど変化が見られない。
- 全体的に、対訳文数に比例して翻訳精度が向上する。

6 考察

6.1 鳥バンクを用いた実験の考察

4章の実験より、多量の対訳句が使用可能な場合、少量の対訳文があれば、ある程度の翻訳精度が得られる。これは、鳥バンクの多量の対訳句が翻訳に必要な単語をほとんどカバーしているため、翻訳を行う際にほとんど未知語が発生していないことが原因だと考えられる。また、文の構造を表現する情報は、言語モデル作成の際に用いる単一言語文と翻訳モデル作成の際に用いる少量の対訳文でカバーできているためであると考えられる。

6.2 英辞郎を用いた実験の考察

5章の実験より、英辞郎の対訳句を用いた場合、学習データが増加するのに比例して翻訳精度が向上している。しかし、少量の対訳文と組み合わせた場合に、鳥バンクを用いた場合と同様な急激な変化は見られなかった。これは、英辞郎には、文とみなされる句が存在している。そのため、そのような句が存在することにより多量の対訳句と少量の対訳文が組み合わせられている構造が自然に発生しているためであると考えられる。英辞郎の対訳句において文とみなされる句の例を表16に示す。

表16 文とみなされる句の例

お客様 以外 の ご 使用 は ご 遠慮 願います。
Rest rooms are for patrons only .

6.3 鳥バンクと英辞郎の比較

本節では、鳥バンクを用いた実験と英辞郎を用いた実験の結果の比較を行う。対訳句数は鳥バンク(約55万句)、英辞郎(約137万句)で一定数に保ち、対訳文数を変化させてそれぞれの翻訳精度の比較を行う。

鳥バンクと英辞郎における対訳文数の変化に伴うMETEOR値の変化を図4に示す。さらに、実験Aと実験Dの対比較評価結果を表17、実験Bと実験Eの対比較評価結果を表18、実験Cと実験Fの対比較評価結果を表19に示す。さらに、実験Aと実験Dの対比較評価例を表20、実験Bと実験Eの対比較評価例を表21、実験Cと実験Fの対比較評価例を表22に示す。

図4、表17、表18、表19より、対訳文数が少量の場合には0文である場合を除いて、鳥バンクの対訳句との組み合わせが有効であることがわかる。しかし、対訳文が増加していくにつれて、鳥バンクと英辞郎の差が小さくなっていくことがわかる。これは、少量の対訳文を用いる場合には、テストデータと同分野である鳥バンクの対訳句が効果的であるためであると考えられる。

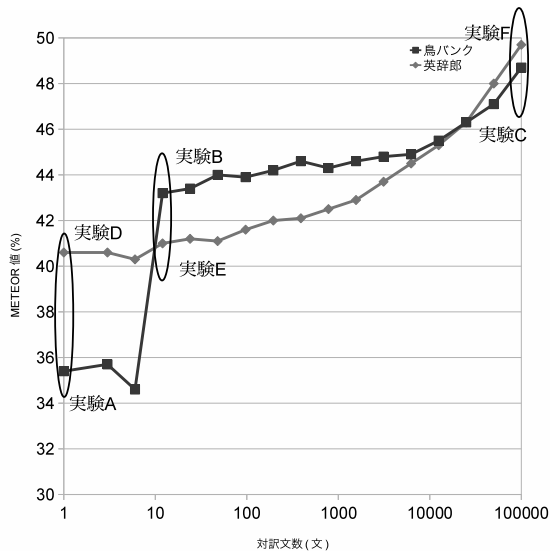


図 4 鳥バンクと英辞郎の比較

表 17 実験 A VS 実験 D

実験 A > 実験 D	実験 A < 実験 D	差なし	一致
2 文	6 文	92 文	0 文

表 18 実験 B VS 実験 E

実験 B > 実験 E	実験 B < 実験 E	差なし	一致
6 文	2 文	92 文	0 文

表 19 実験 C VS 実験 F

実験 C > 実験 F	実験 C < 実験 F	差なし	一致
9 文	7 文	71 文	13 文

表 20 実験 A VS 実験 D

実験 A > 実験 D	
入力文	私は学校に間に合った。
正解文	I made school in time .
実験 A	I was just in time to school .
実験 D	I time for school .
実験 A < 実験 D	
入力文	この風邪薬はよく効く。
正解文	This cold medicine is effective .
実験 A	The medicine is very effective .
実験 D	This medicine is effective for a cold .

表 21 実験 B VS 実験 E

実験 B > 実験 E	
入力文	自動車が動かない。
正解文	The car will not start .
実験 B	The car is not working .
実験 E	car is out of order .
実験 B < 実験 E	
入力文	彼女はいつもクラスの首席を占めている。
正解文	She is always at the top of her class .
実験 B	She is always the class will be the account for the top of his class .
実験 E	She is always at the top of his class .

表 22 実験 C VS 実験 F

実験 C > 実験 F	
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 .
実験 C	Please put the correct mail rates .
実験 F	Make a right postage .
実験 C < 実験 F	
入力文	彼の声は落ち着き払っていた。
正解文	His voice was even .
実験 C	He was very in his voice .
実験 F	His voice was composure .

7 おわりに

本研究では、学習データとして多量の対訳句と少量の対訳文を組み合わせた対訳データを用いることによる翻訳精度の調査を行った。その結果、多量の対訳句が使用可能である場合、少量の対訳文があれば、ある程度の翻訳精度が得られることがわかった。また、一定数以上の対訳文を用いることにより、翻訳精度は対訳文数に比例して精度が向上する。さらに、対訳文数が少量の場合は、テストデータと同分野の対訳句を用いることによって、より良い精度が得られることがわかった。

今後は、今回の実験で得られた結果の誤り分析を行い、さらなる精度の向上を目指したい。

参考文献

- [1] Popović Maja, and Ney Hermann “Statistical Machine Translation with a small amount of bilingual training data”, 5th LREC SALT MIL Workshop on Minority Languages, pp.25-29. 2006.
- [2] 日野聡子: “統計翻訳における対訳句コーパスの効果”, 2013 年修論
- [3] MeCab <http://mecab.sourceforge.net/>
- [4] tokenizer.sed <http://www.cis.upenn.edu/~treebank/tokenizer.sed>
- [5] 鳥バンク <http://unicorn.ike.tottori-u.ac.jp/toribank/>
- [6] 英辞郎: <http://www.alc.co.jp/>
- [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] Moses: “Open Source Toolkit for Statistical Machine Translation”, Proceedings of the ACL 2007 Demo and Poster Sessions, pp.177-180. 2007.
- [9] BLEU: “a Method for Automatic Evaluation of Machine Translation”, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp.311-318. 2002.
- [10] Meteor: Lavie Alon, and Denkowski Michael “An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments”, Proceedings of the Second Workshop on Statistical Machine Translation, pp.228-231. 2007.
- [11] Hideki Isozaki, “Automatic Evaluation of Translation Quality for Distant Language Pairs”, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp.944-952. 2010.
- [12] 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)