

文構造の利用による翻訳への効果の調査

坂田純 村上仁一 徳久雅人 村田真樹

鳥取大学大学院工学研究科情報エレクトロニクス専攻

{d112004,murakami,tokuhisa,murata}@eecs.tottori-u.ac.jp

1 はじめに

現存の機械翻訳方式のうち、統計翻訳 (SMT) は一般に、解析で得られた文構造 (句構造解析結果等) を利用しない方法であった。しかし近年では構文解析結果を利用する「統語情報に基づく統計翻訳 (syntax-based SMT)」も頻繁に用いられている。特に日英などの性質の大きく異なる言語間では、その翻訳精度の高さが注目されている。ただし、利用する文構造の翻訳精度の向上への寄与の仕方等は、まだ明らかではない。

通常、機械翻訳では文構造で句構造や係り受け構造のことを指すが、非階層的に一文全体をカバーする文パターン対も、文構造の一種とみなすことができる。文パターン対を使用する手法 (文パターン翻訳) では、文パターンが適合した場合においては、翻訳精度が高いと報告されている。以上より syntax-based SMT と文パターン翻訳共に、一文全体の文構造を利用することで高い翻訳精度を得ているとみなせる。

そこで本論文では、両機械翻訳方式の使用文構造の違いに着目し、どのようにして文構造等の情報が翻訳精度の向上へ寄与するかを調査する。なお文パターン翻訳方式では重文複文型パターン辞書 [1] を使用し、syntax-based SMT には tree-to-string の SMT [2] を使用する。両翻訳システム間の大きな違いは、翻訳ルールの作成が手動か自動であるか、文構造が非階層的であるか多階層的であるかの二点である。

2 文パターン翻訳

2.1 文パターン対

本論文の文パターン翻訳では、重文複文型パターン辞書 [1] の文パターン対を使用する。この文パターン対は、日本語重文複文とその英語対訳文の約 12 万文対を対象に、それぞれの対訳文対ごとに人手により作成されている。文パターン対は、変数、関数、字面、記号の 4 種類の記述形式を用いて記述されている。変数は別の要素に置き換え可能な要素であり、各日英対訳文対の間で、明確な対応を取ることのできる自立語が変数に置き換えられている。関数は、語形の指定や、述部語尾表現の表記の揺れを吸収するための汎化等の役割を持つ。変数化または関数化されない文要素は、字面のまま文パターンの中に記述される。また、記号は適合範囲の拡大等の役割を持つ。表 1 に文パターン対の具体例を示す。

2.2 文パターン翻訳の手順

本論文の文パターン翻訳方式は、重文複文型パターン辞書を用いる [3] のシステムを使用する。文パターン

翻訳は、一文全体の文構造の抽出と変換を、文パターン対を用いて一括して行う方法だとみなせる。システム全体のモデル図を図 1 に、手順とその概要を以下に示す。

1. 日本語パターン照合システムにより、入力文に対しパターン照合を行い、適合した日本語文パターン、日本語文パターンに対応する英語文パターン、変数情報を得る。
2. 意味属性コードを用いて、日本語文パターンの絞込みを行う。
3. 英文生成システムにより、日英文パターン対、単語辞書を用いて英語文 (翻訳候補文) を生成する。訳語選択では、単語連鎖確率 (単語 tri-gram) と単語翻訳確率を用いて、最尤となる訳語の組み合わせを選択する。
4. 複数の翻訳候補文が得られた場合は、確率値が最大の候補文の一つ、出力文として選択する。

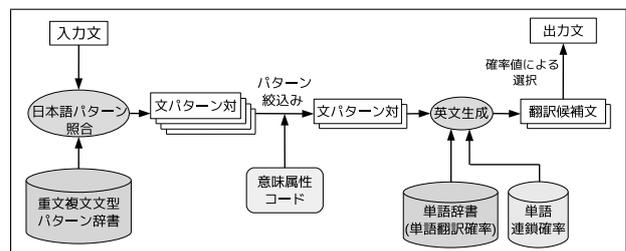


図 1 重文複文型パターン辞書を用いた文パターン翻訳方式

3 tree-to-string の SMT

本論文では、統語情報に基づく統計翻訳のうち、tree-to-string の SMT を使用する [4, 5]。この方式は事前に目的言語側で構文解析を行った後に、翻訳ルールを学習する。そのため作成された翻訳ルールは、目的言語側で品詞情報を持つ木構造として作成される。そして SCFG を用いた手法 (統語情報に基づく統計翻訳の 1 手法) より、各翻訳ルールが高い表現能力を持つとされている。ただし翻訳精度は、構文解析器の解析能に強く依存している。また単語アライメント情報を元に、原言語側と目的言語側の対応をとるため、単語アライメントの精度にも依存している。

4 翻訳実験

まず文パターン翻訳と tree-to-string の SMT の翻訳精度を調査する。

表1 文パターン対の記述例

日本語文パターン原文	彼のお母さんがああ若いとは思わなかった。
英語文パターン原文	I never expected his mother to be so young.
日本語文パターン	<N1 は> N2(彼;N123:他称 (単数/男),N124:男) の N3(お母さん;N180:母,N149:女) が ああ AJ4(若い;NY5:属性) とは V5(思わ;NY32:思考動作,NY31:感情動作).hitei.kako.
英語文パターン	<I N1> never V5^past N2^poss N3 to be so AJ4.

4.1 実験条件

入力文に重文複文型パターン辞書の日本語文パターン原文 (表1 参照) を用いて, 日英機械翻訳の実験を行う。

4.1.1 文パターン翻訳

本実験では 10 万文パターン対を使用する。この 10 万文パターン対の日英原文対をもとに, 単語辞書と単語翻訳確率の学習を行う。単語連鎖確率 (単語 tri-gram 値) の学習は, この 10 万文パターン対の英語文パターン原文 10 万文を用いて行う。学習に使用しない日本語文パターン原文から, 5,000 文を抽出しテスト文に用いる。

4.1.2 tree-to-string の SMT

翻訳システムには, Travatar を用いる。日本語句構造解析器には Ckylark[6] を用い, 単語アライナーには GIZA++[7] を用いる。翻訳モデルの学習は, 文パターン対の学習と同じ日英原文 10 万文対を用いる。言語モデルの学習は, この 10 万文パターン対の英語文パターン原文 10 万文を用いて行う。パラメータはデフォルト値を使用する。また, 重文複文型パターン辞書の対訳文 1,000 文対を使用してパラメータチューニングを行う。

4.2 実験結果

文パターン翻訳では, 一文全体に適合する文パターンが得られたときのみ翻訳を行う。入力 5,000 文のうち, パターン照合と意味属性コードを用いたパターン絞込みに成功した文は, 572 文であった。重文複文型パターン辞書において, 日本語原文が同じでも対訳英文が異なる場合は, 異なる対訳文対とみなしている。したがって学習文が, 入力文と同じ日本語原文を含んでいる場合がある。文パターンが適合した 572 文中, 学習文と同じ日本語文を持つのは 193 文であり, 持たない文は 379 文であった。本実験では, これら 379 文のみを対象として評価を行う。

4.2.1 評価結果

上記 379 文の翻訳結果に対し, BLEU[8], TER[9], METEOR[10], RIBES[11] の 4 つの評価法を用いて自動評価を行った。自動評価の結果を 2 に示す。

次に上記 379 文からランダムに 100 文抽出し人手評価を行った。人手評価は adequacy による 5 段階評価を用いた。人手評価の基準を表 3 に, 人手評価結果を表 4 に示す。表 4 において, 各評価値における数値は, その評価値の文数を示している。表 2 と表 4 より, 自動評価と人手評価共に, 文パターン翻訳が Travatar よりも翻訳精度が高いことがわかる。

表2 自動評価結果

	BLEU	TER	METEOR	RIBES
文パターン翻訳	0.290	0.574	0.568	0.771
Travatar	0.217	0.641	0.489	0.756

表3 人手評価の基準

評価 5	入力文の意味を正しく理解できる
評価 4	一部不適切な部分があるが, 概ね理解できる
評価 3	入力文の意味が何となく読み取れる
評価 2	部分的な理解にとどまる, または意味が入力文と大きく異なる
評価 1	ほとんど意味を理解できない

表4 人手評価結果

	評価 1	評価 2	評価 3	評価 4	評価 5	平均値
文パターン翻訳	10	24	18	7	41	3.45
Travatar	29	36	14	3	18	2.45

5 文構造の利用による翻訳への効果

本論文で使用する両翻訳システムでは, 翻訳に使用する文構造に違いがある。本節ではそれぞれの文構造がどのように翻訳文に影響を与えているか調査する。

5.1 日英文パターン対の翻訳に対する効果

文パターン翻訳において, 出力文に使用された文パターン対がその入力文の翻訳に対し適切であったか調査した。その際, 英語文パターンを用いて, 適切な翻訳文を出力可能だと判断できる場合を“適切”, そうでないときは“不適切”とした。人手評価を行った 100 文のうち, 未知語を含まない 87 文における, 英語文パターンの適切さと人手評価値の関係を表 5 に示す。

表 5 において, 英語文パターンが適切な文の割合は約 80%(70/87 文) であり, そして適切な場合は約 69%(48/70 文) の文で人手評価が高い (評価 4 または 5)。したがって多くの入力文において, 文パターン対により, 入力文の構造を正しく英語文の構造に変換でき, そして高い翻訳精度を得ているとみなせる。

表5 英語文パターンの適切さと人手評価値の関係

	評価 1	評価 2	評価 3	評価 4	評価 5	合計	平均
適切	1	10	11	7	41	70	4.1
不適切	4	7	6	0	0	17	2.1

5.2 tree-to-string の SMT において作成される文構造

tree-to-string の SMT である Travatar においても, 入力文に適合する句構造と, それに対応する英語文構造が

作成される。また文パターン対における変数は翻訳ルールにおける非終端記号に相当し、各翻訳ルールは大別して次の3つに分類できる。

- ルール 1 文全体の翻訳ルール
- ルール 2 非終端記号を含む句翻訳ルール
- ルール 3 非終端記号を含まない句翻訳ルール

図 2 に Travatar において作成される日英の文構造の例を示す。日本語の文構造は品詞情報を持つ句構造であるが、それらは記述の簡略化のため省略した。比較のため同入力文で使用された文パターン対を図 2 の下に示している。

この例は root と終端記号を含め、5 階層の木構造となっている。ルール 1 とルール 2 による階層構造が文パターン対と対比可能であり、文パターン対はルール 2 を持たない 3 層限定 (root と終端記号を含む) の木構造とみなすことができる。ルール 3 は、本翻訳方式における形態素の翻訳およびその訳語選択と、ほぼ同等とみなせる。

5.2.1 比較システムにおける文構造の階層数

人手評価を行った 100 文のうち、未知語を含まない文は 68 文であった。この 68 文中、句構造解析に失敗していた文は 12 文であり、この 12 文を除いた 56 文に対し文構造の階層数を調査した。

56 文の階層数による分類を表 6 に示す^{*1}。最小で 3 階層、最大で 7 階層であり、平均 4.6 階層であった。3 階層で構成された文は 5 つだけであったため、ほとんどの文構造が、ルール 1 の適用後にルール 2 を適用して多階層で構成されていることを示している。また階層数による人手評価値の違いは、ほとんどみられなかった。

表 6 文構造における階層数とその内訳

	階層数				
	3	4	5	6	7
文数	5	20	27	3	1
人手評価値平均	3.2	2.6	3.0	3.0	2.0

5.2.2 使用された翻訳ルールの適切さ

次に使用された翻訳ルールが適切であったか調査した。上記 56 文に対する、使用された不適切な翻訳ルールの種類 (ルール 1 から 3) と人手評価値との関係を表 7 に示す。各値は文数を示しており、“なし” は不適切なルールの使用がない文数を示している。翻訳ルールの探索は root から下層に向かって行われるため [4, 5]、上層で不適切な翻訳ルールが適合した場合は、その下層においてどの翻訳ルールを用いても適切に翻訳できない可能性がある。そのため複数の不適切なルールの使用がある場合は、上層から下層のルールの順に、つまりルール 1 から 3 の順に振り分ける。なお、不適切な翻訳ルールに

は、日英で対応の取れていない誤ったルール^{*2}と、ルールそのものとしては良いがその文における使用では不適切なルールの二通りがある。場合によっては、誤った翻訳ルールが組み合わさって適切な翻訳文となることもある。

表 7 において、不適切なルールの使用により不適切な文構造が作成されたのは 36 文 (ルール 1 の 15 文とルール 2 の 21 文) であり、その大半の 33 文で翻訳精度が低い。そして残り 20 文 (ルール 3 の 5 文となしの 15 文) のうち 15 文で高い翻訳精度が得られていることから、文構造 (句構造) の利用による高い翻訳精度を得る効果がみられる。ただしあまりにも適切な文構造を得る確率が低いいため、現時点では文構造の利用が翻訳精度の向上にさほど寄与していない。

表 7 使用された不適切な翻訳ルールの種類による分類

評価値	ルール 1	ルール 2	ルール 3	なし	合計
5	2	1	0	13	16
4	0	0	0	2	2
3	3	6	1	0	10
2	4	9	4	0	17
1	6	5	0	0	11
合計	15	21	5	15	56

5.2.3 多階層の文構造を使用する場合の問題点

表 7 で不適切な翻訳ルールが使用された 41 文中において、45 個の不適切な翻訳ルールの使用が生じていた。この 45 個の内訳は、対応のとれていない誤ったルールの使用が 33 個で、残り 12 個がその文における使用では不適切な場合であった。句構造解析は 68 文中 56 文で成功しているため (5.2.1 節)、句構造解析失敗が誤った翻訳ルール作成の主要因とは考え難い。このため学習における単語アライメントの精度の低さが、不適切な翻訳ルールおよび不適切な文構造の作成される最大の原因と考えられる。したがって単語アライメントの精度の向上により、翻訳精度が向上する可能性が高い。ただし、対応のとれた翻訳ルールが多数作成されたとしても、その入力文においては不適切な翻訳ルールが選択される可能性が残る。

また上記 12 個の内訳を調べたところ、ルール 2 が 8 個であった。図 2 における、「X0 に従った → followed X0」の翻訳ルールがこれにあたる。入力文は提案の文であり、動詞 “follow” が現在分詞か、“had better” に続くときは原型であるべきだが、翻訳確率により過去形の “followed” が選択されている。これは階層を下に降るときに、文法的な制約が伝達されず消失してしまうことによると考えられる。この例は、多階層の文構造を用いて翻訳を行う場合に、階層間において文法的制約を伝達する必要があることを示唆している。

^{*1} root 直下の「日:X0. → 英:X0.」の翻訳ルールは階層数のカウントから除いた。

^{*2} 日英翻訳では、日本語で省略された要素を、何かしらの翻訳結果として英語文に挿入する必要が多々見られる。そのため厳密には日英で対応がとれていない翻訳ルールでも、例えば root 直下の「日:X0 → 英:he X0」のような一部のルールは、対応の取れた正しいルールとみなした。

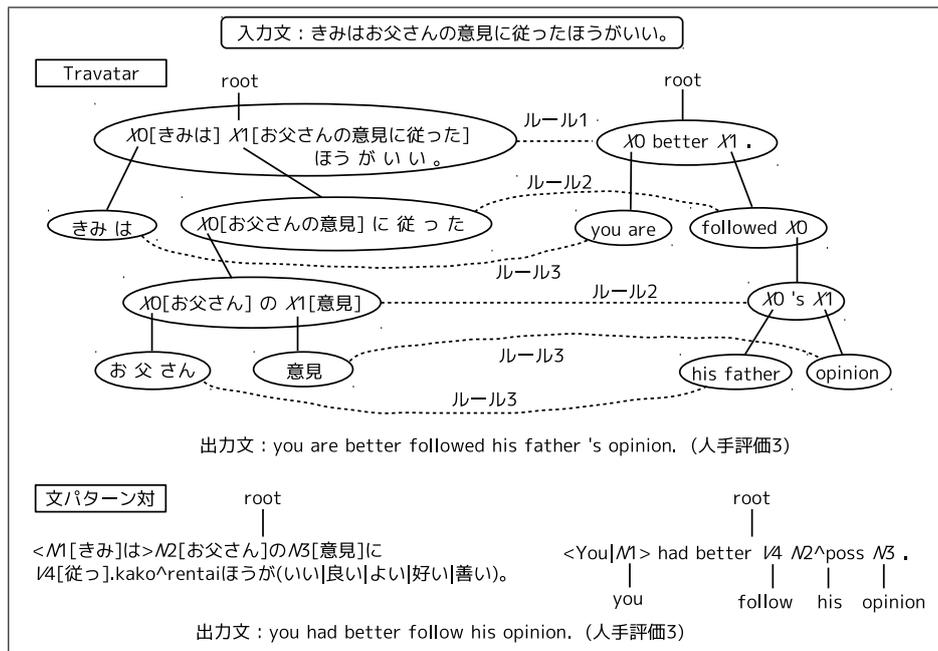


図2 Travatar で作成される日英の文構造

5.3 文構造の翻訳への効果

表5と表7の結果より、両翻訳システム共に文構造の使用により高い翻訳精度を得ていることが確認された。ただし Travatar では、日英で対応のとれていない翻訳ルールが多数作成されているため、全体の翻訳精度の向上にはあまり寄与していない。このため多階層的に木構造を作成することによる、翻訳精度への具体的な効果は、明らかにはならなかった。しかし分析結果より、多階層的な木構造を利用する場合は、階層間における文法的制約の伝達の必要性が示唆されている。多階層的に木構造を作成する効果は、翻訳精度の向上よりも翻訳可能な文の増加に貢献していると考えられる。

一方、文パターン対は少数階層(3階層)限定の木構造とみなせる。表5の結果は、3階層限定の文パターン対でも、高い翻訳精度を得ることが十分可能であることを示している。一文全体の文構造の変換を、一つの変換ルール(文パターン対)で行うため、階層間の文法的制約の伝達等を考慮しなくてよいことが、翻訳精度の高い要件の一つと考えられる。

6 おわりに

本論文では、文パターン翻訳と tree-to-string の SMT が使用する文構造の違いに着目し、どのようにして文構造等の情報が翻訳精度の向上へ寄与するかを調査した。分析の結果、両翻訳システム共に、文構造の使用により高い翻訳精度を得ていることが明らかになった。しかし、tree-to-string の SMT では自動作成した翻訳ルールの精度が低く、文構造の使用が全体の翻訳精度の向上にはあまり寄与していない。tree-to-string の SMT における多階層的に木構造を作成する効果は、翻訳精度の向上よりも翻訳可能な文の増加に貢献していると考えられる。一方、文パターン対は3階層限定の木構造とみなせ、少数

階層限定の文パターン対でも、高い翻訳精度を得ることが十分可能であることが明らかになった。

参考文献

- [1] 池原悟, 阿部さつき, 徳久雅人, 村上仁一. 非線形な表現構造に着目した重文と復文の日英文型パターン化. 自然言語処理, Vol.11, No.3, pp.69-95, 2004.
- [2] Graham Neubig, Travatar: A Forest-to-String Machine Translation Engine based on Tree Transducers. Proc. of the ACL Demonstration Track, pp.91-96, 2013.
- [3] 坂田純, 徳久雅人, 村上仁一, 村田真樹. 意味的等価変換方式に基づく単語レベルパターン翻訳方式の評価. 言語処理学会第20回年次大会発表論文集, pp.298-301, 2014.
- [4] Yang Liu, Qun Liu, and Shouxun Lin. Tree-to-String Alignment Templates for Statistical Machine Translation. Proc. of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL, pp.609-616, 2006.
- [5] Haitao Mi, Liang Huang, and Qun Liu. Forest-Based Translation. Proc. of the ACL-08:HLT, pp.192-199, 2008.
- [6] Yusuke Oda, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. Ckylark: A More Robust PCFG-LA Parser. Proc. of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT) (Software Demonstrations), pp.41-45, 2015.
- [7] Franz Josef Och, and Hermann Ney. Giza++: Training of statistical translation models. 2000. (<http://code.google.com/p/giza-pp/>)
- [8] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. 40th Annual meeting of the Association for Computational Linguistics, pp.311-318, 2002.
- [9] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A Study of Translation Edit Rate with Targeted Human Annotation. Proc. of the 7th Conference of the Association for Machine Translation in the Americas, pp.223-231, 2006.
- [10] Sanjeev Banerjee and Alon Lavie. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. Proc. of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, pp.65-72, 2005.
- [11] 平尾努, 磯崎秀樹, Kevin Duh, 須藤克仁, 塚田元, 永田昌明. RIBES: 順位相関に基づく翻訳の自動評価法. 言語処理学会第17回年次大会発表論文集, pp.1115-1118, 2011.