

# 類推用例翻訳における文長ごとの翻訳品質の評価

木村 竜矢      松岡 仁      西川 裕介   ルパージュ・イヴ

早稲田大学大学院 情報生産システム研究科

{tatsuya-kimura@ruri.,jinmatsuoka@akane.,y\_nishikawa@asagi.,yves.lepage@}waseda.jp

## 1 はじめに

現在の機械翻訳の主流はコーパスベースの機械翻訳である。コーパスベースの機械翻訳は大きく2つに分類できる。1つめが統計的機械翻訳、2つめが用例翻訳である。近年の統計的機械翻訳の研究の主流は、句に基づく翻訳モデルである。句に基づく翻訳モデルにおける句の並べ替えを表現する多くの歪みモデルが提案されている。用例翻訳では、構造的言語処理を指向する用例ベース機械翻訳システム [1] や、類推関係に基づいた用例翻訳 [2] などが提案されている。統計翻訳では、並べ替え (リオーダーリング) の問題を解決する為に複雑な歪みモデルを用いているが、それに比べ用例翻訳ではよりシンプルな翻訳手法で精度の高い翻訳が可能である事が示されている [2]。そこで、本論文では統計翻訳のような複雑な歪みモデルを用いなくとも、用例翻訳がどの程度、統計翻訳に近い振る舞いをおこなう事ができるのかを検証する。

## 2 類推関係に基づいた用例翻訳

本論文で使用する用例翻訳システムは、類推関係に基づいた用例翻訳システム [2] である。類推とは式 (1) で示すように、A と B の関係は C と D の関係に等しい事を表す。

$$A : B :: C : D \Rightarrow \begin{cases} d(B, D) = d(A, C) \\ d(C, D) = d(A, B) \\ |A|_a + |D|_a = |B|_a + |C|_a, \forall a \end{cases} \quad (1)$$

ここで、 $d(A, B)$  は A と B の編集距離を表し、 $|A|_a$  は A における文字 a の個数を表している。この関係を翻訳に応用すると式 (2) のように表現できる [2]。

$$\left. \begin{array}{l} \hat{A} \text{ translates } A, \\ \hat{B} \text{ translates } B, \\ \hat{C} \text{ translates } C, \\ A : B :: C : D \end{array} \right\} \Rightarrow \begin{array}{l} \text{Any } \hat{D} \text{ such that} \\ \hat{A} : \hat{B} :: \hat{C} : \hat{D} \\ \text{translates } D \end{array} \quad (2)$$

本論文で用いる類推関係に基づいた用例翻訳システムでは、可切性を用いて得られた翻訳テーブルを使用する [3]。以下に例として、"she eats the hamburger" を英語から日本語に翻訳する過程について説明する。まず、図 1 のように可切性を計算し、文章を木構造化する [4]。次に、その木構造に基づいて文章を以下の 1 から 3 のように分割する。その後、式 (2) を用いて小

1. "the hamburger"
2. "eats the hamburger"
3. "she eats the hamburger"

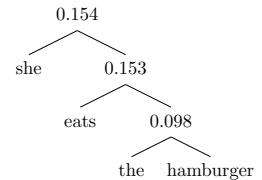


図 1: Binary tree

さな部分木から翻訳を行う。ここでは、まず始めに "the hamburger" から翻訳を始める。もし、翻訳したいものが翻訳テーブルに存在すればそのまま結果を出力する。"the hamburger" は表の翻訳テーブルに存在するので "ハンバーガーを" と翻訳することができる。

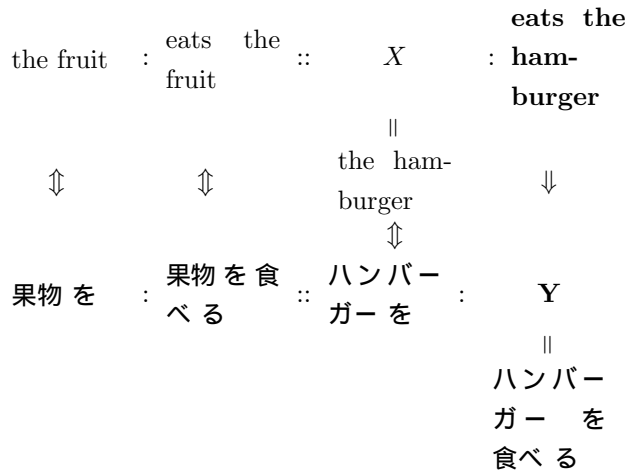
表 1: 可切性を用いた翻訳テーブル

元言語 (英語)	目的言語 (日本語)
the fruit	果物を
eats the fruit	果物を食べる
the hamburger	ハンバーガーを
she eats the fruit	彼女が果物を食べる
⋮	⋮

次に、"eats the hamburger" を翻訳する。もし、翻訳したいものが翻訳テーブルに存在しなければ類推関係に基づいて翻訳を行う。ここで、"eats the hamburger" は表 1 の翻訳テーブルに存在しないので、類推関係に基づいて翻訳する。

表 2: 訓練データとテストデータ

		フランス語	英語	フィンランド語	フランス語	ポルトガル語	スペイン語
訓練データ	文の数	347,614		347,614		347,614	
	単語の数	10,959,243	9,945,400	7,180,028	10,959,243	10,302,370	10,472,185
	文の長さ: 加算平均 ± 標準偏差	31.56±17.16	28.65±15.49	20.70±11.25	31.56±17.16	29.63±16.30	30.17±16.73
テストデータ	文の数	1,000		1,000		1,000	
	単語の数	28,812	27,182	20,665	29,238	27,541	27,590
	文の長さ: 加算平均 ± 標準偏差	28.85±12.81	27.21±12.49	20.71±10.69	29.28±13.07	27.53±13.30	27.62±12.74



ここで、類推関係により得ることができた言語資源は表 3 のように新たに翻訳テーブルに追加する。最後に "she eats the hamburger" を類推関係に基づいて翻訳する。ここでは、表 3 の新たに言語資源を追加した翻訳テーブルを使用する。以上のように、小さい部分木から翻訳を行う事で、翻訳を行うにつれて翻訳テーブルの質が高くなり、翻訳出来る可能性を向上させる事が出来る。

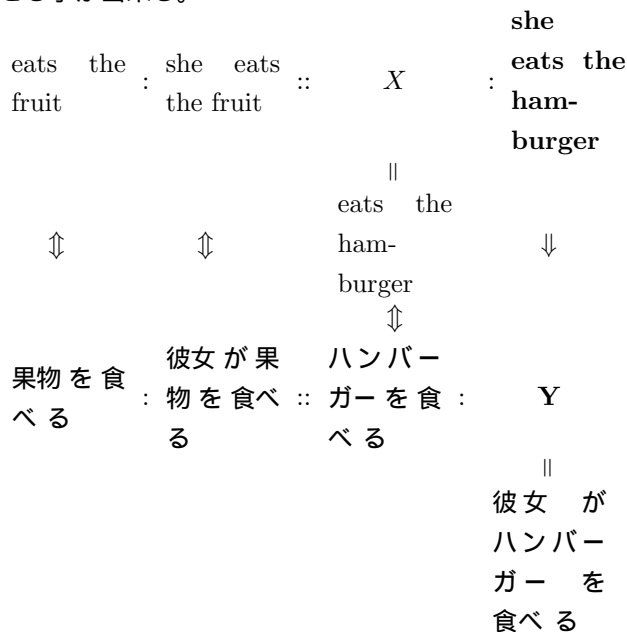


表 3: 翻訳結果を追加した翻訳テーブル

元言語 (英語)	目的言語 (日本語)
the fruit	果物を
eats the fruit	果物を食べる
the hamburger	ハンバーガーを
she eats the fruit	彼女が果物を食べる
⋮	⋮
eats the hamburger	ハンバーガーを食べる

### 3 実験

本論文の実験では、訓練データとして Europarl コーパスを使用した [5]。使用した訓練データは 347,614 文で、テストデータは文の長さを  $1 \leq 60$  単語の間でサンプリングした 1,000 文である (表 2)。言語ペアとしては、ヨーロッパの主要な 11 言語間での全ての組み合わせにおいて、通常最も BLEU スコアが低いフランス語とフィンランド語、通常最も BLEU スコアが高いスペイン語とポルトガル語、それに加え、英語とフランス語に関しての実験を行った。現在の統計翻訳の主流である GIZA++/Moses[6] の翻訳テーブルを用いて Moses デコーダー [7] で翻訳をおこなった結果、可切性により得られた翻訳テーブル [3] を用いて類推関係に基づいた用例翻訳システム [2] で翻訳をおこなった結果と GIZA++/Moses[6] の翻訳テーブルを用いて類推関係に基づいた用例翻訳システム [2] で翻訳をおこなった結果の 3 つを比較した。出力された結果をそれぞれ 1 文ずつ BLEU スコア [8] を計算し、グラフ化したものが図 2 から図 7 である。ここでは、複数出現する同じ長さの文の BLEU スコアについては加算平均で表している。表 5 は 1,000 文全体の BLEU スコアを示す。

表 4: 翻訳テーブルのエントリー

	仏-英 (fr-en)	仏-芬 (fr-fi)	葡-希 (pt-es)
可切性のエントリー数	1,255,774	1,045,670	1,337,194
長さ: 加算平均 ± 標準偏差	10.04±14.96	11.40±17.40	10.47±15.56
GIZA++のエントリー数	1,329,320	597,235	1,681,573
長さ: 加算平均 ± 標準偏差	3.03±1.27	3.03±1.47	3.23±1.37



図 2: フランス語から英語の翻訳結果

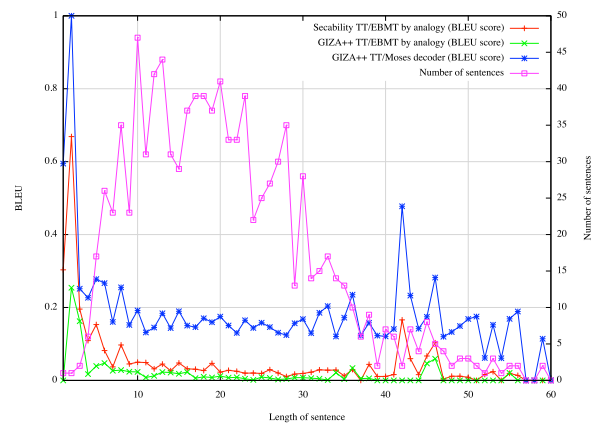


図 5: フィンランド語からフランス語の翻訳結果

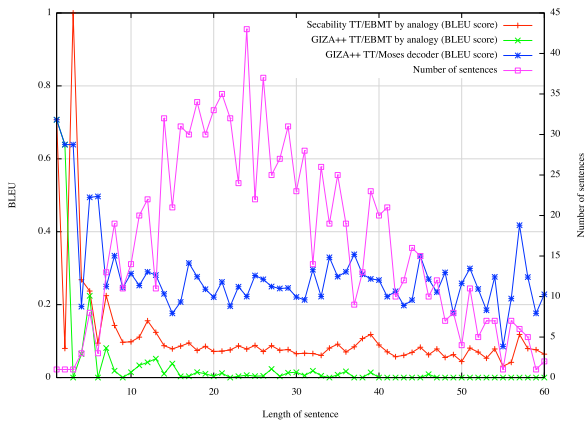


図 3: 英語からフランス語の翻訳結果

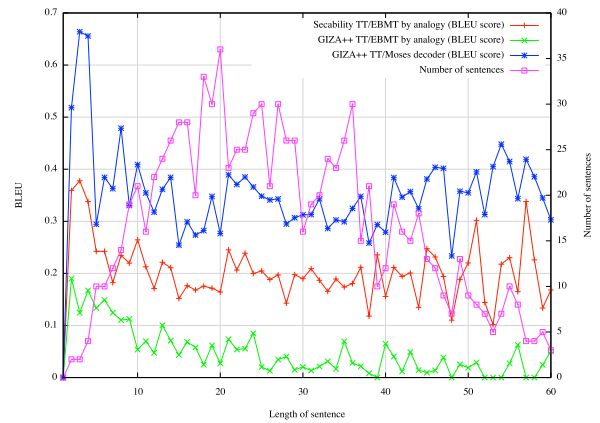


図 6: ポルトガル語からスペイン語の翻訳結果

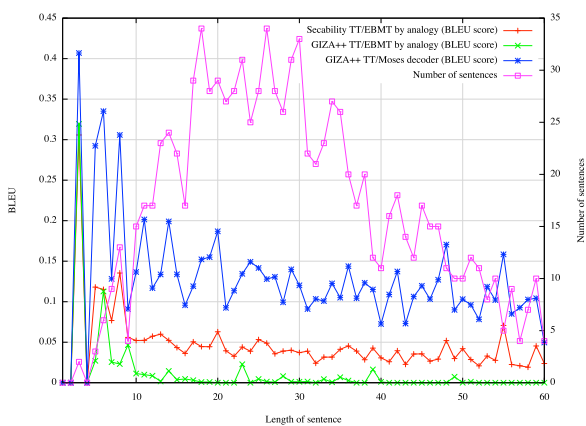


図 4: フランス語からフィンランド語の翻訳結果

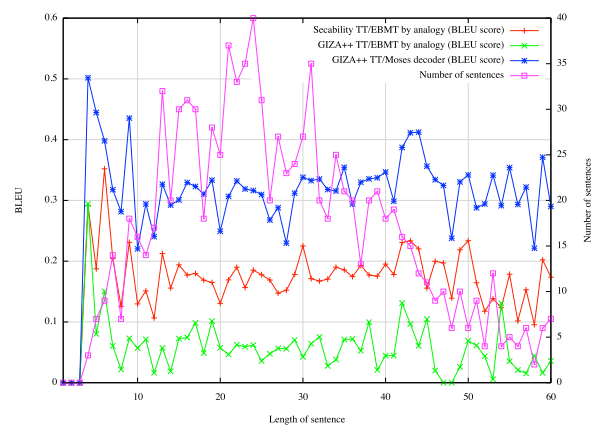


図 7: スペイン語からポルトガル語の翻訳結果

表 5: BLEU 自動評価結果

	fr-en	en-fr	fr-fi	fi-fr	pt-es	es-pt
可切性の翻訳テーブルを用いた類推用例翻訳	11.94	7.78	2.95	1.88	20.96	19.81
GIZA++の翻訳テーブルを用いた類推用例翻訳	0.13	0.06	0.09	0.14	1.38	4.16
GIZA++の翻訳テーブルを用いた統計翻訳	31.13	34.10	12.02	16.61	36.13	34.63

## 4 考察と結論

可切性の翻訳テーブルを用いた類推用例翻訳では、文の長さが 10 までは比較的高い BLEU スコアをだしているが、その後減少し、ある一定値を行き来していることがわかる。文の長さが 10 から 60 の長さにおいては BLEU スコアは変動が少なく、用例翻訳システムでも長い文において翻訳結果の品質にあまり影響がないことが読み取れる。しかし、従来手法である GIZA++ の翻訳テーブルを用いた類推用例翻訳では、文が長くなるほど翻訳品質が低い、もしくは翻訳が出来ていない。これは、表 4 からわかるように翻訳テーブルのエントリーの長さが短い事が原因だと考えられる。また、フランス語とフィンランド語の BLEU スコアが総体的に低い原因としては、表 2 からわかるように他の言語（特にフランス語）に比べて単語の数が著しく少ないことが読み取れる。これは、他の言語では 2 語や 3 語で表現する単語をフィンランド語では 1 語など短い語数で表現しているため、翻訳の際に連語が上手く翻訳出来ていない事が影響していると考えられる。総体の結論としては、可切性の翻訳テーブルを用いた類推用例翻訳は文の長さによる BLEU スコアの影響は無く、長い文においても安定した BLEU スコアを出す事が出来る事が証明出来た。

## 5 おわりに

本論文では、類推関係に基づいた用例翻訳における文長ごとの翻訳品質を評価した。その結果、可切性の翻訳テーブルを用いた類推用例翻訳では長い文においても短い文とさほど変わらない BLEU スコアを出せることがわかった。

今後の課題としては MultiWord Expression (MWE) などで単語のまとめあげをおこない、翻訳テーブルの質を向上させることで、より精度の高い翻訳を用例翻訳でもおこなえる事ができると考えられる。

## 謝辞

本稿は早稲田大学特定課題研究助成費（課題番号 2013A-6336）による研究成果の一部である。

## 参考文献

- [1] Nakazawa, Toshiaki and Yu, Kun and Kawahara, Daisuke and Kurohashi, Sadao, Example-based machine translation based on deeper NLP, *Proceedings of IWSLT*, pages 64-70, 2006
- [2] Y. Lepage and E. Denoual. Purest ever example-based machine translation: Detailed presentation and assessment. *Machine Translation*, 19(3): pages 251-282, 2005b.
- [3] T. Kimura, J. Matsuoka, Y. Nishikawa and Y. Lepage, Generation of Translation Tables Adequate For Example-Based Machine Translation by Analogy, In *Proceedings of the International Conference on Artificial Intelligence and Software Engineering (AISE2014)*, Phuket, Thailand, 2014.
- [4] C. Chenon. Vers une meilleure utilisabilité des mémoires de traduction, fondée sur un alignement sous-phrastique *PhD thesis*, Université Joseph Fourier-Grenoble 1, 2006.
- [5] P. Koehn. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the Tenth Machine Translation Summit (MT Summit X)*, pages 79-86, Phuket, Thailand, 2005.
- [6] F.J. Och and H. Ney. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19-51, 2003.
- [7] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, Moses: Open source toolkit for statistical machine translation, *Proceedings of 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, Prague, Czech Republic, pages 177-180, 2007.
- [8] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL 2002)*, pages 311-318, 2002.