

統計翻訳における日本語省略補完の効果の分析

平 博順 須藤 克仁 永田 昌明

NTT コミュニケーション科学基礎研究所

{taira.hiroto, sudoh.katsuhito, nagata.masaaki}@lab.ntt.co.jp

1 はじめに

日本語の話し言葉では主語や目的語などが頻繁に省略される。日本語の話し言葉を統計翻訳を用いて英語等の省略が少ない言語に翻訳する場合、源言語の日本語における省略現象は、翻訳精度を低下させる原因の一つになると考えられる。日本語と同様の省略が多い中国語、韓国語に関して文献 [1] では、Chinese Treebank と Korean Treebank を対象として、英語側にあって中国語、韓国語側にはない要素（空要素）を補完した場合、中英、韓英のフレーズベース翻訳の翻訳精度が向上することが報告されている。また、日本語に関しては文献 [3] において、日本語の辞書例文を対象にして、日英翻訳における主語補完の有効性が報告されている。

一方、日本語の話し言葉翻訳に対する省略補完に焦点を当てた研究はこれまで少ないため、本研究では、既存の話し言葉コーパスに対し人手で省略補完を行ったデータを作成し、このデータを用いたフレーズベースの日英統計翻訳の実験を行って、省略補完を行う場合と行わない場合での精度の比較および分析を行った。この分析は、今後自動的に省略補完を行って統計の日英翻訳を行う際の精度向上の上限、精度向上に効果のある省略補完方法について知見を得ることを目的としている。

2 人手省略補完データの作成

2.1 本研究で分析対象とする省略現象

本研究では、日本語文中に含まれる述語に対する必須格で表層に現れなかったものを「省略」と見なし、その必須格を手手で補完し省略補完データを作成した。ただし、日本語での述語の必須格と任意格の区別は自明ではないため、日本語語彙大系 [7] の構文辞書に現れる用言の格を必須格と見なして省略補完したデータを作成した。なお今回は、述語そのものが省略されて

表 1: 省略補完の例

日本語原文：	連休中に帰ればいいので。
構文辞書：	(N1 が)(N2 から)(N3 に) 帰る
省略補完文：	exo1 が exo3 に exo3 から連休中に帰ればいいので。

いるような場合等、その他の省略現象については分析対象外とした。

2.2 省略補完データ作成方法

本研究で分析対象とした ADD データの日英対訳文は、直訳調でない訳も多く含まれていたため、英語を参照して日本語の補完をすることはせず、日本語側の述語に注目して省略補完を行った。本節では、「連休中に帰ればよいので。」という日本語文を例にして人手省略補完データの具体的な作成方法について説明する(表 1)。まず、原文に含まれる用言「帰る」に対して、日本語語彙大系の構文辞書に登録されている構文パターンを調べると「(N1 が)(N2 から)(N3 に) 帰る」というエントリがある。これにしたがって「～が」、「～から」、「～に」を必須格とみなして、省略要素の人称を表す省略タグ(exo_1、exo_2、exo_3、exo-ph)と助詞の組の省略補完を行う。ここでは、「～が」で省略されているものが一人称、「～から/より」および「～に/へ」で省略されているものが三人称の省略と考えられるので、対応する省略タグを exo_1、exo_3 を補完する。補完する際には、助詞も付与して補完を行う。ここで、exo_1 は一人称(私、私達、etc.) exo_2 が二人称(あなた、あなた達、etc.) exo_3 が三人称(それ、それら、etc.) exo-ph が他の文に含まれるフレーズまたは文を表す。

表 2: ADD データの例

日:	少々伺いたいことがあります。
英:	I have some questions to ask .
日:	はい、どのようなご用件でしょうか。
英:	Yes , what is it ?
日:	新婚旅行に海外に行きたいのですが。
英:	We'd like to go overseas for our honeymoon .

2.3 実験対象データ

実験用対訳コーパスとして、ADD データ (ATR Dialogue Database) [2] を用いた。このデータは、自動翻訳電話研究のために目的指向型の模擬会話が収集されたもので、旅行に関する旅行会社と客の対話、国際会議の申し込みに関する参加者と事務局の対話の二つのタスクのデータから構成されている。表 2 に ADD データに含まれる日本語、参照英訳の例を示す。

ADD データ中のキーボード会話の内、旅行会話 (日日) の前半約 1 万文、国際会議 (日日) 約 6000 文、国際会議 (日英) 約 6600 文を訓練データ、旅行会話 (日日) の後半約 1000 文をチューニング用データ、旅行会話 (日英) 約 1400 文をテスト用データとした。各データに対して、省略補完データを独自に人手で作成した。また、省略補完されていないデータを訓練データに追加した場合の傾向について調べるためのデータとして、IWSLT2007¹ の訓練データ約 4 万文も用いた。表 3 に我々が用いたデータセットの内訳を示す。

また、表 3 には、省略タグを含む文、すなわち省略補完を適用した文の数についても示してある。省略補完を行った文は、訓練データでは補完対象の ADD データ 21,888 文の内、13,835 文 (63.2%)、テストデータでは、1,402 文中 905 文 (64.5%) であった。構文パターンで示された格がすべて充足されている文であった場合、および「どのようなご用件でしょうか。」「さようなら」といった慣用的に使われる言い回しの場合に省略補完作業を行わなかった。

3 統計翻訳実験

3.1 実験の概略

今回の統計翻訳実験の手順の概略を図 1 に示す。まず、日本語文と対訳英文からなる学習用対訳データを

¹<http://iwslt07.fbk.eu/menu/tasks.html>

表 3: 実験対象データの内訳

	train	dev	test
ADD	21,988	1,129	1,402
IWSLT07	39,953	-	-
合計 (文ペア)	61,941	1,129	1,402
exo_1 を含む文	8,710	359	586
exo_2 "	6,581	260	411
exo_3 "	8,417	594	594
exo_ph "	1,122	58	38
省略を含む文全体	13,835	777	905

用意し、日本語文に対して人手で省略補完文を作成する。次に、省略補完文と対訳英文を入力として、フレーズベースの統計翻訳システム Moses [4] で学習を行い、翻訳モデルおよび言語モデルを得る。その後、同様に人手で省略補完を行った日本語文のテスト用データをデコーディングし、システム出力 (英語) を得る。システム訳とテスト用データの英語参照訳とを比較し翻訳精度 (BLEU 値 [6]) を求める。省略補完の効果については、省略補完を行わなかった場合の実験の結果との比較で行った。

3.2 実験設定

文の分かち書きについては、日本語文に対しては MeCab² を、英語文に対しては Moses 付属のトークナイザ³ を用いて行った。学習時は、学習に用いる文の長さ (単語長) の上限値を 80 に設定した他はデフォルトのパラメータを使用した。また、MERT [5] を用いてチューニングをした場合の実験も行ったが、再現性を保障するために今回の報告ではチューニングを行わない場合の結果について示した。基本的には、チューニングにより、省略補完の有無に関わらず、ほぼ同じような上がり幅での精度向上が見られた。

4 実験結果

4.1 訓練およびテストデータへの省略補完結果

まず、ADD データのみを使用し、訓練データおよびテストデータに省略補完を行った場合についての影響

²<http://mecab.sourceforge.net/>

³<http://www.statmt.org/moses/>

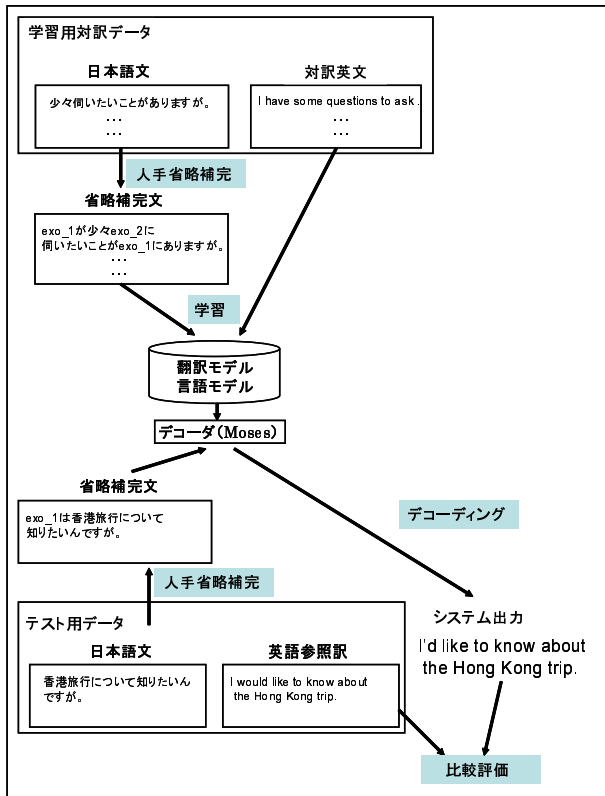


図 1: 翻訳実験の概略

を調べた。省略補完タグ付き ADD データの 1 人称タグ (exo_1) を「私」、2 人称タグ (exo_2) を「あなた」、3 人称タグ (exo_3) および exo_ph タグを「それ」に置き換えた上で mooses によりフレーズベース翻訳の学習を行った場合の結果を表 4 に示す。ここで「ベースライン」が省略補完前の原文の日本語文を使用した場合、「train, test 補完」は、省略補完を行ったデータを用いた場合の BLEU 値である。この結果からは、省略補完した場合の方が、0.0011 ポイント精度が向上しており、省略補完に一定の効果があることが伺えた。

表 4: 訓練およびテストデータへの省略補完効果

手法	BLEU
ベースライン	0.0834
train, test 補完	0.0845

4.2 テストデータのみへの省略補完結果

次に、テストデータのみについて省略補完データを使用した場合の省略補完の影響について調べた。古市らの主語補完の研究 [3] で、訓練データとテストデータ

両方に対して主語補完を行った場合より、テストデータのみ主語補完を行った方が精度向上が大きい、という報告があった。古市らの研究では、翻訳対象が辞書例文であることと、省略補完については人手による補完でなく、自動補完であるという違いがあるが、今回の話し言葉の翻訳に関しても同様の傾向があるかどうかについて調べた。

テストデータのみについて省略補完を行ったデータを使用した場合の結果を表 5 に示す。BLEU 値で比較する限り、テストデータのみに対して主語補完を行った場合、省略補完を行わない場合より翻訳精度が高くなっていた。また、訓練データにも省略補完を行った場合に比べても、精度の上昇幅が大きかった。

今回、省略補完を行う際、構文辞書に登録された構文パターンを用いて省略されている格を強制的に補完したが、補完されてきた文は、日本語として必ずしも自然な文ではないことが多い。そのため、訓練データに省略補完を行った場合、訓練データ中の学習がうまくいっている自然な文に対して、省略補完された文が悪影響を及ぼしているのではないかと考えられる。

表 5: テストデータのみでの省略補完結果

手法	BLEU
ベースライン	0.0834
test のみ補完	0.0850

4.3 訓練データを増やした場合の省略補完結果

一般的に、統計翻訳の精度を向上させるためには、訓練データの数を増やすことが重要である。省略補完をしない訓練データの数を増やしていても、省略補完データによる翻訳精度の向上が見られるか確認を行った。省略補完されていない訓練データとして、IWSLT2007 の訓練データ約 4 万文を追加して実験を行った。表 6 に

- 省略補完を行わない場合 (ベースライン)
- 訓練の一部 (ADD データ) とテストデータ両方に補完を行った場合
- テストデータのみ補完を行った場合

の 3 種類の実験の結果について示す。

その結果、まず省略補完を行わないベースラインにおいて、約 4 万文の訓練データの増加により、BLEU

値で、0.0088 ポイントの上昇が見られた。さらに、テストデータにのみ省略補完を行うことで、0.0034 ポイント BLEU 値が上昇した。しかし、訓練データにも省略補完データを入れた場合には、逆に精度が 0.0067 ポイント下がってしまった。この結果から、今回使用したデータに限って言えば、訓練データに対しては、単純に省略補完を行っても効果が無く、テストデータに対しては、省略補完により一定の精度向上の効果があることが分かった。

表 6: 訓練およびテストデータへの省略補完効果 (訓練データ追加)

手法	BLEU
ベースライン	0.0923
train, test 補完	0.0889
test のみ補完	0.0957

5 アラインメントに基づく分析

これまでに述べてきた省略補完の実験で、最も省略補完の効果があつた、テストデータにのみ省略補完を行って、フレーズベース翻訳を行った場合について、ベースラインの省略補完を行わなかった場合の翻訳結果と比較して、フレーズ翻訳のレベルで比較して、どのような場合に翻訳の質が向上しているのかについて、人手で分析を行った。テストデータのみを省略補完したときの結果について、テストデータのデコード結果からランダムに 50 サンプルを選び、人手で分析した結果を表 7 に示す。その結果、必ずしも省略補完によって翻訳の質が向上する訳ではないが、省略補完をすることによって、かえって翻訳の質が低下する場合は少ないことが分かった。

表 7: 翻訳の質の向上 (人手評価)

翻訳の質は良いまま	14
翻訳の質が向上	15
翻訳の質が悪いまま	18
翻訳の質が悪化	3
合計	50

6 おわりに

本研究では、日本語話し言葉において、用言の必須格とみなせる部分についての省略補完を行うことにより、どの程度、統計翻訳の精度向上が見込めるかについての評価を、人手で省略補完を行ったデータを用いた日英統計翻訳の実験により行った。その結果、省略補完により日英翻訳の精度向上が見込めるが、訓練データ、テストデータ両方に省略補完を行うのではなく、テストデータにのみ省略補完を行った方が、精度向上が見込める可能性が高いことが分かった。今後、省略補完により精度向上が見込める文のタイプについて解析を進める予定である。

参考文献

- [1] Tagyoung Chung and Daniel Gildea. Effects of empty categories on machine translation. In *Proc. of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010)*, pp. 636–645, 2010.
- [2] 江原暉将, 小倉健太郎, 森元逞. 電話対話データベースの構築. 情報処理学会 第 40 回全国大会講演論文集, pp. 486–487, 1990.
- [3] 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 発表論文集, pp. 163–166, 2011.
- [4] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. Moses: open source toolkit for statistical machine translation. In *Proc. of the 45th Annual Conference of the Association for Computational Linguistics (ACL-07), Demonstration Session*, pp. 177–180, 2007.
- [5] Franz Josef Och. Minimum error rate training for statistical machine translation. In *Proc. of the ACL*, 2003.
- [6] Kishore Papineni, Salim Roukos, Todd Ward, and WeiJing Zhu. BLEU: A method for automatic evaluation of machine translation. In *Proc. of the 40th Annual Conference of the Association for Computational Linguistics (ACL-02)*, 2002.
- [7] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系. 岩波書店, 1997.