

句ベース統計翻訳における構文情報の効果

今村 賢治, 大熊英男, 隅田英一郎

ATR 音声言語コミュニケーション研究所

{kenji.imamura, hideo.okuma, eiichiro.sumita}@atr.jp

1 はじめに

Brown et al. (1993) によって提案された統計翻訳も、単語を単位とした翻訳モデルから、句 (複数単語列) を単位とした翻訳モデルに移行しつつある (Koehn et al., 2003; Vogel et al., 2003; Zens and Ney, 2004)。これを句ベース統計翻訳 (Phrase-based SMT) と呼ぶ。句ベース統計翻訳の場合、句の内部での語順は局所的に変更されているため、語順調整のコストは低くなる。しかし、句そのものの順序が大幅に異なる言語間では、句の順序を調整しなければ適切な翻訳文とはならない。従来の句ベース統計翻訳は、句の順序調整を平坦な構造上で行っており、構文情報は使っていない。

一方、構文情報を利用し、木構造をマッピングすることにより翻訳を行う統計翻訳も提案されてきている (Graehl and Knight, 2004; Melamed, 2004)。しかしこれらは未だ理論にとどまっており、実際に翻訳したときの特徴は明らかになっていない。

本稿では、構文トランスファ方式統計翻訳を提案する。構文トランスファ方式は、古くから用いられている翻訳方式であり、構文構造が異なる言語間の翻訳に適していると考えられる。本稿で提案する方式は、階層的に語順を調整するだけでなく、句ベース統計翻訳で用いられているような長単位句を単位とすることができる特徴である。

2 構文トランスファ方式統計翻訳

2.1 概要

構文トランスファ方式機械翻訳は、入力文を構文解析し、構文木を出力の構文木にマッピングすることにより翻訳文を生成する。構文トランスファ方式による日英翻訳の例を図 1 に示す。

実際には、構文木全体を直接マッピングすることはできないため、通常、構文木のノード間でマッピングを行う。両言語間の構文ノードは一対一に対応しており構文ノードが翻訳単位となり、すべての部分木が

translation equivalence となる。これは、訓練時に対応づかない構文ノードを、あらかじめ削除しておくからである。

IBM model (Brown et al., 1993) 等の単語翻訳方式の統計翻訳では、translation equivalence を生成するため、fertility model, NULL model を用いて、単語の挿入・削除を行う。構文トランスファ方式では、単語の挿入・削除は、構文ノード間のマッピング時に自動的に行われる。たとえば、図 1 の例を生成する際には、(NP → バス) の変換時、自動的に「the」を挿入し、(NP → the bus) を生成する。また、語順変更は、変換時の子ノードの順序として表される。子の部分木は複数の単語から成り立っているため、構文トランスファ方式は、句を単位とした語順変更を階層的に行うことができる。

2.2 モデル

統計翻訳は、入力単語列 f が与えられたとき、確率を最大化する出力単語列 e を、すべての可能な組み合わせの中から探索することにより、翻訳を行う。

$$\begin{aligned}\hat{e} &= \operatorname{argmax}_e P(e|f) \\ &= \operatorname{argmax}_e P(e)P(f|e).\end{aligned}\quad (1)$$

$P(e)$ は言語モデル確率、 $P(f|e)$ は翻訳モデル確率と呼ばれる。構文トランスファ方式の統計翻訳は、翻訳モデル中に隠れ変数として原言語・目的言語の構文木 (それぞれ \mathcal{E} , \mathcal{F} と示し、単語列 e , f を生成する) を仮定し、木構造同士のマッピングを行うことにより、翻訳文を生成する。

$$P(f|e) = \sum_{\mathcal{E}, \mathcal{F}} P(f, \mathcal{F}, \mathcal{E}|e).\quad (2)$$

本稿では、単言語の確率的構文解析で用いられている内側確率を、2 言語に拡張した形でモデルを定義する。具体的には、翻訳モデルを目的言語 (入力) の木構造モデル、原言語 (出力) の木構造モデル、木構造マッピングモデルの 3 つのサブモデルに分割し、その総積で翻訳確率を表現する。

⁰ 本研究は独立行政法人 情報通信研究機構の研究委託「大規模コーパスベース音声対話翻訳技術の研究開発」により実施したものである。

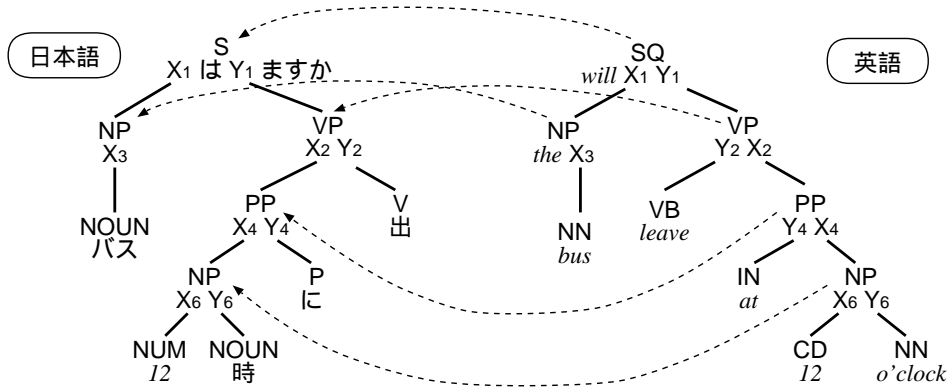


図 1: 構文トランスファによる翻訳過程の例 (日英翻訳)

目的言語木構造テーブル				木構造マッピングテーブル			原言語木構造テーブル			
名前	親ノード	確率	子ノード列	目的言語	確率	原言語	名前	親ノード	確率	子ノード列
θ_f^1	S	$2.8e^{-1}$	X_{NP} は Y_{VP} ます か	θ_f^1	$2.0e^{-1}$	θ_e^1	θ_e^1	SQ	$3.2e^{-1}$	will X_{NP} Y_{VP}
θ_f^2	VP	$3.8e^{-1}$	X_{PP} Y_V	θ_f^2	$1.2e^{-1}$	θ_e^2	θ_e^2	VP	$4.3e^{-2}$	Y_{VB} X_{NP}
θ_f^3	VP	$4.8e^{-2}$	X_{NP} に Y_V	θ_f^3	$4.3e^{-3}$	θ_e^3	θ_e^3	VP	$1.0e^{-2}$	Y_{VB} X_{PP}
θ_f^4	NP	$5.3e^{-1}$	X_{NOUN}	:			θ_e^4	NP	$1.2e^{-2}$	the X_{NN}

図 2: 構文トランスファ方式における翻訳モデルの例

たとえば、構文ノード N^i ($N_e^i \in \mathcal{E}$, $N_f^i \in \mathcal{F}$, $\text{yield}(\mathcal{E}^i) = e^i$, $\text{yield}(\mathcal{F}^i) = f^i$) が、 j 個の部分木 ($\mathcal{E}^{i+1} \dots \mathcal{E}^{i+j}$ および $\mathcal{F}^{i+1} \dots \mathcal{F}^{i+j}$) から成り立っているとすると、

$$\begin{aligned}
 P(f^i, \mathcal{F}^i, \mathcal{E}^i | e^i) &= P(\theta_e^i) P(\theta_f^i) P(\theta_f^i | \theta_e^i) \\
 &\cdot \prod_{k=1}^j P(f^{i+k}, \mathcal{F}^{i+k}, \mathcal{E}^{i+k} | e^{i+k}) \quad (3)
 \end{aligned}$$

で表す。ただし、 θ_e^i , θ_f^i は、それぞれ構文木 \mathcal{E}^i , \mathcal{F}^i を構成する文脈自由文法規則である。

3 訓練

3.1 句アライメント

本稿で用いる句アライメントは、Alignment Template (Och et al., 1999) とほぼ同様なアプローチをとる。ただし、異なる点は、Alignment Template が単語アライメントの連続性だけを基に句を抽出するのに対して、本稿では、構文木を制約として使い、原言語、目的言語ともに部分木として成り立ちうる句のみを対応づける点にある。

たとえば、単語アライメントと構文木が図 3 のとおり与えられた場合、単語アライメント (2)(3) に着目すると、これだけを含み、他の単語アライメント (i.e. (1)(4)(5)(6)) を含まない句としては、($NP \rightarrow 12$ o'clock), ($NP \rightarrow 12$ 時) があるので、これは句アライメントとして抽出される。しかし、単語アライメント (4)(5) に着目した場合、これだけを含み、他の単語アライメントを含まない部分木はないため、「leave at」と「に出」は句アライメント結果には含まれない。

このように抽出された句アライメントは、translation equivalence と見なされる。構文ラベルが付いているため、そのまま構文解析器に適用することが可能である。本稿では、これを長単位句と呼ぶ。

また、句アライメント結果は階層性を持っているため、直下のアライメントを非終端記号とすると、日本語と英語の文脈自由文法規則が得られる (図 2)。

3.2 パラメータ推定

本稿では、モデルの確率を、すべて相対頻度によって与える。たとえば、原言語の木構造モデルのパラメータは、以下の式で与える。

$$\begin{aligned}
 P(\theta_e^i) &= P(N_e^{i+1} \dots N_e^{i+j} | N_e^i) \\
 &= \frac{\text{count}(N_e^{i+1} \dots N_e^{i+j}, N_e^i)}{\text{count}(N_e^i)}, \quad (4)
 \end{aligned}$$

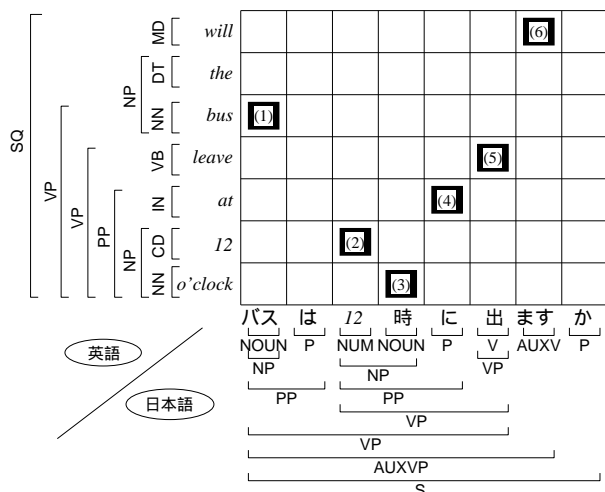


図 3: 句アライメント

ただし、 $count(N)$ は、構文ノード N がコーパス中出现する頻度で、 N_e^i は θ_e^i の親ノードのカテゴリ、 $N_e^{i+1} \dots N_e^{i+j}$ は、子ノードのカテゴリ列を表す。目的言語の木構造モデルのパラメータも同様に算出する。

木構造マッピングモデルのパラメータも同様に以下の式で与える。

$$P(\theta_f^i | \theta_e^i) = \frac{count(\theta_f^i, \theta_e^i)}{count(\theta_e^i)}. \quad (5)$$

4 デコード

構文トランスファ方式は、まず入力文の構文解析を行う。したがって、基本的には文脈自由文法を用いたパーザに、変換部、生成部を付加したものとして実現される。本稿では、ボトムアップチャートパーザを基にする。デコードの手順は以下のとおりである。

1. まず、目的言語の木構造モデルを用いて、入力文をボトムアップに構文解析する。
2. 入力の部分木ができた時点で、木構造マッピングモデルおよび原言語の木構造モデルを参照し、出力の部分木構造を作成する。
3. 出力の部分木を展開し、部分単語列を作成する。単語列の確率は、式 (3) に言語モデル確率を乗算したものによって与えられる。
4. 入力の部分木のカバー範囲が同じ、かつ入出力の部分木の構文ラベルが同じものについて、出力単語列リストをマージし、確率の上位 N 個のみを、その部分木の翻訳結果とする。
5. 以上ステップ 1~4 を、入力文全体が構文解析完了するまで繰り返す。

入力文の構文解析結果に曖昧性があるとき、または変換候補が複数存在するときは、個別に変換を行い、ステップ 4 で出力の単語列リストをマージする。長単位句の翻訳は、構文解析の曖昧性の一つとして扱われる。つまり、終端記号のみから成り立つ規則を各モデルに追加しておく、上記処理により、単語・句を区別せずに翻訳される。

5 実験

本稿では、日英翻訳を対象に評価を行う。

5.1 実験設定

コーパス: 本稿では対訳コーパスとして、評価型ワークショップ IWSLT (Akiba et al., 2004)¹ と同じコーパスを用いる。これは旅行会話に頻出する表現を集めた基本表現集である。訓練用として 20,000 対訳、テスト用として 500 文を使用した。

モデル: 翻訳モデルを構築するための単語アライメントとして、GIZA++ (Och and Ney, 2003) の IBM モデル 4 の Viterbi Alignment を使用した。また、英語は Charniak (2000) の解析器を用い、日本語は、内部で開発された句構造解析器を用いて訓練した。

言語モデルは単語 bigram および trigram を併用した。

評価方法: 自動評価法として、BLEU (Papineni et al., 2002), NIST (Dodington, 2002), mWER (multiple Word Error Rate, (Nießen et al., 2000)) を用いた。参照訳は 1 文あたり 16 である。また、主観評価として、A (完全訳), B (部分訳), C (理解可能訳), D (不可訳) の 4 段階評価を用いた。なお、mWER のみ、低いスコアが良質な翻訳を意味する。

5.2 実験結果

5.2.1 翻訳品質

まず、提案方法による翻訳品質を測定した。結果を表 1 に示す。構文情報、長単位句の効果を独立に測定するため、以下の 2 方式でも実験を行った。

- 翻訳モデルから、長単位句を削除した場合 (w/o phrases)。最も短い終端記号列から成り立つ規則だけでデコードした。
- 構文情報なしでデコードを行った場合。ここでは、USC ISI が開発した句ベースのビーム探索デコーダ、Pharaoh (Koehn et al., 2003)² を使用した。

¹<http://www.slt.atr.jp/IWSLT2004/>

²<http://www.isi.edu/licensed-sw/pharaoh/>

表 1: 翻訳品質 (Japanese-to-English)

システム	自動評価			主観評価		
	mWER	BLEU	NIST	A	A+B	A+B+C
提案方式 (w/ phrases)	0.462	0.422	8.41	43.6%	55.2%	64.6%
提案方式 (w/o phrases)	0.475	0.404	8.08	40.6%	54.2%	63.6%
Pharaoh (w/ phrases)	0.523	0.382	7.65	32.0%	45.8%	58.2%

なお、長単位句については、提案方式と同じものを使用した。

各方式の翻訳品質を比較すると、すべての評価基準に対し、提案手法（構文 + 長単位句）が最もよい性能を示した。長単位句と構文情報は併用することが可能であり、両者を併用することが最も性能を向上させることができる。

5.2.2 構文解析失敗

本提案方式は、構文解析を用いるため、下位の構造で構文解析に失敗した場合、それより上位の構造を作成することができない。つまり、獲得されたモデルが data sparceness 等により十分なカバレッジを持っていない場合、構文解析が失敗する。本実験では 110 文 (22%) が構文解析に失敗し、部分翻訳結果の組み合わせで翻訳された。

本方式は単言語の構文解析に比べ、(1) 原言語・目的言語の双方が構文解析できなければならない。(2) translation equivalence となる規則しか構文解析に利用できない。ため、どうしても失敗が多くなる。さらに翻訳品質を向上されるためには、構文解析失敗を抑えることが必要である。

6 まとめ

本稿では、構文トランスファ方式統計翻訳を提案した。本稿の提案方式は、従来の句ベース統計翻訳で用いられている長単位句と構文情報を組み合わせることができ、構文トランスファ単体、長単位句単体の翻訳に比べ、品質が向上することを示した。

今回、確率推定に相対頻度を用いたが、Inside-outside アルゴリズムも適用可能である。今後はこのような最適化の検討を行うことを考えている。

参考文献

Yasuhiro Akiba, Marcello Federico, Noriko Kando, Hiromi Nakaiwa, Michael Paul, and Jun'ichi Tsujii. 2004. Overview of the IWSLT04 evaluation campaign. In *IWSLT 2004 Proceedings*, pages 1–12.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of NAACL-2000*, pages 132–139.

George Doddington. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the HLT Conference*.

Jonathan Graehl and Kevin Knight. 2004. Training tree transducers. In *HLT-NAACL 2004: Main Proceedings*, pages 105–112.

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *HLT-NAACL 2003: Main Proceedings*, pages 127–133.

I. Dan Melamed. 2004. Statistical machine translation by parsing. In *Proceedings of ACL 2004, Main Volume*, pages 653–660.

Sonja Nießen, Franz Josef Och, Gregor Leusch, and Hermann Ney. 2000. An evaluation tool for machine translation: Fast evaluation for MT research. In *Proceedings of LREC 2000*, pages 39–46.

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

Franz Josef Och, Christoph Tillmann, and Hermann Ney. 1999. Improved alignment models for statistical machine translation. In *Proceedings of EMNLP 1999*, pages 20–28.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL 2002*, pages 311–318.

Stephan Vogel, Ying Zhang, Fei Huang, Alicia Tribble, Ashish Venugopal, Bing Zhao, and Alex Waibel. 2003. The CMU statistical machine translation system. In *Proceedings of MT Summit IX*, pages 402–409.

Richard Zens and Hermann Ney. 2004. Improvements in phrase-based statistical machine translation. In *HLT-NAACL 2004: Main Proceedings*, pages 257–264.