

主語補完による日英統計翻訳の翻訳精度の改善

古市将仁 村上仁一 徳久雅人 村田真樹
鳥取大学 工学部 知能情報工学科

{s072047, murakami, tokuhisa, murata} @ ike.tottori-u.ac.jp

1 はじめに

近年の機械翻訳では、対訳データから自動的に翻訳規則を生成し翻訳を行う統計翻訳が注目されている [1]. 統計翻訳において、主語が省略されている日本語文の翻訳精度が低いことは、問題点の1つである. 猪澤らは、文献 [2] において、主語が省略されている日本語文の日英翻訳は、翻訳精度が低いことを報告している.

そこで本研究では、主語が省略されている日本語文に対し、主語を付与することで、翻訳精度の向上を目指す. まず、主語を付与した日本語文と、主語が省略されている日本語文に対し、それぞれ日英統計翻訳を行う. 次に、翻訳モデルの確率と言語モデルの確率を掛け合わせた翻訳確率が最大となる出力文を選択する. そして、選択した文を、最終的な出力文とする.

実験の結果、自動評価と人手評価の両方において翻訳精度の向上が確認でき、提案手法の有効性を確認できた.

2 関連研究

日本語文において省略されている格要素の補完を行う研究として、中岩らの研究 [3][4][5][6] がある. 中岩らは、日本語において省略されている格要素の補完を行った. 省略補完のために、格の意味的制約・用言意味属性・様相表現・接続語などを考慮したルールを構築した. そして、構築したルールを用いて省略格要素の指示対象を推定した. 実験の結果、95%の省略格要素が適切に補完できた.

また、省略補完を統計翻訳へ適用した研究としては、我々の研究 [7] や、平らの研究 [8] がある. 平らは、日英統計翻訳において、日本語文の省略格要素を手で補完し、翻訳を行った. 実験の結果、テストデータのみ省略補完を行った場合、人手評価では、50文中15文において翻訳精度が向上し、3文において翻訳精度が低下した. また、自動評価では、BLEU 値において 0.0016 向上した.

3 主語の定義

文献 [9] において、日本語文は、主語の定義が曖昧であることを示している. そこで本研究では、単文において、名詞の後に助詞“は”、“が”、“も”がある句を主語と定義する. 表1の下線は、主語の例である.

表1 主語の例
私は 急いで朝食を飲みこんだ。
水が 腐っている。
彼も ついに音をあげた。

4 主語省略文の判断条件

日本語文において、主語を省略している文 (以後、主語省略文と呼ぶ) が多く存在する. また、命令文は主語を含まないことが多い. そこで、本研究では以下の2つの条件を満たす日本語文を主語省略文とする.

条件1 主語を含まない文

条件2 動詞が命令形ではない文

表2に主語省略文の例を示す.

表2 主語省略文例
昼食をたっぷり取った。
サッカーをした。

5 提案手法

提案手法では、まず、主語が省略されているテストデータに対し、8種類の主語を補完する. 次に、主語が省略されているテストデータ (以後、原文と呼ぶ) と、主語補完を行った8文に対して、それぞれ統計翻訳を行う. そして、出力文9文から、翻訳モデルの確率と言語モデルの確率を掛け合わせた翻訳確率が最大となる文を選択する. 最後に、選択した文を最終的な出力文とする. 図1に、提案手法の手順を示す. また、以下に図1の手順の詳細を示す.

手順1 テストデータに対し、表3に示す8種類の主語を補完する.

表3 補完する8種類の主語			
私は	彼は	彼女は	私たちは
それは	あなたは	彼らは	誰かが

図1の“原文 + 主語補完を行った8文”は、主語補完の例である.

手順2 主語補完を行う前の原文と手順1で主語補完を行った8文、合計9文に対して統計翻訳を行う.

手順3 手順2で翻訳した9文から、翻訳確率が最大となる文を出力文として選択する.

6 実験環境

6.1 実験データ

統計翻訳の前処理として、実験データの日本語文に対して、MeCab[10]を用いて分かち書きを行う. また、実験データの英語文に対して、tokenizer.perl[11]を用いて分かち書きを行う. 本実験で用いる実験データの詳細を以下に示す.

1) 学習データ

辞書の例文より抽出した単文コーパス 181,988 文 [12] から、学習データとして 100,000 文を用いる. 主語省略文は、100,000 文中 10,459 文である. なお、文献 [7] では、学習データに対して主語補完を行った場合、翻訳精度の低下を報告している. よって本研究では学習データの主語省略文に対して主語補完を行わない.

2) テストデータ

テストデータとして、単文 10,000 文中に含まれる主語省略文 1,501 文を用いる.

3) ディベロップメントデータ

ディベロップメントデータとして、単文 1,000 文を用いる. 主語省略文は、1,000 文中 153 文である. 主語省略文 153 文に対し、対訳英語文の主語を参照して主語補

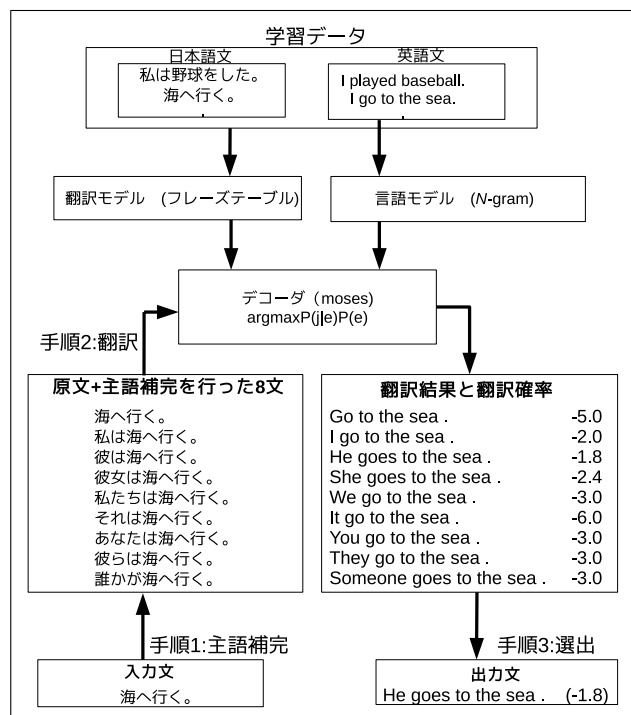


図 1 実験の手順

完を行う。

6.2 デコーダ

デコーダには、Moses[11]を用いる。本研究では、デベロップメントデータ 1,000 文を利用し、Moses のパラメータを最適化する。

6.3 評価方法

6.3.1 自動評価

本研究では、出力文の自動評価として、BLEU[13], METEOR[14], NIST[15], RIBES[16] を使用する。

6.3.2 人手評価

本研究では、主語補完を行わない翻訳をベースラインとする。本研究では、人手評価として、提案手法とベースラインの対比較評価を行う。対比較評価には、テストデータの出力文よりランダムに抽出した 200 文を用いる。対比較評価の評価基準を表 4 に示す。

表 4 対比較評価の評価基準

ベースライン○	提案手法の翻訳精度がベースラインの翻訳精度よりも低い場合
提案手法○	提案手法の翻訳精度がベースラインの翻訳精度よりも高い場合
差なし	提案手法の翻訳精度とベースラインの翻訳精度に差がない場合
同一出力	提案手法の出力とベースラインの出力が完全に一致する場合

7 実験結果

7.1 自動評価結果

自動評価結果を表 5 に示す。

表 5 自動評価結果

	BLEU	METEOR	NIST	RIBES
ベースライン	0.1166	0.4084	3.9628	0.6677
提案手法	0.1210	0.4156	4.0061	0.6864

表 5 より、提案手法はベースラインと比較し、翻訳精度が高いことが確認できる。

7.2 人手評価結果

ベースラインと提案手法の人手評価結果を表 6 に示す。

表 6 人手評価結果

ベースライン○	提案手法○	差なし	同一出力
3	18	49	130

表 6 より、提案手法はベースラインと比較し翻訳精度が高いことが確認できる。

8 考察

8.1 提案手法の有効性の理由

実験結果より、提案手法の有効性が確認できた。提案手法が有効である理由は、以下の 3 種類を考えている。

- 主語の出力
- 日英間のフレーズ対応の改善
- 出力文の選択による改善

この 3 種類の理由について、例を挙げて説明する。

a) 主語の出力

主語の出力によって翻訳精度が改善した例を表 7 に示す。ただし、表中の翻訳確率*は、log 表記であり、底は 10 である。また、表 7 における提案手法は、“原文 + 主語補完を行った 8 文”に対する翻訳結果において、翻訳確率が最大となった出力文を示している。

表 7 主語の出力の例

入力文	イングランドの王となった。
正解文	He became King of England .
ベースライン (翻訳確率*)	became the king of England . (-2.167)
提案手法 (翻訳確率*)	He became the king of England . (-1.929)

表 7 において、ベースラインは主語を出力していない。一方提案手法は、“彼は”を補完した入力文に対する出力文を選択し、主語を出力していることが確認できる。

b) 日英間のフレーズ対応の改善

日英間のフレーズ対応の改善によって翻訳精度が改善した例を表 8 に示す。また、表 8 中の、ベースラインにおける日英の句のアライメントと、提案手法におけるアライメントを表 9 に示す。

表 8 適切な動詞の出力の例

入力文	ナイフを彼の腹に突き刺した。
正解文	She stuck the knife in his stomach .
ベースライン (翻訳確率*)	I was in his belly with a knife . (-2.616)
提案手法 (翻訳確率*)	He stabbed in his belly with a knife . (-2.240)

表 9 のベースラインでは、アライメントが不適切である。その結果、適切な動詞を出力していない。一方提案手法では、“彼は”を補完した文に対する出力文を選択し、アライメントが適切である。その結果、日本語フレーズ

表9 表8におけるアライメント

ベースライン	提案手法
ナイフ knife	彼 He
を a	は ナイフ を a knife
彼の his	彼の 腹 his belly with
腹 belly	に in
に in	突き刺した stabbed
突き刺し with	。 .
た I was	
。 .	

“突き刺した”と英語フレーズ “stabbed”が対応し、適切な動詞を出力している。

c) 出力文の選択による改善

ベースラインにおいて、主語省略文を適切に翻訳することがある。この場合、主語補完が翻訳に悪影響を与えることがある。一方提案手法では、原文に対する出力文を翻訳確率で選択するため、主語補完による翻訳への悪影響を防ぐことができる。例を表10に示す。表11に、提案手法の“原文”における翻訳確率及びアライメントを示す。表12に、提案手法の“私は”を補完した文の出力文における翻訳確率及びアライメントを示す。なお、表12の“私は”の出力文は、提案手法において、主語省略文に対し“私は”を補完した文の出力文である。

表10 主語省略文に対する出力文の出力

入力文	試験の結果に がっかりした。
正解文	I was disappointed by my examination results .
ベースライン	I was disappointed with the result of the examination .
(翻訳確率*)	(-1.415)
提案手法	I was disappointed with the result of the examination .
(翻訳確率*)	(-1.415)

表11 提案手法の“原文”の出力文における翻訳確率及びアライメント

“原文” の出力文 (翻訳確率*)	I was disappointed with the result of the examination . (-1.415)
アライメント	
試験の結果に the result of the examination がっかりした I was disappointed with	

表12 提案手法の“私は”の出力文における翻訳確率及びアライメント

“私は” の出力文 (翻訳確率*)	I was disappointed with the result of the examination for me . (-2.120)
アライメント	
私は for me 試験の結果に the result of the examination がっかりした I was disappointed with	

表10において、提案手法は、原文を選択している。表10の例では、表11より、“原文”の出力文におけるアライメントが適切である。“原文”の出力文は主語を出力し、翻訳精度が高く、翻訳確率も高い。一方、表12より、主語補完を行った場合、入力文の主語のアライメントが不適切である。“私は”の出力文は、誤ったフレーズ “for me”を出力し、翻訳確率は低い。

8.2 提案手法が有効でない例

統計翻訳では、翻訳精度が低い、翻訳確率が高い出力文が存在する。一方提案手法では、翻訳確率を用いて出

力文を選択する。よって、提案手法では、出力文の選択が適切ではない文が存在する。表13に、例を示す。また、提案手法における翻訳候補と翻訳確率を表14に示す。なお、表14の“私は”の出力文は、提案手法において、原文に対し“私は”を補完した文における出力文であり、“彼は”の出力文、“彼女は”の出力文なども同様である。

表13 ベースライン○の例

入力文	とうとう 雨 になった。
正解文	It began to rain at last .
ベースライン (翻訳確率*)	Eventually it began to rain . (-1.298)
提案手法 (翻訳確率*)	He finally to rain . (-1.297)

表14 表13における提案手法の翻訳確率

“原文”の出力文 (翻訳確率*)	Eventually it began to rain . (-1.298)
“私は”の出力文 (翻訳確率*)	I was finally to rain . (-1.713)
“彼は”の出力文 (翻訳確率*)	He finally to rain . (-1.297)
“彼女は”の出力文 (翻訳確率*)	She was finally to rain . (-1.627)
“私たちは”の出力文 (翻訳確率*)	We were finally to rain . (-2.075)
“それは”の出力文 (翻訳確率*)	It was finally to rain . (-1.738)
“あなたは”の出力文 (翻訳確率*)	You 're finally to rain . (-2.246)
“彼らは”の出力文 (翻訳確率*)	They were finally to rain . (-1.816)
“誰かが”の出力文 (翻訳確率*)	Someone was finally to rain . (-2.204)

表14より、提案手法の翻訳候補において、“原文”の出力文が、最も翻訳精度が高い。しかし、提案手法では、“彼は”を補完した出力文の翻訳確率が最大となる。よって、表13における提案手法の翻訳精度は、ベースラインよりも低い。

表13と同様にベースライン○であった例を表15に示す。

表15 ベースライン○の例

入力文	やがて 夏休み になります。
正解文	The summer vacation is just around the corner .
ベースライン (翻訳確率*)	I will soon be during the summer vacation . (-2.118)
提案手法 (翻訳確率*)	He will soon during the summer vacation . (-2.075)
入力文	すべて ひどい 混乱 状態 に あった。
正解文	Everything was in terrible confusion .
ベースライン (翻訳確率*)	Everything was in a terrible shambles . (-2.369)
提案手法 (翻訳確率*)	He was in a terrible shambles . (-2.319)

8.3 提案手法で選択した主語の数

テストデータに対する正解文の主語の調査を行った。主語ごとの文数を表16に示す。また、提案手法において、選択した出力文の各主語の文数を表17に示す。

表16より、正解文の主語は、“I”が最も多い。一方、表

表 16 正解文の主語

文頭単語	正解文 (文)
I	408
He	320
She	115
We	107
They	74
It	46
You	9
Someone	1
Anyone	1
Somebody	1
Anybody	1
その他	418

表 17 補完した主語の文数

選択した主語	文数
主語付与なし	821
彼は	554
私は	64
彼女は	42
彼らは	13
それは	6
誰かが	1
私たちは	0
あなたは	0

17 より, 提案手法では“彼は”を補完した出力文を最も選択している。よって, 本研究において, 正解文の主語と提案手法の主語が不一致である文が多く存在する。

ところで, 自動評価において, 出力文の主語と, 正解文の主語が一致しない場合, 翻訳精度が低くなる。よって, 本研究において, 自動評価結果は実際の翻訳精度よりも低くなる。表 18 に例を示す。

表 18 主語不一致の文例

入力文	昼食を たっぷり 取っ た。
正解文	I had a big lunch .
提案手法	He had a big lunch .
提案手法の BLEU 値	0.7598

一方, 人手評価においては, 提案手法の出力文の主語と正解文の主語の不一致に関して, 無視して評価を行っている。

8.4 複文に対する提案手法の適用

複文において, 単文と同様の実験を行った。複文は, 主語省略文を自動的に抽出することが困難である。そこで, 複文における実験では, 複文の主語省略文 20 文を人手で抽出し, テストデータとして用いた。表 19 に, テストデータ 20 文における人手評価結果を示す。また, 複文の翻訳例を表 20 に示し, 提案手法における翻訳候補と翻訳確率を表 21 に示す。なお, 表 21 は, 表 14 と同様の表記方法である。

表 19 人手評価結果

ベースライン○	提案手法○	差なし	同一出力
0	0	3	17

表 20 複文における翻訳例

入力文	より すぐれた 兵器 を 必要 と している。
正解文	We need better arms .
ベースライン出力文	We need a superior weapons .
提案手法出力文	We need a superior weapons .

表 21 複文における翻訳候補と翻訳確率

“原文”の出力文 (翻訳確率*)	We need a superior weapons . (-3.752)
“彼は”の出力文 (翻訳確率*)	He is superior weapons . (-4.311)

表 19 の人手評価結果より, 複文において提案手法は有効性は確認できない。しかし, 表 21 において, “原文”の

出力文は, 翻訳精度が高く翻訳確率も高い。一方, “彼は”の出力文は, 翻訳精度が低く翻訳確率も低い。よって, 翻訳確率を用いた提案手法の選択方法は有効であると考えられる。今後, テストデータの文数を増やし, 翻訳確率を用いた主語補完の選択方法の有効性を調査する必要がある。

9 おわりに

本研究では, 主語が省略されている日本語文に対し, 主語を付与することで, 翻訳精度の向上を目指した。まず, 主語が省略されている日本語文と主語を付与した日本語文に対し, それぞれ日英統計翻訳を行った。次に, 翻訳確率が最大となる出力文を選択した。そして, 選択した文を最終的な出力文とした。実験の結果, 人手評価において, 18 文の翻訳精度が向上し, 3 文の翻訳精度が低下した。また, 自動評価では, BLEU 値において 0.0044 向上し, 提案手法の有効性が確認できた。今後, 提案手法のより有効な選択方法を提案することで, 翻訳精度の向上を目指す。

参考文献

- [1] Richard Zens, Franz Josef Och, Hermann Ney “Phrase-based Statistical Machine Translation”, KI 2002, pp35-56, 2002.
- [2] 猪澤雅史, 村上仁一, 徳久雅人, 池原悟, “文節区切りの学習データを用いた, 日英統計翻訳の検討”, 言語処理学会年次大会発表論文集, pp1022-1025, 2010.
- [3] 中岩浩巳, 池原悟, “日英翻訳システムにおける用言意味属性を用いたゼロ代名詞照応解析”, 情報処理学会論文誌 34(8), pp1705-1715, 1993.
- [4] 中岩浩巳, 池原悟, “語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析”, 自然言語処理, Vol.3, No.4, pp50-65, 1996.
- [5] 中岩浩巳, “日英対訳コーパスを用いたゼロ代名詞とその指示対象の自動認定”, 自然言語処理研究会 (NL-123), pp33-40, 1998.
- [6] 中岩浩巳 “日英機械翻訳におけるゼロ代名詞照応解析に関する研究”, 博士論文, 2002.
- [7] 古市将仁, 村上仁一, 徳久雅人, 村田真樹, “日英統計翻訳における主語補完の効果”, 言語処理学会第 17 回年次大会, pp163-166, 2011.
- [8] 平 博順, 須藤 克仁, 永田 昌明, “統計翻訳における日本語省略補完の効果の分析”, 言語処理学会第 18 回年次大会, 2012.
- [9] 三上 章, “日本語の構文”, くろしお出版, 1993.
- [10] MeCab <http://mecab.sourceforge.net/>
- [11] Moses, moses-2010-08-13.tgz <http://www.statmt.org/moses/>
- [12] 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.
- [13] K.Papineni *et al.*: “BLEU: a method for automatic evaluation of machine translation”, 40th Annual meeting of the Association for Computational Linguistics, pp. 311-318, 2002.
- [14] B.Satanjeev *et al.*: “METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments”, ACL-2005, pp.65-72, 2005.
- [15] NIST : Automatic Evaluation of Machine Translation Quality Ising N-gram Co-Occurrence Statistics <http://www.itl.nist.gov/iad/mig/test/mt/>
- [16] RIBES : Rank-based Intuitive Bilingual Evaluation Measure <http://www.kecl.ntt.co.jp/icl/lirg/ribes/>