

# マルチソースニューラル機械翻訳における 翻訳時の原言語欠落補完

西村 優汰<sup>1</sup> 須藤 克仁<sup>1</sup> Graham Neubig<sup>2,1</sup> 中村 哲<sup>1</sup>

<sup>1</sup>奈良先端科学技術大学院大学

<sup>2</sup>Carnegie Mellon University

{nishimura.yuta.nn9, sudoh, s-nakamura}@is.naist.jp  
gneubig@cs.cmu.edu

## 1 はじめに

機械翻訳の精度は対訳コーパスのデータ量に大きな影響を受けるため、十分な資源のない言語対では高精度な機械翻訳の実現が困難である。このような問題を解決するため、大規模コーパスのある言語対と組み合わせる翻訳を行う多言語機械翻訳手法 [1, 2, 3] が提案されている。特に、原言語が複数、目的言語が1つのものはマルチソース機械翻訳 (multi-source MT) と呼ばれ、単言語対で翻訳を行う手法よりも高い精度が得られている [4, 1, 5]。Multi-source MT では、欠落している対訳が存在しないコーパスからの学習を前提としているが、多言語の対訳文が存在するコーパスではある文に対して全ての言語に対応する文が揃っていない状況は非常に限定される。本研究では、このように複数言語の対訳コーパスにおいて対訳文が完全に揃っていないもの (本稿では欠落コーパスと呼ぶ) を活用する multi-source MT に着目し、そのためにニューラル機械翻訳 (NMT) の一手法である multi-encoder NMT を利用する。Multi-source MT における欠落の問題は学習時と翻訳 (テスト) 時の両方で起こり得る。学習時における問題は、欠落コーパスにおける欠落が存在する対訳文も学習に使用する手法として、欠落を特殊記号で置換する手法 [6] や学習済み multi-encoder NMT を用いた擬似対訳で置換する手法 [7] が提案され、その有効性が示されている。一方で、翻訳時における問題については未着手である。

そこで、本稿では、学習時ではなく翻訳時に multi-encoder NMT を利用することによって翻訳精度の向上を図る手法を提案する。具体的には、小資源言語対の翻訳精度向上を図るべく、大規模な対訳文が存在する中間言語を介して翻訳を行うピボット機械翻訳手法 [8] を参考にした multi-encoder NMT 手法を提案する。ピボット機械翻訳では、ある言語対においてその言語対で機械翻訳を行うよりも容易に機械翻訳が行えるような中間言語

を介する事によって翻訳精度を向上させている。複数ある原言語側に欠落がある入力文に対する multi-encoder NMT でも、欠落している言語を中間言語と考え、中間言語の擬似対訳文を生成し、欠落を補完することによって翻訳精度が向上すると考えられる。そこで本稿では、通常の単言語入力 of NMT によって欠落している言語の擬似対訳文の複数候補を生成し、multi-encoder NMT において最適となる擬似対訳文候補を選択して翻訳する手法を提案する。欠落が存在するテストセットを利用した実験においていくつかの欠落補完手法を比較し提案手法の有効性を示した。

## 2 Multi-encoder NMT

### 2.1 Multi-encoder NMT

Zoph ら [1] は、原言語を複数、目的言語を1つ (Many-to-One) 用いた multi-encoder NMT を提案している。この手法では、エンコーダを原言語の数、デコーダを1つ用いることで機械翻訳を行なっている。multi-encoder NMT を行うことで単言語対での機械翻訳よりも精度向上に有効であることが示されている。ここで、multi-encoder NMT におけるエンコーダとデコーダの接続部分について簡潔に述べる。原言語が2種類ある場合を考える。各エンコーダの最終 hidden 層をそれぞれ  $h_1, h_2$  とし、cell も同様に  $c_1, c_2$  とする。この時、デコーダの hidden 層の初期状態  $h$ , cell の初期状態  $c$  は次のように与えられる。

$$h = \tanh(W_c[h_1; h_2]) \quad (1)$$

$$c = c_1 + c_2 \quad (2)$$

また、Attention についても簡潔に述べる。タイムステップ  $t$  の時の各エンコーダの context vector を  $c_t^1, c_t^2$ , デコーダの hidden 層を  $h_t$  とした時のデコーダの softmax 層に送る前の最後の hidden 層  $\tilde{h}_t$  は次のように与えられる。

$$\tilde{h}_t = \tanh(W_c[h_t; c_t^1; c_t^2]) \quad (3)$$

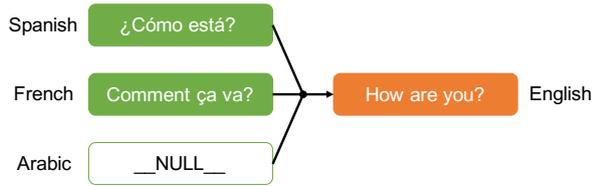


図 1: 欠落を特殊記号で置換した multi-encoder NMT: アラビア語の対訳文が欠落している。

## 2.2 欠落を考慮した multi-encoder NMT モデルの学習

Multi-encoder NMT の学習時に欠落コーパスを活用する手法について述べる。

**特殊記号による置換** 我々は, multi-encoder NMT の精度を向上させるために, 欠落コーパスであっても利用可能な対訳文は余すところなく学習に利用する手法を提案し, 有効性を示した [6]. 図 1 で示されているように, コーパス中の欠落している部分に欠落部分であることを示す特殊記号“\_NULL\_”を文の代わりに挿入することによって欠落部分を埋めている。

**Multi-encoder NMT による疑似対訳での補完** 我々は, 上記の手法を改良して, 学習済みの multi-encoder NMT を用いて疑似対訳を生成し, 欠落部を補完するという手法を提案し, 有効性を示した [7]. 図 2 を用いて, 簡単に説明を行う. この図における使用言語は英語, スペイン語, フランス語の 3 言語であり, 最終的な目的は, スペイン語の翻訳を得ることである. これは以下の 3 つの処理によって実現される. まず初めに, フランス語の疑似対訳を得るために multi-encoder NMT の学習 (原言語: 英語, スペイン語, 目的言語: フランス語) を行う. この時, 原言語側に欠落している対訳がある場合は, 特殊記号“\_NULL\_”で置換して学習を行う. 次に, コーパス中の欠落しているフランス語対訳を補完するために, 学習を行った multi-encoder NMT モデルで疑似対訳を生成し, 欠落を補完する. 最後に, 欠落を補完したコーパスを用いて, 新しく multi-encoder NMT の学習 (原言語: 英語, フランス語, 目的言語: スペイン語) を行う.

しかし, この手法では学習時の原言語側の欠落を補完するだけで, 最終的に翻訳したい文の原言語側に欠落がある場合には適用できないという問題点がある。

## 3 提案手法

最終的に翻訳したい文の原言語側に欠落がある場合, 我々の先行研究 [7] では, 欠落を疑似対訳で補完することができないという問題があった. そこで, multi-encoder NMT における原言語のみを用いた one-to-one NMT に

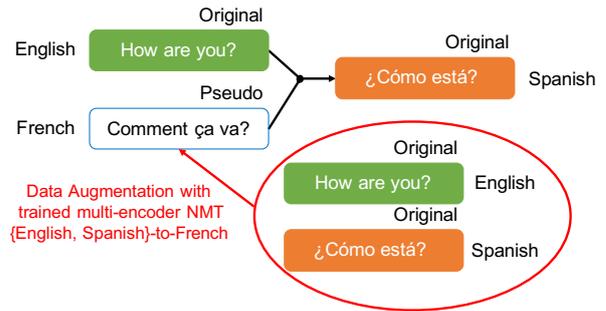


図 2: 欠落コーパスを用いた multi-encoder NMT における対訳文補完の例. この例では, フランス語の対訳文が欠落している。

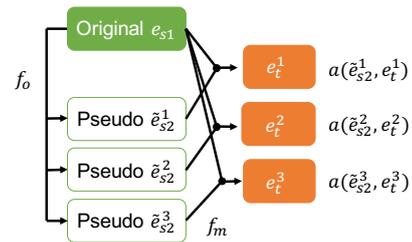


図 3: 複数の疑似対訳候補を考慮した multi-encoder NMT

よって疑似対訳の複数候補を生成し, 欠落を補完する最適な疑似対訳を選択する手法を提案する. 提案手法を図 3 を用いて説明する. ここで, 原言語の対訳文が  $e_{s1}$  と  $e_{s2}$  で, 目的言語の対訳文が  $e_t$  である multi-encoder NMT  $f_m$  を想定し, この時, 対訳文  $e_{s2}$  は欠落していると仮定している. まず初めに, 原言語が  $s_1$  であり, 目的言語が  $s_2$  である one-to-one NMT  $f_o$  の訓練を行う. 次に, 訓練を行った  $f_o$  を用いて欠落部分を補完するために, ビーム探索によって n-best の疑似対訳  $\tilde{E}_{s2} = \{\tilde{e}_{s2}^1, \dots, \tilde{e}_{s2}^n\}$  を生成する. この時, n-best の疑似対訳の文の生成確率を  $P_s = \{p(\tilde{e}_{s2}^1|f_o, e_{s1}), \dots, p(\tilde{e}_{s2}^n|f_o, e_{s1})\}$  とする. 生成した n-best の疑似対訳, それぞれを用いて multi-encoder NMT の出力文  $E_t = \{e_t^1, \dots, e_t^n\}$  を得る. Multi-encoder NMT の出力文の生成確率を  $P_t = \{p(e_t^1|f_m, e_{s1}, \tilde{e}_{s2}^1), \dots, p(e_t^n|f_m, e_{s1}, \tilde{e}_{s2}^n)\}$  とする. 最後に, 以下の式によって最終的な翻訳文  $e_t$  を決定する.

$$a(\tilde{e}_{s2}^n, e_t^n) = \varepsilon \ln p(\tilde{e}_{s2}^n|f_o, e_{s1}) + (1 - \varepsilon) \ln p(e_t^n|f_m, e_{s1}) \quad (4)$$

$$e_{s2}, e_t = \arg \max_{e_{s2}^n, e_t^n \in E_{s2}, E_t} a(\tilde{e}_{s2}^n, e_t^n) \quad (5)$$

また, 式 4 中の  $\varepsilon$  はハイパーパラメータであり, one-to-one NMT を用いた疑似対訳の出力と multi-encoder NMT の出力のどちらを重視するかを調整するためのものである. 本提案手法によって, 様々なパターンの疑似対訳候補が考慮されることが期待される。

表 1: 訓練文数, 欠落文数, テストにおける文数. 各目的言語において, “train”は訓練文数, “missing” は各言語の組み合わせの中で欠落が存在する言語における欠落している訓練文数と欠落度合い, “test” は欠落が存在するテストの対訳文数を示している.

Pair	Trg	train	missing	test
en-hr/sr	hr	115127	34116 (29.6%)	1145
	sr	129461	48450 (37.4%)	896
en-sk/cs	sk	58109	16772 (28.9%)	602
	cs	97488	56151 (57.6%)	1966
en-vi/id	vi	150829	81945 (54.3%)	1405
	id	77936	9052 (11.6%)	333

## 4 実験

翻訳時の原言語側の欠落を擬似対訳で補完する手法として提案手法が有効であることを示すため, 欠落が存在するテストセットにおいていくつかの欠落補完手法を比較する実験を行った.

### 4.1 実験データ

本実験には, TED talks の多言語コーパスを用いた. この多言語コーパス中におけるそれぞれの対訳文の量は言語によって大きく異なる. 本実験では, 以下に示す 7 つの言語セットを用いた.

- 英語 (en), クロアチア語 (hr), セルビア語 (sr)
- 英語 (en), スロバキア語 (sk), チェコ語 (cs)
- 英語 (en), ベトナム語 (vi), インドネシア語 (id)

また, TED talks のコーパスから 1 文の長さが 40 語より少ない文を抽出し実験に使用した. TED talks では, 原則として英語の講演が元になっているので, 必然的に英語は原言語の一つとなり, 各言語対の中で英語以外の言語をそれぞれ目的言語, 原言語の 1 つとした. それぞれの言語対, 目的言語で使用可能な訓練文数, テスト文数を表 1 に示す.

### 4.2 実験設定

実験に使用した NMT モデルの詳細を以下に示す. NMT モデルは, Luong ら [9] によって提案された Global Attention と Input Feeding を使用し, さらにエンコーダでは, Bahdanau らの手法 [10] に使用された Bidirectional Encoder を使用した. hidden 層と embed 層のユニット数はそれぞれ 512 とした. Multi-encoder NMT の学習手法としては, 我々の先行研究 [7] を用いた. SentencePiece [11] を用いてサブワード分割を行い, 各言語対における全ての言語の訓練文をまとめて統一のサブワードモデルを作成し, サブワード語彙数は 16000 とした. モデルの

最適化のアルゴリズムとして Adam を使用し, gradient clipping を 5 に設定した. 評価手法として, BLEU [12] を使用し, 評価ツールとして SacreBLEU<sup>1</sup> [13] を用いた. 開発データでの Log Perplexity が最小となった時点でのパラメータを保存し, テストデータで評価した. 提案手法における n-best を本実験では 5-best とした.

### 4.3 Baseline 手法

提案手法との比較として, 以下に示す 2 つの baseline 手法との比較を行った. まず 1 つ目が, 原言語を英語とした one-to-one NMT である. 次に 2 つ目が, 提案手法と同様に補完を行うが, 複数候補でなくビーム探索による 1-best のみを利用する multi-encoder NMT である.

### 4.4 実験結果

表 2 に実験結果を示す. 表 2 中の “proposed” における BLEU は, それぞれの目的言語において最適なハイパーパラメータ  $\epsilon$  を用いた文から算出したものである. 最適なハイパーパラメータ  $\epsilon$  は, 0 から 1 の間で値を 0.05 ずつ変えていき, BLEU が最大となるものとした. まず, 提案手法は, 1-best の擬似対訳で欠落を補完した時よりも高い翻訳精度を得られていることがわかる. このことから, one-to-one NMT における最適な出力結果が multi-encoder NMT に用いる対訳文の欠落を補完する文として必ずしも最適ではないことがわかり, また, 提案手法は n-best の出力結果からどの出力文が欠落を補完するのに適切であるかを選択する手法として有効性を示すことができた. しかし, 目的言語がセルビア語とチェコ語の場合では, 原言語が英語の one-to-one NMT における BLEU よりも提案手法における BLEU の方が低いが, 表 1 において各言語対内でセルビア語とチェコ語の訓練文数はクロアチア語とスロバキア語の訓練文数より多いことから, 提案手法は各言語対内で訓練文数が少ない言語に対して有効に働くのではないかと考えられる. また, 言語対が en-sk/cs の場合は one-to-one NMT と proposed method での BLEU の差が大きい, 言語対が en-hr/sr の場合は差が小さいことがわかり, 表 1 からスロバキア語とチェコ語での訓練文数の差は大きい, クロアチア語とセルビア語での訓練文数の差は小さいことがわかることから, 言語対での各目的言語における訓練文数の違いが大きいほど, 提案手法は有効に働くと考えられる.

### 4.5 分析

式 5 で定義したように, 提案手法ではハイパーパラメータ  $\epsilon$  を用いている.  $\epsilon$  の値によって翻訳精度がどのよう

<sup>1</sup>[https://github.com/aws-labs/sockeye/tree/master/sockeye\\_contrib/sacrebleu](https://github.com/aws-labs/sockeye/tree/master/sockeye_contrib/sacrebleu)

表 2: BLEU による実験結果

Pair	Trg	Baseline		Proposed (5-best)
		one-to-one (En-to-Trg)	1-best	
en-hr/sr	hr	22.58	22.55	<b>22.62</b>
	sr	<b>16.38</b>	15.71	16.17
en-sk/cs	sk	14.16	16.57	<b>16.91</b>
	cs	<b>15.13</b>	13.63	13.89
en-vi/id	vi	22.62	22.96	<b>23.42</b>
	id	26.41	26.23	<b>26.97</b>

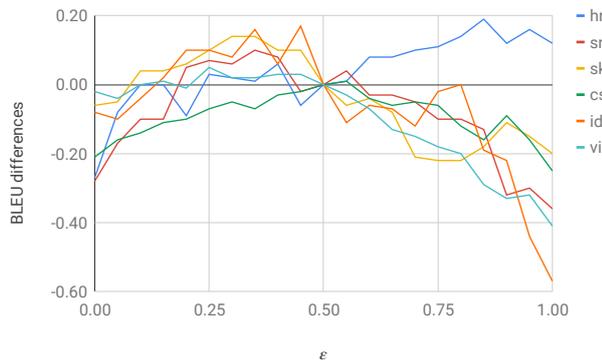


図 4:  $\epsilon$  を変化させた時の BLEU の変化.  $\epsilon$  が 0.5 である時の BLEU を基準とし, 各  $\epsilon$  での BLEU との差分を示している.

に変化するかを調べるために,  $\epsilon$  を 0.05 ずつ変化させた時の BLEU の変化を調査した. 図 4 は, それぞれの目的言語において  $\epsilon$  を変化させた時の BLEU の変化を示している. この図では,  $\epsilon$  が 0.5 である時の BLEU を基準とし, 各  $\epsilon$  での BLEU との差分を示している. 目的言語がチェコ語, クロアチア語以外の全ての言語では  $\epsilon$  が 0.5 以下である時の方が高い BLEU を得られていることから, これらの目的言語では multi-encoder NMT での出力文の生成確率の方がより重要であることがわかる. また, one-to-one NMT での出力文の生成確率がより重要であるクロアチア語では, 提案手法と 1-best で擬似対訳を生成した時の BLEU の差が他の言語対に比べて小さいことが表 2 からわかる. このことから, multi-encoder NMT での出力文の生成確率の方がより重要である場合の方が, 提案手法がより有効に働くと考えられる.

## 5 おわりに

本研究では, 多言語の欠落コーパスを翻訳時に活用する multi-encoder NMT に着目し, 通常の単言語入力による NMT によって欠落している言語の擬似対訳文の複数候補を生成し, multi-encoder NMT において最適となる擬

似対訳文候補を選択して翻訳する手法を提案し, 実験により有効性を示した. しかし, 言語の組み合わせや訓練文数の違い, 欠落の度合いなどによって, 結果にばらつきがあるので更なる調査が必要である.

## 謝辞

本研究の一部は JSPS 科研費 JP16H05873 と JP17H06101 の助成を受けたものである.

## 参考文献

- [1] Barret Zoph and Kevin Knight. Multi-Source Neural Translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 30–34, San Diego, California, June 2016. Association for Computational Linguistics.
- [2] Daxiang Dong, Hua Wu, Wei He, Dianhai Yu, and Haifeng Wang. Multi-Task Learning for Multiple Language Translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pages 1723–1732, Beijing, China, July 2015. Association for Computational Linguistics.
- [3] Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Vigas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics*, vol. 5, pages 339–351, 2017.
- [4] Franz Josef Och and Hermann Ney. Statistical Multi-Source Translation. In *Proceedings of the eighth Machine Translation Summit (MT Summit VIII)*, pages 253–258, September 2001.
- [5] Ekaterina Garmash and Christof Monz. Ensemble Learning for Multi-Source Neural Machine Translation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1409–1418, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.
- [6] Yuta Nishimura, Katsuhito Sudoh, Graham Neubig, and Satoshi Nakamura. Multi-Source Neural Machine Translation with Missing Data. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 92–99. Association for Computational Linguistics, July 2018.
- [7] Yuta Nishimura, Katsuhito Sudoh, Graham Neubig, and Satoshi Nakamura. Multi-source neural machine translation with data augmentation. In *15th International Workshop on Spoken Language Translation (IWSLT)*, Bruges, Belgium, October 2018.
- [8] Hua Wu and Haifeng Wang. Pivot language approach for phrase-based statistical machine translation. *Machine Translation*, 21(3):165–181, 2007.
- [9] Thang Luong, Hieu Pham, and Christopher D. Manning. Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal, September 2015. Association for Computational Linguistics.
- [10] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of the 3rd International Conference on Learning Representations*, May 2015.
- [11] Taku Kudo. Subword regularization: Improving neural network translation models with multiple subword candidates. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 66–75. Association for Computational Linguistics, 2018.
- [12] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–318, Philadelphia, July 2002.
- [13] Matt Post. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, WMT 2018, Belgium, Brussels, October 31 - November 1, 2018, pages 186–191, 2018.