

# 依存構造木に基づく機械翻訳における柔軟な非終端記号の挿入位置の選択モデル

Insertion Position Selection Model for Flexible Non-Terminals in Dependency Tree-to-Tree Machine Translation

京都大学大学院情報学研究科 **中澤 敏明**

2010年京都大学大学院情報学研究科知能情報学専攻博士課程修了。博士（情報学）。機械翻訳の研究に従事。

✉ nakazawa@nlp.ist.i.kyoto-u.ac.jp ☎ 075-753-5346

京都大学大学院情報学研究科教授 **黒橋 禎夫**

1994年京都大学大学院工学研究科電気工学第二専攻博士課程修了。博士（工学）。2006年4月より京都大学大学院情報学研究科教授。自然言語処理、知識情報処理の研究に従事。

## 1 はじめに

木構造を利用した翻訳モデルは非終端記号を用いることで目的言語の語順を表現することができるため、単語列による翻訳モデルで必須である単語並べ替えモデルが不要である。例えば図1に示すような翻訳ルールを与えられたとすると、「公園 - park」を非終端記号に置き換えることで、「公園」の代わりに「図書館」がある新たな入力文を翻訳することは非常に容易である。

一方で、訓練対訳文に存在しない単語や句（部分木）を挿入する必要がある場合は問題である。本稿ではこの

ような部分木を浮動部分木と呼ぶ。例えば図1のように入力文に「突然」が存在する場合、この浮動部分木をどのように翻訳したら良いかを図1の翻訳ルールは知らないため、このルールをそのまま使うことができない。

この問題を解決するために、これまでの研究では、糊ルールを使う方法<sup>[1]</sup>や依存構造木の形に制約を与える方法<sup>[2]</sup>などが提案されてきた。Richardsonら<sup>[3]</sup>は浮動部分木に対して、可能性のある挿入箇所を全て考慮する「柔軟な非終端記号」を導入することでこの問題を解決した。可能性のある挿入箇所は以下の制約を満たす：

- ・可能性のある挿入箇所は元言語側の浮動部分木の親に対応する目的言語側の単語の子でなければならない
- ・可能性のある挿入箇所は依存構造木上で交差してはならない

例えば図1では浮動部分木「突然」に対する可能性のある挿入箇所が灰色の矢印で示されている。「突然」は「電話する」の子供なので、可能性のある挿入箇所も「電話する」の対応先である“called”の子供でなければならない。

柔軟な非終端記号は浮動部分木を扱うのに非常に強力であり、翻訳精度も大きく向上すると報告されているが、翻訳時の計算コストが増大するという問題がある。我々の実験では柔軟な非終端記号を利用することで、翻

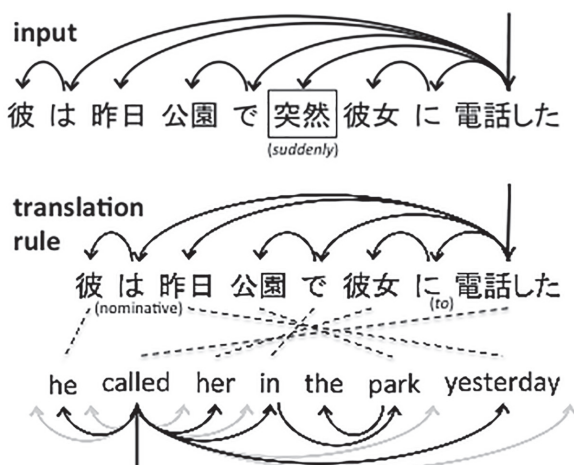


図1 入力文と翻訳ルールの例

訳にかかる時間が3から6倍になった。これは最適な挿入箇所がデコーディング時に動的に選択されるためである。しかし我々は両言語の木構造の情報を利用することで、あらかじめ可能性のある挿入箇所を減らし、さらには1つに限定してしまうことも可能であると考え、これを行う手法を提案する。提案する手法により、柔軟な非終端記号を利用して計算コストを抑えることができ、翻訳にかかる時間が著しく長くなることを避けることができるだけでなく、探索空間が小さくなることでより最適な解が見つかりやすくなり、翻訳精度の向上も見込める。

## 2 挿入箇所の選択

最適な挿入箇所は、挿入される単語 ( $I$ ) とその原言語側と目的言語側の文脈を考慮することで選択できると仮定した。原言語側の文脈として  $I$  の親の単語 ( $P_s$ ) と、 $I$  と  $P_s$  との距離 ( $D_s$ ) を利用する。目的言語側の文脈としては各挿入箇所の親の単語 ( $P_t$ ) と、挿入箇所と  $P_t$  との距離 ( $D_t$ )、挿入箇所の前 ( $S_p$ ) と後ろ ( $S_n$ ) の兄弟の単語を利用する。距離は単語列上の単語数ではなく、依存構造木上の兄弟の数で計算し、挿入箇所が親の左側ならば正の値、右側ならば負の値とする。図1の“park”と“yesterday”の間の挿入箇所を例にとると、 $I$  = 「突然」、 $P_s$  = 「電話した」、 $D_s$  = +2、 $S_p$  = “park”、 $S_n$  = “yesterday”、 $P_t$  = “called”、 $D_t$  = -3となる。これらの手がかりをニューラルネットワークモデルの入力とし、挿入箇所の選択問題を解く。

### 2.1 ニューラルネットワークモデル

図2に挿入箇所選択問題を解くためのニューラルネットワークモデルを示す。 $k$  番目の挿入箇所候補に対して、手がかりとなる単語 ( $I$ 、 $P_s$ 、 $P_t$ 、 $S_p^k$ 、 $S_n^k$ ) は3つの埋め込み層 (embedding layer) によりベクトル表現に変換される。3つの層はそれぞれ表層語 (200次元)、品詞 (10次元)、係り受けタイプ (10次元) の埋め込みを行い、合わせて220次元のベクトルを生成する。これらの単語とそれらの距離を使って、原言語側と目的言語側の文脈ベクトル  $c_s^k$  と  $c_t^k$  を生成する。次に各言語の文脈ベクトルから挿入箇所候補に対する文脈ベクトル  $c_i^k$  を生成する。最後に挿入箇所候補のスコ

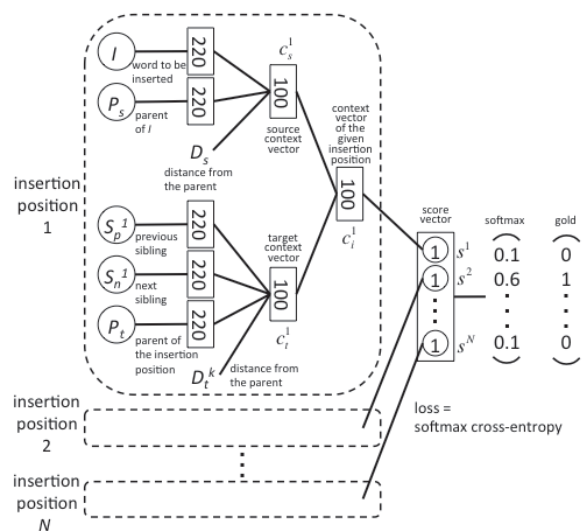


図2 挿入箇所選択のためのニューラルネットワークモデル (四角内の数字はベクトルの次元数)

ア  $s^k$  を  $c_i^k$  より計算する。各ベクトルは以下の式で計算される：

$$\begin{aligned}
 c_s^k &= \tanh(W_{c_s} [I; P_s; D_s]) \\
 c_t^k &= \tanh(W_{c_t} [S_p; S_n; P_t; D_t^k]) \\
 c_i^k &= \tanh(W_{c_i} [c_s^k; c_t^k]) \\
 s^k &= W_s c_i^k
 \end{aligned}$$

ここで “;” はベクトルの結合を意味する。 $c_s^k$ 、 $c_t^k$ 、 $c_i^k$  のサイズはそれぞれ100次元とした。

同一のネットワークを全ての挿入箇所候補に適用し、それぞれの挿入箇所のスコアを得る。これらのスコアをソフトマックス関数により正規化し、交差エントロピー誤差関数により誤差を計算する。各層は全結合のフィードフォワードネットワークで、学習時には50%ドロップアウトを用いた。

### 2.2 訓練データの生成

ニューラルネットワークモデルを訓練するためのデータは、単語アライメント済みの対訳コーパスからアルゴリズム1により自動的に生成することができる。ALIGNMENT関数は与えられた原言語の単語に対応する目的言語の単語を返し、ISPARENTCHILD関数はPtがCtの親ならば真を返し、そうでなければ偽を返す。

GENERATEDATA関数は  $P_s$ 、 $C_s$ 、 $P_t$  とそれぞれの文脈を使って  $C_t$  の位置を予測するような一つの訓練データを、 $C_t$  を目的言語の依存構造木から取り除くことで生成する。もともと  $C_t$  があった位置が正解となり、それ以外は不正解となる。なお  $C_s$  は図2の  $I$  に相当する。

### アルゴリズム 1 訓練データの生成

```

for all  $P_s \in$  words in source tree do
   $P_t =$  ALIGNMENT( $P_s$ )
  for all  $C_s \in$  CHILDREN( $P_s$ ) do
     $C_t =$  ALIGNMENT( $C_s$ )
    if ISPARENTCHILD( $P_t$ ,  $C_t$ ) then
      GENERATEDATA( $P_s$ ,  $C_s$ ,  $P_t$ ,  $C_t$ )
    end if
  end for
end for
end for

```

## 2.3 翻訳における挿入箇所の選択

訓練されたニューラルネットワークモデルを使うことで翻訳ルール内の浮動部分木に対する各挿入箇所候補のスコアが計算でき、最もスコアが高いものを最適な挿入箇所として選択することができる。ここで、多くの場合、翻訳ルールには元の対訳依存構造木の一部しか含まれていない。つまり挿入箇所の選択に利用できる文脈情報は、ニューラルネットワークモデルの訓練に利用したものと異なることになる。例えば図 1 において入力文に「公園で」がないとすると、挿入箇所候補数は 6 となり、文脈として “in” が使えないことになる。しかしながら翻訳における挿入箇所選択に利用できる文脈と似たような文脈や全く同じ文脈は他の対訳文で出現する可能性が高いと考えられるため、大きな問題ではない。

## 3 実験

ASPEC<sup>[4]</sup> を利用して、挿入箇所選択と翻訳の 2 つの実験を行った。ASPEC に含まれる文の数を表 1 に示す。日本語文の解析には JUMAN<sup>[5]</sup> と KNP<sup>[6]</sup> を用いた。英語文の解析はまず nlparsen<sup>[7]</sup> を用いて句構造に変換し、人手で整備したルール<sup>[8]</sup> により依存構造木に変換

した。中国語の解析には KKN<sup>[9]</sup> と SKP<sup>[10]</sup> を利用した。単語アライメントには教師あり単語アライメントツール Nile<sup>[11]</sup> を用いた。

依存構造木に基づいた翻訳デコーダーとして KyotoEBMT<sup>[12]</sup> を用いた。ニューラルネットワークモデルの構築と訓練には Chainer<sup>[13]</sup> を用いた。

表 1 : ASPEC の文数

	日 ⇄ 英	日 ⇄ 中
訓練	2,020,106	667,520
開発	1,789	2,115
テスト	1,812	2,174

## 3.1 挿入箇所選択実験

ニューラルネットワークモデルの訓練データと開発データ及びテストデータは 2.2 節の方法により自動的に生成した。ASPEC より生成されたデータの大きさと、不動部分木ごとの挿入箇所候補数の平均を表 2 に示す。トレーニングは 100 エポック行い、開発データにおいて最も精度が高かったパラメータをテストに利用した。

テストデータにおける平均誤差および精度を表 2 に示す。結果より、提案するモデルは非常に高精度に正しい挿入箇所を選択できていることがわかる。目的言語が日本語の場合の精度が逆方向の場合の精度よりも高くなっているが、これは日本語が head-final 言語であり、ほとんどの場合、単語は左から右にかかるため、問題が簡単なためであると考えられる。

テストデータにおいては誤りとされている挿入箇所を選択した場合であっても、文法的に正しく、文の自然さとしても問題ない場合もあると考えられるため、実際の精度はこれよりも高いと推定される。例えば挿入箇所のスコア上位 2 つまでの精度を見ると、目的言語が日本語の場合は 99.5%、それ以外の場合でも 99.0% 程度となる。

表 2 挿入箇所選択実験のデータサイズおよび実験結果

	日 → 英	英 → 日	日 → 中	中 → 日
訓練	15.7M		5.7M	
開発	160K		58K	
テスト	160K		58K	
挿入箇所の平均数	3.39	3.15	3.72	3.41
最適 epoch 数	89	71	61	79
平均誤差	0.089	0.058	0.105	0.056
精度 (%)	97.08	97.72	96.51	97.99

### 3.2 翻訳実験

以下の3つの設定で、ASPECを用いて翻訳実験を行った：

- ・設定1：柔軟な非終端記号を使わず、単純な糊ルールを使って用例を結合する
- ・設定2：柔軟な非終端記号を使うが、挿入箇所を選択は行わない（ベースライン）
- ・設定3：ニューラルネットワークモデルにより最適と判定された挿入箇所のみを使う（提案手法）

翻訳の精度は自動評価尺度であるBLEU<sup>[14]</sup>とRIBES<sup>[15]</sup>によって評価し、有意差検定も行った<sup>[16]</sup>。RIBESはBLEUよりも語順の正しさに敏感であり、提案手法も語順の正しさにより効果があると考えられるため、RIBESの向上が期待される。また設定1を基準とした設定2および3の相対的な翻訳時間も調べた。

実験結果を表3に示す。結果を見ると、英日翻訳のBLEUスコアを除いて、すべての言語対において提案手法の翻訳精度がベースラインと比較して有意に向上した。また翻訳にかかる時間も、ベースラインと比較しておよそ60%低減された。

## 4 まとめと今後の課題

本稿では依存構造木に基づいた機械翻訳において柔軟な非終端記号を用いる際に計算コストが大きくなってしまいう問題に対して、ニューラルネットワークモデルを利用して最適な挿入箇所をあらかじめ決定することでデコード時の探索空間を小さくする手法を提案した。提案したモデルは最適な挿入箇所を高精度で選択することができ、これにより翻訳にかかる時間の低減および翻訳精度の向上を達成することができた。

現在は文脈として周辺の単語のみを用いているが、その単語を根とした部分木全体の情報を利用することで、より精度が向上する可能性がある。例えば図1において“in the park”という部分木全体の情報を用いる方が

“in”のみを用いるよりも良いと考えられる。これは特に目的言語が日本語の場合に重要である。なぜなら日本語では動詞の子供は多くの場合格助詞であるが、格助詞の情報だけでは最適な挿入箇所の選択には不十分であると考えられるからである。部分木全体をベクトル表現に変換する手法としては、[17]や[18]などの既存手法が利用できる。これは今後の課題である。

[1] David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL' 05), pages 263-270. Association for Computational Linguistics.

[2] Libin Shen, Jinxi Xu, and Ralph M Weischedel. 2008. A new string-to-dependency machine translation algorithm with a target dependency language model. In Association for Computational Linguistics.

[3] John Richardson, Fabien Cromie`res, Toshiaki Nakazawa, and Sadao Kurohashi. 2016. Flexible non-terminals for dependency tree-to-tree reordering. In NAACL.

[4] Toshiaki Nakazawa, Manabu Yaguchi, Kiyotaka Uchimoto, Masao Utiyama, Eiichiro Sumita, Sadao Kurohashi, and Hitoshi Isahara. 2016. ASPEC: Asian scientific paper excerpt corpus. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016), pages 2204-2208, Portoroz, Slovenia, May.

表3 翻訳実験結果。\*は設定2と比較して設定3が有意 ( $p < 0.01$ ) に良いことを示す。

	日 → 英			英 → 日			日 → 中			中 → 日		
	BLEU	RIBES	Time	BLEU	RIBES	Time	BLEU	RIBES	Time	BLEU	RIBES	Time
設定1	20.28	65.08	1.00	28.77	75.21	1.00	24.85	66.60	1.00	30.51	73.08	1.00
設定2	21.61	69.82	6.28	<b>30.57</b>	76.13	3.30	28.79	78.11	5.16	34.32	77.82	5.28
設定3	<b>22.07*</b>	<b>70.49*</b>	2.25	30.50	<b>76.69*</b>	1.27	<b>29.83*</b>	<b>79.73*</b>	2.21	<b>34.71*</b>	<b>79.25*</b>	1.89



- [5] Sadao Kurohashi, Toshihisa Nakamura, Yuji Matsumoto, and Makoto Nagao. 1994. Improvements of Japanese morphological analyzer JUMAN. In Proceedings of The International Workshop on Sharable Natural Language, pages 22-28.
- [6] Sadao Kurohashi and Makoto Nagao. 1994. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. *Computational Linguistics*, 20(4):507-534.
- [7] Eugene Charniak and Mark Johnson. 2005. Coarse-to- fine n-best parsing and maxent discriminative reranking. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL' 05), pages 173-180.
- [8] Michael Collins. 1999. Head-Driven Statistical Models for Natural Language Parsing. Ph.D. thesis, University of Pennsylvania.
- [9] Mo Shen, Hongxiao Liu, Daisuke Kawahara, and Sadao Kurohashi. 2014. Chinese morphological analysis with character-level pos tagging (short paper). In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014), Baltimore, USA.
- [10] Mo Shen, Daisuke Kawahara, and Sadao Kurohashi. 2012. A reranking approach for dependency parsing with variable-sized subtree features. In Proceedings of 26th Pacific Asia Conference on Language Information and Computing, pages 308-317.
- [11] Jason Riesa, Ann Irvine, and Daniel Marcu. 2011. Feature-rich language-independent syntax-based alignment for statistical machine translation. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 497-507. Association for Computational Linguistics.
- [12] John Richardson, Fabien Cromieres, Toshiaki Nakazawa, and Sadao Kurohashi. 2014. KyotoEBMT: An example-based dependency-to-dependency translation framework. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 79-84. Association for Computational Linguistics.
- [13] Seiya Tokui, Kenta Oono, Shohei Hido, and Justin Clayton. 2015. Chainer: a next-generation open source framework for deep learning. In Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS).
- [14] Kishore Papineni, Salim Roukos, Todd Ward, and WeiJing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In ACL, pages 311-318.
- [15] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010. Automatic evaluation of translation quality for distant language pairs. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10



10, pages 944-952, Stroudsburg, PA, USA.  
Association for Computational Linguistics.

[16] Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In Dekang Lin and Dekai Wu, editors, Proceedings of EMNLP 2004, pages 388-395, Barcelona, Spain, July. Association for Computational Linguistics.

[17] Yang Liu, Furu Wei, Sujian Li, Heng Ji, Ming Zhou, and Houfeng WANG. 2015. A dependency-based neural network for relation classification. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 285-290. Association for Computational Linguistics.

[18] Lili Mou, Hao Peng, Ge Li, Yan Xu, Lu Zhang, and Zhi Jin. 2015. Discriminative neural sentence modeling by tree-based convolution. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2315-2325, Lisbon, Portugal, September. Association for Computational Linguistics.