

ニューラル日英機械翻訳における欠落語に関する分析

竹野 峻輔[†] 永田 昌明[‡] 山本 和英[†][†]長岡技術科学大学[‡]NTT コミュニケーション科学基礎研究所[†]{takeno, yamamoto}@jnlp.org, [‡]nagata.masaaki@ntt.co.jp

1 はじめに

ある対訳文での対応関係を考えたとき、単語対応する語をもう一方の言語中に持たない不対応語 (unaligned words) が存在する。例えば下記の対訳文においては英語側の代名詞 *me*, *you* (ゼロ代名詞) や限定詞 *the*, 日本語側の終助詞「ね」には明確な単語対応を決めることは難しい。

(日) 博物館 についたら 知らせ て ね。

(英) tell me when you get to the museum .

機械翻訳において、とりわけ問題となるのが目的側言語における不対応語である。これらの語は原言語側に明示的な手がかりがないため、原言語側の文脈情報または目的言語側の文法情報から間接的に生成する必要がある。本稿では、これを不対応語と区別して欠落語 (missing words) と呼ぶ。これまでこの欠落語について機械翻訳では、ゼロ代名詞や冠詞生成といった個別の形で研究がされてきた。[5, 11, 12, 1, 15, 14, 13]

IBM 翻訳モデルにおいて、原言語側の不対応語は fertility として単語ごと個別にモデル化されるのに対し、欠落語は NULL からの生成 (NULL-generated) として一元的にモデル化される。

句に基づく機械翻訳 (Phrase-Based Statistical Machine Translation; PBSMT) では、目的言語側は句テーブルに格納された単語しか生成できないので、欠落語 (主に冠詞) を含む句を動的に生成したり [13], 欠落語 (主にゼロ代名詞) に対応するトークンを原言語側に挿入することで解決が試みて来た [1, 15]。

近年の主流となったニューラル機械翻訳 (Neural Machine Translation; NMT) では、日本語と英語のような語順が大きく異なる言語対においても流暢な翻訳が得られるようになった。

しかしながら、NMT においてこの欠落語の問題が、どの程度解決されたかは現状において明らかではない。本稿ではこの調査のため先行研究において提案した翻訳文と参照訳の単語対応に基づいて欠落語を同定し、PBSMT と NMT における欠落語に関する翻訳誤りを定量的に分析を行ったのでこれを報告する。

2 注視型エンコーダデコーダモデル

近年の機械翻訳研究では、ニューラル機械翻訳手法の登場により、日英間のような語順構造が大きく異なるような言語間においても流暢な結果が得られるようになった。本節では、現在のニューラル機械翻訳において標準的なエンコーダ-デコーダモデルについて説明する。

現在のニューラル機械翻訳システムの標準的な実装であるエンコーダ・デコーダモデルでは、入力シンボルの系列 $x = x_0 \dots x_n$ に対して、出力シンボルの系列 $y = y_0 \dots y_n$, モデルパラメータを θ をしたとき、入力に対する出力系列は、以下の条件付き対数尤度を利用して表現される。

$$\log p(\mathbf{y}|\mathbf{x}) = \sum_{j=1}^m \log p(y_j|y_{<j}, \mathbf{x}; \theta) \quad (1)$$

入力系列 x は、このモデルにおいて非線形変換を介して語彙次元の 2 値ベクトルから隠れ層の次元数の埋め込み表現空間に写像される。この非線形変換のパラメータをエンコーダ・デコーダモデルは、下式の系列尤度を目的関数として、誤差逆伝播法によりモデルの最適化を行う。

$$J_t = \sum_{(x,y) \in D} \log p(\mathbf{y}|\mathbf{x}) \quad (2)$$

これまで様々なネットワークの形態が提案されているが、本稿では Luong et al. (2015) により提案された注視型エンコーダ・デコーダモデルを利用する [6]。注視型エンコーダ・デコーダモデルでは y_j を計算する際に、新たに注視 (attention) 層と呼ばれる層を設ける。この層は入力系列の単語 x_i におけるエンコーダの最終層の各出力 \tilde{h}_t とデコーダにおける y_{j-1} の埋め込み表現を連結させたものを入力としてとり、入力系列 x_i の重みを計算する。この出力は y_i の出力を行う時の入力系列の重要度を表すことから統計翻訳における単語対応の役割を持つと考えられているが、その性質は未だ十分に明きらかになっていない。特に単語対応がつけられない欠落語に関して、この注視機構が

どのように振る舞うかは不明瞭であり，明らかにする必要があると考えられる．

3 対訳コーパスに基づく欠落語検出

我々は対訳文中の欠落語の同定方法に関して，単語対応から欠落語を同定し，単語対応と言語モデルに基づいて原言語文の適切な位置に欠落語を配置する方法を先行研究において提案している [17]．本節ではこの欠落語の同定法を簡潔に説明する．

まず IBM モデル [7] に基づく双方向の単語対応を intersection による対称化を行い，単語対応を決定する．NULL に対する双方向の単語翻訳確率を利用し，次式のスコアにより欠落語のしやすさを計算する．ここで双方向の単語翻訳確率を利用するのは低頻度に対処するためである．

$$S_u(w) = p(e = w | f = NULL) * p(f = NULL | e = w)$$

目的言語の語彙において上式のスコアが高いものから上位 N 語を欠落語候補とする．話し言葉の日英対訳コーパスから獲得された欠落語の例を表 1 に示す．

日本語から英語への翻訳では，*a/an* や *the* といった限定詞に加え，特に話し言葉コーパスで問題になる日本語のゼロ代名詞に対応する英語の代名詞が欠落語に含まれることが分かる．このほか *it*，*there* などの虚辞や *would*，*will*，*could* などの助動詞も含まれる．同様に英日方向における欠落語には「が」，「を」，「に」などの格助詞，「か」，「ね」などの終助詞ますなどの丁寧語が含まれていることが分かる．

これらの欠落語候補を利用し，単語対応を用いて，欠落語候補に含まれ，かつ，原言語に対応する単語がない目的言語の単語を対訳文単位で特定し，これを欠落語とした．

表 1: 話し言葉対訳コーパスにおいて獲得された欠落語候補

日英翻訳における欠落語例						
'd	'll	'm	's	't	"	,(comma)
a	about	all	and	are	as	be
could	do	for	get	give	go	going
how	i	if	in	is	it	like
may	me	of	one	out	please	take
that	the	there	they	to	want	we
will	would	you	can	have	make	want

英日翻訳における欠落語例						
て	の	が	た	は	に	しれ
い	し	ん	まし	で	、	など
を	こと	も	ませ	と	。	枚
です	から	いる	？	ます	ない	よう
れ	する	なり	なっ	あり	つい	方
いい	な	か	ね	さ	ある	何
ば	人	だ	もの	選手	お	たら
年	ください	，	それ			

4 欠落語に関する調査

4.1 データセットとツール

統計的機械翻訳 (PBSMT) およびニューラル機械翻訳 (NMT) における欠落語の翻訳誤りを分析するため，IWSLT 2005 における日英機械翻訳のデータセットを用いた [2]．この対訳コーパスでは訓練データ，開発データおよび評価データとして，それぞれ 19,972 文対，500 文対，1,000 文対の対訳データが与えられる．小規模ではあるが，ゼロ代名詞をはじめとする欠落語を多く含む話し言葉の対訳コーパスであることから，今回の問題設定に適していると考えた．しかし，訓練データ数が約 2 万文対と少ないため調査のためには不十分であると考え，訓練データに 2 種類の話し言葉コーパスを追加して実験を行った．一つはストレートワード社¹により販売されている日常会話の用例日英対訳コーパス 50,709 文対である．もう一つはハルピン工業大学により作成された音声翻訳向けの中日英対訳コーパス 62,727 文対である [16]．これらを IWSLT2005 の訓練データに加えた合計 133,408 文対のうち，語数が 3 語以上 60 語未満の対訳文 (訓練データ 129,087 文，開発データ 493 文，評価データ 960 文) を実験に使用した．

前処理として，英語の単語分割器には Stanford Parser に付属の分割器²を用いた．日本語の単語分割には辞書に UniDic を利用した形態素解析器である MeCab³を用いた．単語対応の計算には GIZA++[8]を用いる．比較対象となる翻訳手法として，句に基づく統計的機械翻訳の標準的な実装である Moses[4]，および，注視型エンコーダデコーダモデル [6] に基づくニューラル機械翻訳の実装の一つである seq2seq-attn を用いた⁴．

実験においてハイパパラメータは標準設定を使用するが，NMT に関しては開発セットをもとに BLEU[9] が最大となる時点で学習を止めたモデルを使用する．欠落語を同定する際には，訓練，開発・評価のデータをひとまとめにして intersection により単語対応を求め，grow-diag-final-and による単語対応から単語翻訳確率の計算した．このときの欠落語候補は上位 100 単語の単語を対象とした．

自動評価尺度を用いて PBSMT と NMT の翻訳精度を比較した結果を表 2 に示す．自動評価尺度としては，最も標準的な BLEU，日英間で人手評価と相関が高いとされる RIBES[3]，および，後編集の作業量と相関があるとされる翻訳編集率 TER[10] の 3 種類を使用した．学習データは約 12 万文と少ないが，NMT

¹<http://www.straightword.jp/>

²<http://nlp.stanford.edu/software/lex-parser.shtml>

³<http://taku910.github.io/mecab/>

⁴<https://github.com/harvardnlp/seq2seq-attn>．現在は OpenNMT として開発が継続されている

が全ての自動評価尺度によって PBSMT を大きく上回ることがわかった。BLEU, RIBES, TER において、それぞれ+3.2 ポイント, +4.1 ポイント, -5.4 となることがわかった。次節以降では、これらのシステムの翻訳出力を用いて欠落語に関する誤り分析を行う。

表 2: IWSLT2005 日英機械翻訳における評価結果

method	IWSLT2005 拡張データ		
	BLEU	RIBES	TER
PBSMT	32.9	75.7	52.0
NMT	36.1	81.6	46.6

4.2 欠落語に関する誤り分析

欠落語に関する誤り分析では、まず翻訳編集率 TER を求めるソフトウェアを用いてシステム出力文から参照訳へ変換するのに必要な最小限の編集操作を求め、次に表 1 に示す欠落語候補ごとに編集操作の内容を集計する。結果を表 3 に示す。

TER における編集操作はシステム出力文に対して単語の削除と挿入、および参照訳中の単語への言い換え(置換)からなる。この内、欠落語が関与する編集操作は全体で NMT において 44.3%(1,440/3,248)、PBSMT において 44.1%(1,552/3,518)であり、平均的には PBSMT よりも NMT は欠落語の翻訳誤りは少ない。欠落語に関する翻訳誤りのうち、最も出現頻度が高いのは、代名詞、限定詞、前置詞などである点は両者とも共通している。

代名詞に関する翻訳誤りは、NMT においては全体の 33.9%(488/1,440) を占めている。PBSMT ではシステム出力文から代名詞が欠損する(代名詞の挿入)場合が多い(180件)のに対して、NMT においては100件にまで減少している。しかし、代名詞同士の言い換えを含む編集操作は NMT(268件)の方が PBSMT(254件)より多い。NMT は流暢な(文法的に正しい)文を出力することにより、代名詞の挿入や削除に関する誤りは減っているが、*i* から *you* への言い換えの人称の問題は減っていないことがわかった。

限定詞に関する翻訳誤りは代名詞の翻訳誤りの次に大きな割合を占める。PBSMT と比べると、NMT においては平均的には改善されているが(68件減少)、*a/an* と *the* の使い分け、*all* や *some/any* の使い分けの誤りは解決されていない。また前置詞に関する翻訳誤りは、絶対数において NMT と PBSMT に大差はないことも判明した。

以上よりニューラル機械翻訳では、統計的機械翻訳に比べて欠落語の問題は改善されているものの完全な

解決にいたっていないことがわかった。加えて、ゼロ代名詞のように原言語文における文法情報の認識の問題、代名詞の人称や限定詞の使い分けのように目的言語の文法事項に関する問題は、依然、課題として残っていることがわかった。

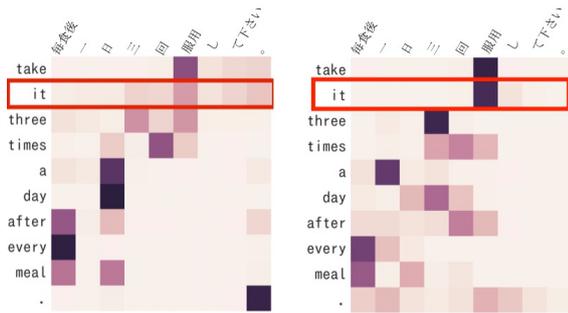
表 3: TER に基づく欠落語の誤り分析

	PBSMT	NMT
翻訳編集率	52.0	46.6
翻訳編集回数	3,518	3,248
欠落語に関する誤り	1,552	1,440
代名詞に関する翻訳誤り		
全体	488	526
代名詞の挿入	100	180
代名詞の言い換え	254	268
代名詞の削除	134	78
限定詞に関する翻訳誤り		
全体	262	312
限定詞の挿入	70	72
限定詞の言い換え	152	160
限定詞の削除	40	80
前置詞に関する翻訳誤り		
全体	192	270
前置詞の挿入	48	44
前置詞の言い換え	114	104
前置詞の削除	30	52
その他		
全体	498	444

4.3 ニューラル機械翻訳における欠落語の注視に関する調査

ニューラル機械翻訳モデルにおいては、注視機構により統計的機械翻訳における単語対応を実現しているとされている。しかし、そもそも入力文中の単語と単語対応がつかない出力文中の単語である欠落語について、注視層がどのような振る舞いをするかは明らかではない。ニューラル機械翻訳における注視機構は、文を生成する方向がモデルに組み込まれている点が、統計的機械翻訳における単語対応とは異なる。そこで我々は、目的言語を通常通りに学習した場合(順方向)と逆方向に並び替えて学習させた場合の2種類について注視機構の重みを観察した。図 1 に例を示す。

一般に、図 1(a) の単語 *it* のように、原言語側に対応する単語が存在しない目的言語側の単語(欠落語)は、



(a) 目的語順方向モデルの注視 (b) 目的語逆方向モデルの注視

図 1: 参照文に関する注視機構の振る舞いの一例

原言語側のある単語を中心として、重みを広く分散させる傾向にあることが分かった。文頭から単語を予測する場合、*take* の次に *it* を生成するための決定的な手がかりがないために、このような結果になったものと考えられる。

また欠落語に関しては、順方向と逆方向で重みの分布が大きく変化することがある。図 1(b) の単語 *it* は、なぜか逆方向では重みが *take* に集中している。おそらく文末から単語を予測する場合、*it* を生成する際に文脈情報が豊富にあるため、*it* の主辞である *take* に対応する単語「服用」に対して注視を働かせることができるのだと思われる。このように注視機構の振る舞いは方向性を持っており、特に欠落語に関するモデル化には改善の余地があると考えられる。

5 おわりに

機械翻訳における欠落語の問題について、句に基づく統計的機械翻訳手法と近年に主流であるニューラル機械翻訳の欠落語に関する分析を行った。

この結果としてニューラル機械翻訳では句構造で見られていたような主語の欠落の問題が大きく減少していることが分かった。ニューラル機械翻訳において出力される文は比較的流暢であるものの代名詞の人称誤りや冠詞をはじめとする限定詞の誤りの問題が依然として残っていることがわかった。

これらの問題は対訳文の単純な学習による機械翻訳システムでは取り扱いが難しく、文の前後の情報を利用した機械翻訳システムの実現が期待される。今後の予定として文脈情報を注入可能な機械翻訳モデルの検討を行っていきたいと考えている。

参考文献

[1] Tagyoung Chung and Daniel Gildea. Effects of Empty Categories on Machine Translation. In *Proceedings of the EMNLP2010*, pages 636–645, 2010.

[2] Matthias Eck and Chiori Hori. Overview of the IWSLT 2005 evaluation campaign. In *Proceedings of the IWSLT2005*, pages 1–22, 2005.

[3] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. Automatic evaluation of translation quality for distant language pairs. In *Proceedings of the EMNLP2010*, pages 944–952, 2010.

[4] Philipp Koehn and Hieu Hoang. Factored Translation Models. In *Proceedings of the EMNLP-CoNLL2007*, pages 868–876, 2007.

[5] Taku Kudo, Hiroshi Ichikawa, and Hideto Kazawa. A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation. In *Proceedings of the ACL2014*, pages 557–562, 2014.

[6] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. In *EMNLP2015*, 2015.

[7] Franz Josef Och. Minimum error rate training in statistical machine translation. In *Proceedings of the ACL2003*, pages 160–167, 2003.

[8] Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, 2003.

[9] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the ACL2002*, pages 311–318, 2002.

[10] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A study of translation edit rate with targeted human annotation. *Proceedings of Association for Machine Translation in the Americas*, 2006.

[11] Hirotoshi Taira, Katsuhito Sudoh, and Masaaki Nagata. Zero pronoun resolution can improve the quality of JE translation. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 111–118, 2012.

[12] Kristina Toutanova and Hisami Suzuki. Generating case markers in machine translation. In *Proceedings of the NAACL-HLT2007*, pages 49–56, 2007.

[13] Yulia Tsvetkov, Chris Dyer, Lori Levin, and Archana Bhatia. Generating English determiners in phrase-based translation with synthetic translation options. In *Processing of the 8th Workshop on Statistical Machine Translation*, pages 271–280, 2013.

[14] Longyue Wang, Zhaopeng Tu, Xiaojun Zhang, Hang Li, Andy Way, and Qun Liu. A Novel Approach to Dropped Pronoun Translation. In *Proceedings of the NAACL2016*, pages 983–993, 2016.

[15] Bing Xiang, Xiaoqiang Luo, and Bowen Zhou. Enlisting the ghost: Modeling empty categories for machine translation. In *Proceedings of the ACL2013*, pages 822–831, 2013.

[16] Muyun Yang, Hongfei Jiang, Tiejun Zhao, and Sheng Li. Construct trilingual parallel corpus on demand. In *Chinese Spoken Language Processing*, pages 760–767. Springer, 2006.

[17] 竹野 峻輔, 永田 昌明, and 山本 和英. 機械翻訳に向けた欠落語の投射によるオラクル入力文の作成. 116 (379):135–140, 2016.