

# 階層型統計翻訳の調査

久保田 裕介 村上仁一 徳久雅人 村田真樹  
鳥取大学 工学部 知能情報工学科  
{s062029, murakami, tokuhisa murata}@ike.tottori-u.ac.jp

## 1 はじめに

渡辺ら [18] は日本語, 英語間における特許情報の翻訳を行っている. その研究の中で階層型統計翻訳と句に基づく統計機械翻訳の比較を行っている. しかし特許情報は, 文が長いと人手評価が困難である. そこで, 本研究では短い単文, 重文複文の日英翻訳において階層型統計翻訳と句に基づく統計機械翻訳の比較調査を行う. なお, 本論文では句に基づく統計機械翻訳は PSMT, 階層型統計翻訳は HSMT と表記する.

## 2 句に基づく統計翻訳 (PSMT)

日英統計翻訳は, 日本語文  $J$  が与えられたとき, 全ての組み合わせから確率が最大となる英語文  $E$  を探索し翻訳を行う. 翻訳モデルは  $P(j|e)$ , 言語モデルは  $P(e)$  である. 図 1 に日英統計翻訳の手順を図 1 に示す.

$$E = \arg \max_e P(e|j) \approx \arg \max_e P(j|e)P(e)$$

- 手順 1: 日英対訳学習文より翻訳モデルを学習
- 手順 2: 日英対訳学習文の英語文より言語モデルを学習
- 手順 3: デコーダで手順 1, 2 で学習した翻訳モデルと言語モデルを用いて翻訳を行う.

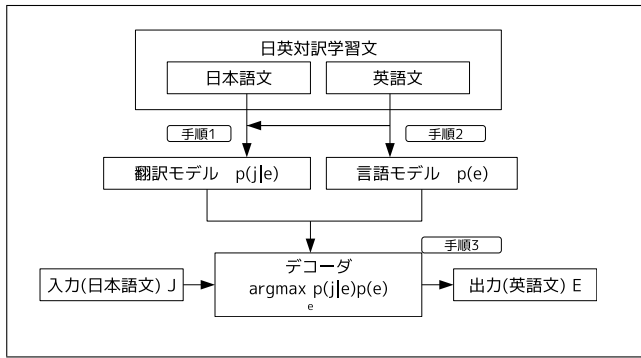


図 1 日英統計翻訳の手順

### 2.1 翻訳モデル

翻訳モデルは, 英語の単語の列から単語の列へ確率的に翻訳を行うためのモデルである. 翻訳モデルはフレーズテーブルで管理されている. 表 1 に例を示す. 左から, 日本語フレーズ, 英語フレーズ, フレーズの日英翻訳確率, 単語の日英翻訳確率の積, フレーズの英日翻訳確率, 単語の英日翻訳確率の積である.

表 1 フレーズテーブルの例

彼は		he is		0.23	0.05	0.04	0.01
10 パーセント		a ten parsent		1	0.02	0.25	0.005

### 2.2 言語モデル

言語モデルには,  $N$ -gram を用いる.

### 2.3 デコーダ

デコーダは, 与えられた文から言語モデルと翻訳モデルの確率が最大となる文を探索し翻訳を行う.

### 2.4 パラメータチューニング

パラメータチューニングには, MERT[8] (Minimum Error Rate Training) が用いられる. MERT は, 目的の自動評価 (一般的に BLEU) を最大にするような翻訳結果を出力するようにパラメータを調整する.

## 3 階層型統計翻訳 (HSMT)

階層型統計翻訳では, ルールテーブルを利用する. 与えられた日本語文にルールテーブルを用いて文脈自由文法的に翻訳を行う方式である. ツールとして moses[6] や joshua[7] がある.

### 3.1 glue-grammer

glue-grammer とは, デコーディングで与えられた文を翻訳開始から翻訳終了へ導くためのルールである. rule-grammer を作成するための翻訳規則である. 以下に, glue-grammer を表 2 に示す. 左から非終端記号, 日本語パターン, 英語パターンを示している.

表 2 glue-grammer の例

[S]		[X,1]		[X,1]		
[S]		[S,1]	[X,2]		[S,1]	[X,2]

### 3.2 rule-table

rule-table は, HSMT における翻訳モデルである. 日本語パターンから英語パターンへの翻訳を行うためのモデルが管理されている. rule-table の例を表 3 に示す. 左から非終端記号, 日本語パターン, 英語パターン, 日英フレーズ翻訳確率, 日英単語辞書確率, 英日単語辞書確率を示している.

表 3 rule-table

[X]		あなたの		your		1.0791	2.5125	0.6910		
[X]		あなたは	[X,1]		You	[X,1]		0.7403	2.5435	1.0440
		[X]		あなたが	[X,1]	を	[X,2]			
		you	[X,2]	the	[X,1]		-0.0	2.7378	0.7686	

## 4 実験の目的

渡辺ら [18] は日本語, 英語間における特許情報の翻訳を行っている. その研究の中で階層型統計翻訳と句に基づく統計機械翻訳の比較を行っている. しかし特許情報は, 文が長いと人手評価が困難である. そこで, 本実験では日本語の単文, 重文複文における HSMT と PSMT の比較調査を行う.

## 5 実験

実験は, 単文と重文複文の 2 種類で行う.

## 5.1 実験データ

### a. 単文コーパス

実験に、辞書の例文より抽出した単文コーパス 182,899 文 [19] から、学習データとして 100,000 文、テストデータとして 10,000 文、ディベロップメントデータとして 1,000 文を用いる。統計翻訳の前処理として、各コーパスの日本語文に対して、MeCab[17] を使用し形態素解析を行う。

### b. 重文複文コーパス

実験には、辞書の例文より抽出した重文複文コーパス 122,719 文 [20] から、学習データとして 100,000 文、テストデータとして 10,000 文、ディベロップメントデータとして 1,000 文を用いる。統計翻訳の前処理として、各コーパスの日本語文に対して、MeCab[17] を使用し形態素解析を行う。

## 5.2 評価方法

評価方法には、自動評価と人手評価を用いる。

### a. 自動評価

自動評価には BLEU[10], NIST[11], METEOR[12], TER[13], WER[13], IMPACT[14], RIBES[15] を用いる。BLEU は、語の並びが近いほど大きなスコアを高いスコアをつける。METEOR は、人手評価との相関が高い場合に高いスコアをつける。TER と WER は、エラー率が多いほど高いスコアをつける。IMPACT は、名詞句のチャンクを用いて翻訳出力に対するスコアをつける。RIBES は、文全体の大局的な並びが近い場合に高いスコアをつける。

### b. 人手評価

単文、重文複文における PSMT と HSMT の翻訳出力を人手で対比較評価する。手順として、はじめに PSMT の翻訳結果 10,000 からランダムで 100 文を抽出する。次に、抽出した文と対応する HSMT の翻訳結果と PSMT の翻訳結果を一文毎に人手評価する。

## 5.3 使用デコーダ

日本語の単文、重文複文における HSMT と PSMT の比較調査に moses[6] を用いる。

## 6 実験結果

### 6.1 単文コーパスを用いた実験

単文コーパスを用いた PSMT の自動評価結果を表 4 に示す。

表 4 単文における PSMT の自動評価結果

BLEU	NIST	METEOR	IMPACT	RIBES
0.1249	4.6376	0.4352	0.4285	0.6944

TER	WER
71.970	74.849

### 6.1.1 単文コーパスを用いた HSMT の自動評価結果

単文コーパスを用いた HSMT の自動評価結果を表 5 に示す。

表 5 単文コーパスを用いた HSMT の自動評価結果

BLEU	NIST	METEOR	IMPACT	RIBES
0.1372	4.8640	0.4595	0.4453	0.7148

TER	WER
70.348	73.049

### 6.1.2 単文コーパスを用いた PSMT と HSMT の人手評価

単文コーパスを用いた HSMT の人手評価結果を表 6 に示す。

表 6 単文コーパスを用いた PSMT と HSMT の人手評価結果

PSMT	HSMT	差なし	同一出力
12	22	54	12

単文コーパスを用いた翻訳結果は、自動評価、人手評価ともに HSMT が高いスコアとなった。

単文コーパスを用いた翻訳評価基準を表 7~10 に示す。

表 7 PSMT の結果が HSMT より優れている例

入力文	我々は早く決定することに意見が一致した。
正解文	We agreed to decide quickly .
PSMT	We agreed on a quick decision .
HSMT	We are agreed to a quick decision .

表 8 HSMT の結果が PSMT より優れている例

入力文	彼は店の前に看板を立てる。
正解文	He will put up a sign in front of his store .
PSMT	He Set up a sign in front of the store .
HSMT	He in front of the store or .

表 9 PSMT の結果と HSMT の結果が変わらない例

入力文	急に走るのは心臓に悪い。
正解文	It is bad for your heart to suddenly start running .
PSMT	I suddenly the heart in the wrong .
HSMT	I suddenly for the heart .

表 10 PSMT と HSMT の結果が同一の文である例

入力文	:彼の飛行機の出発時間を忘れた。
正解文	: I am pledged to secrecy .
PSMT	: I pledged to keep the secret .
HSMT	: I pledged to keep the secret .

### 6.2 重文・複文コーパスを用いた実験

#### 6.2.1 重文複文コーパスを用いた PSMT の自動評価結果

重文複文コーパスを用いた PSMT の自動評価結果を表 11 に示す。

表 11 重文複文コーパスを用いた PSMT の自動評価結果

BLEU	NIST	METEOR	IMPACT	RIBES
0.1032	4.2991	0.3985	0.372	0.6425
TER	WER			
80.070	84.585			

### 6.2.2 重文複文コーパスを用いた HSMT の自動評価結果

重文複文コーパスを用いた HSMT の自動評価結果を表 12 に示す。

表 12 重文複文コーパスを用いた PSMT と HSMT の自動評価結果

BLEU	NIST	METEOR	IMPACT	RIBES
0.1220	4.5702	0.42515	0.3972	0.6802

TER	WER
76.582	80.336

### 6.2.3 重文・複文コーパスを用いた人手評価

重文・複文コーパスを用いた HSMT の人手評価結果を表 13 に示す。

表 13 重文・複文コーパスを用いた人手評価

PSMT	HSMT	差なし	同一出力
7	26	64	3

重文複文コーパスを用いた翻訳結果は、自動評価、人手評価ともに HSMT が高いスコアとなった。

重文複文コーパスを用いた翻訳評価基準を表 14~16 に示す。

表 14 PSMT の結果が HSMT より優れている例

入力文	私は 旅行に出かけたくてたまらなかった。
正解文	I was impatient to start on the trip .
PSMT	I want to go on a trip .
HSMT	I was dying I left on a trip .

表 15 HSMT の結果が PSMT より優れている例

入力文	英語に親しむように努めている。
正解文	I am trying hard to get more familiar with English .
PSMT	began to drink habitually in English .
HSMT	He tried to enjoy English .

表 16 PSMT の結果と HSMT の結果に差がない例

入力文	家中搜したが、見つからなかった。
正解文	I looked all over the house , but could not find it .
PSMT	for the , but I did not find it .
HSMT	He looked all , but I could not find it .

表 17 PSMT と HSMT の結果が同一の文である例

入力文	私は絶対に秘密を守ると誓っている。
正解文	I am pledged to secrecy .
PSMT	I pledged to keep the secret .
HSMT	I pledged to keep the secret .

## 7 考察

### 7.1 日本語の単文・重文複文における HSMT の調査

HSMT と PSMT では、単文コーパス・重文複文コーパスにおいて自動評価・人手評価共に HSMT が良い結果となった。特に、重文複文の人手評価の BLEU 値と人手評価で大きな差があった。この原因について今後調査が必要である。

### 7.2 人手ルール追加実験

人手で作成したルールを追加する HSMT の実験をおこなった。人手で作成したルールとして、重文複文句型パターン辞書（鳥バンク）のルール 1,000 パターンを新たに追加し、デコーダには joshua を用いた。

コーパスは、重複文のディベロップメントデータ 1,000 文、テストデータ 10,000 文、トレーニングデータ 100,000 文を用いた。

手順を以下に示す。

- 手順 1：鳥バンクから、ルールを抽出
- 手順 2：抽出したルールを、glue-grammer へ追加
- 手順 3：ルールを追加した glue-grammer で、パラメータチューニングとデコーディングを行う

なお、ベースラインは、同じコーパスを用いた HSMT の翻訳結果とする。提案手法の結果を表 18 に示す。

実験の結果、テスト文 10,000 文中 5 文の翻訳が変化した。したがって翻訳精度は、ほとんど変化していない。

表 18 鳥バンクルール追加実験自動評価

	BLEU	NIST	METEOR
ベースライン	0.0847	3.9548	0.3927
提案手法	0.0847	3.9547	0.3928

## 8 おわりに

本研究では、階層型統計翻訳の文法が異なる言語間での翻訳精度を期待されているところに注目し、単文と重文複文の日英統計翻訳の場合、句に基づく統計機械翻訳と翻訳精度に違いが出るのかを調査した。結果として、単文、重文複文共に階層型統計翻訳よりも良い結果となった。特に、重文・複文コーパスにおける階層型統計翻訳での実験での人手評価が高かった。

このように差が出た理由として、階層型統計翻訳の翻訳結果は述語がよく補完されていた。このため、主語+述語ができていたために評価が高くなったと考えられる。

## 参考文献

- [1] Chiang, David. "A hierarchical phrase-based model for statistical machine translation." In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL05). pp.263-270.2005.
- [2] Chiang, David. "Hierarchical phrase-based translation," Computational Linguistics, 33(2). 2007
- [3] Tsuyoshi Okita, Andy Way "Statistical Machine Translation with Terminology"
- [4] Michel Galley, Christopher D. Manning. "A Simple and Effective Hierarchical Phrase Reordering Model"
- [5] 乗松 潤矢 "統計的機械翻訳における階層フレーズモデルの書換え規則の検討"
- [6] Koehn, Philipp, H. Hoang, etc. Moses: "Open source toolkit for Statistical Machine Translation." Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, Prague, Czech Republic, pp. 177–180. (2007)
- [7] joshua <http://cs.jhu.edu/~ccb/joshua/>
- [8] Franz Josef Och "Minimum Error Rate Training in Statistical Machine Translation" Association for Computational Linguistics, pp.160-167, (2003)
- [9] giza-pp-v1.0.3.tar.gz  
<http://www.fjoch.com/GIZA++.html>
- [10] Kishore Papineni etc. "BLEU: a Method for Automatic Evaluation of Machine Translation", Association for Computational Linguistics, pp.311-318, (2002)
- [11] "Automatic Evaluation of Machine Translation Quality Using  $N$ -gram CoOccurrence Statistics." <http://www.nist.gov/speech/tests/mt/doc/ngram-study.pdf>.2002.
- [12] METEOR "The METEOR Automatic Machine Translation Evaluation System" <http://www.cs.cmu.edu/~alavie/>
- [13] Gregor Leusch, Nicola Ueffing and Hermann Ney. "A Novel String-to-String Distance Measure with Applications to Machine Translation Evaluation." In Proc. of MT Summit IX, 240–247. TRANSLATION ERROR RATE (TER) 7.0 <http://www.cs.umd.edu/~snover/tercom/> (2003)
- [14] INPACT "Automatic Evaluation for Machine Translation" <http://www.eli.hokkai-s-u.ac.jp/echi>
- [15] RIBES "Rank-based Intuitive Bilingual Evaluation Measure" <http://www.kecl.ntt.co.jp/icl/lirg/ribes>
- [16] SRILM "The SRI Language Model Toolkit" <http://www.speech.sri.com/projects/srilm>
- [17] MeCab <http://mecab.sourceforge.net/>
- [18] Taro Watanabe, Jun Suzuki, Hajime Tsukada and Hideki Isozaki. 2007. "Online Large-Margin Training for Statistical Machine Translation" In EMNLP-CoNLL 2007 pp. 764-773.
- [19] 西山七絵, 村上仁一, 徳久雅人, 池原悟, "単文型パターン辞書の構築", 言語処理学会第11回年次大会, pp.372-375, (2005)
- [20] 村上仁一, 池原悟, 徳久雅人, "日本語英語の文対応の対訳データベースの作成", 「言語, 認識, 表現」第7回年次研究会, (2002)