2008－NL－188 (12)
2008／11／27

統計翻訳における，単文と重文複文の翻訳精度の評価

猪澤雅史　村上仁一
徳久雅人　池原悟

鳥取大学 工学部 知能情報工学科
〒 680-8552 鳥取市湖山町 4-101
E-mail:{s042009,murakami,tokuhisa,ikehara}@ike.tottori-u.ac.jp

あらまし：統計翻訳を用いた研究において，旅行会話タスクと特許翻訳タスクといったドメインの違いによる翻訳精度の報告がされている．しかし，単文と重文といった文の構造の違いによる翻訳精度の報告は見当たらなかった．本研究では，辞書の例文から抽出した対訳データを単文と重文複文を分類し，それぞれをテストデータと学習データに用いて翻訳精度を比較した．その結果，重文複文の翻訳において，学習データには単文が有効であるという結果を得た．

Evaluation of Translation Quality of Single and Compound/Complex Sentence in Statistical Machine Translation

Masafumi Izawa , Jin'ichi Murakami , MasatoTokuhisa and Satoru Ikehara

Faculty of Engineering, Tottori University
Minami 4-101, Koyama-cho, Tottori-shi, 680-8552, Japan
E-mail:{s022033,murakami,tokuhisa,ikehara}@ike.tottori-u.ac.jp

Abstract : In statistical machine translation, the evaluations of translation quality of domains such as ATR's Travel Task and Patent Translation Task was investigated. But the evaluation of translation quality of sentence construction has been investigated. In this study, We classified parallel corpus extracted from example sentence of electronic dictionary into single and compound/complex sentence, and compared their translation qualities. As a result, We obtained the result that single sentence is effective as training data in translation of compound/complex sentence.

1. はじめに

現在，機械翻訳において，対訳データから自動的に翻訳規則を獲得し，翻訳を行う統計翻訳[1]が注目されている．そして，日英統計翻訳では，文の構造が単純な単文の翻訳において，高い翻訳精度が得られている[2]．しかし，重文複文や長文といった構造が複雑な文の翻訳では，翻訳精度が低い．また，旅行会話タスク[3]と特許翻訳タスク[4]といったドメインの違いによる翻訳精度の報告はされているが，単文と重文といった文の構造の違いによる翻訳精度の報告は見当たらなかった．

そこで，本研究では，辞書の例文から抽出した対訳データ[8]を単文と重文複文に分類し，それぞれをテストデータと学習データに用いて翻訳精度の比較を行う．実験の結果，重文複文の翻訳において，学習データとして単文が有効であることがわかった．

2. 日英統計翻訳システム

2.1 基本的な考え方

日英統計翻訳は，日本語文 j が与えられたとき，全ての組み合わせの中から確率が最大になる英語文 $\hat{e}$ を探索することによって翻訳を行う．以下に基本モデルを示す．

$$\hat{e} = argmax_e P(e \mid j)$$
$$\simeq argmax_e P(j \mid e)P(e)$$

$P(j \mid e)$ は翻訳モデル，$P(e)$ は言語モデルと呼ぶ．また，$\hat{e}$ を探索する翻訳システムをデコーダと呼ぶ．

2.2 翻訳モデル

翻訳モデルは日本語の単語列から英語の単語列へ確率的に翻訳を行うためのモデルである．翻訳モデルには，大きくわけて語に基づく翻訳モデルと句に基づく翻訳モデル[5]がある．初期の統計翻訳は，語に基づく翻訳モデルを用いていた．しかし，翻訳精度の高さから，現在は句に基づく翻訳モデルが主流となっている．句に基づく翻訳モデルは表 1 に示すフレーズテーブルと呼ばれる表で管理される．

表 1 フレーズテーブルの例

庭 ||| the garden ||| 0.043 0.38 0.05 0.02
庭 から ||| from the garden ||| 0.5 0.17 1 0.03
庭 で ||| at garden ||| 0.333 0.165 0.0833 0.048

左から，日本語フレーズ，英語フレーズ，フレーズの英日翻訳確率 $P(j \mid e)$，英日方向の単語の翻訳確率 (IBM モデル) の積，フレーズの日英翻訳確率 $P(e \mid j)$，日英方向の単語の翻訳確率 (IBM モデル) の積である．本稿では，日本語フレーズ，英語フレーズ，各種確率の 3 つをまとめて，フレーズ対と呼ぶ．

2.3 言語モデル

言語モデルは単語列に対して，それらが起こる確率を与えるモデルである．日英翻訳では，言語モデルを用いて，訳文候補の中から英語として自然な文を選出する．言語モデルとして代表的なものに N-gram モデルがある．

3. 実験環境

3.1 実験データ

実験には，辞書の例文から抽出した対訳データを，単文と重文複文に分類したコーパスを用いる．分類は日本語文のみを見て行われ，単文コーパスは 182,899 文[7]，重文複文コーパスは 122,719 文[8] である．また，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は chasen[6] を用いて形態素解析を行う．また，英語文に対しては句読点の前後にスペースを入れる．一般に，英語文に対しては，大文字の小文字化を行うが，本研究では行わない．単文コーパスと重文複文コーパス中の対訳文の例を表 2 に示す．

表 2 単文コーパスと重文複文コーパスの例

単文コーパス
ぶどう酒 は 葡萄 より 作ら れる 。
Wine is made from grapes .
花子 は 、 悲し そう に 俯い て い た 。
Hanako appeared sad and downcast .
娘 は 今年 中学校 に 上がっ た 。
My daughter advanced to middle school this year .
重文複文コーパス
彼 は 偏見 が あり その ため 信頼 でき なかっ た 。
He was biased , and so unreliable .
その 鳥 は 山 を 越え て 飛ん で いっ た 。
The bird winged its flight over the hills .
急い で い て 彼女 に 大事 な こと を 言い 忘れ た 。
I was in such a hurry I forgot to tell her the most important thing .

3.2 フレーズテーブルの学習

フレーズテーブルの学習には，多くの方法がある．本研究では，“train-phrase-model.perl[10]” を用いる．このプログラムは IBM model1〜5[1] に基づく，“GIZA++[9]” を利用している．また，フレーズテーブルを生成する際，フレーズテーブル内の日本語と英語のフレーズ中の単語数の上限値として，max-phrase-length が定義されている．例えば，max-phrase-length の値が 7 の場合，日本語か英語のいずれかのフレーズ中の単語数が，8 以上のフレーズ対は生成されない．

3.3 N−gram モデルの学習

言語モデルは，N−gram モデルを用いる．N−gram モデルの学習には “SRILM[11]” の ngram-count を用いる．なお，本研究ではスムージングに “-ndiscount” を用いる．

3.4 デコーダのパラメータ

デコーダは “moses[12]” を使用する．一般に，翻訳モデルの確率 $P(j \mid e)$ を得るために，フレーズテーブルの各種確率の重み “weight-t” は “1.0 0.0 0.0 0.0 0.0” を用いる．しかし，クロスエントロピーをフレーズテーブルの確率としたとき，翻訳精度は高くなる[2]．そこで，本研究では “weight-t” を “0.5 0.0 0.5 0.0 0.0” とする．また，日本語から英語への翻訳は，動詞の位置が大きく変化する．フレーズの並び替え確率の重み “distortion weight” が低いとき，デコーダは日本語から英語へ翻訳するときのフレーズの位置の変化に柔軟に対応できる．そこで，本研究では “distortion weight” を “0.2” とする．

なお，moses のパラメータは MERT を用いて最適化することができる．しかし，本研究では，全ての実験条件を同じにするために最適化を行わない．

3.5 評価方法

出力文の評価には自動評価法である BLEU[13] と METEOR[14] を使用する．BLEU は予め用意された正解文と比較して，語順が正しい場合に高いスコアを出す．また，METEOR は予め用意された正解文と比較して，単語属性が正しい場合に高いスコアを出す．両評価法とも 0 から 1 の間で評価され，1 が最も良い評価である．また，本研究では，入力文 1 文に対して正解文 1 文を用いて評価を行う．

4. 予備実験

4.1 実験データ

4.1.1 単文コーパス

単文コーパス 182,899 文から，1,000 文を Open テスト用のデータとしてランダムに抽出する．残りの 181,899 文は学習に用いる．学習データ量と精度の関係を調べるために 181,899 文から 1,000 文，5,000 文，10,000 文，50,000 文，100,000 文をランダムに抽出する．各学習データ量における，日本語と英語の総単語数を表 3 に示す．

表 3 単文コーパスの総単語数

学習データ (文)	日本語の総単語数	英語の総単語数
1,000	9,432	8,503
5,000	48,200	43,053
10,000	91,460	80,307
50,000	497,893	44,4004
100,000	1,006,954	851,725
181,899	1,916,262	1,648,795

4.1.2 重文複文コーパス

重文複文コーパス 122,719 文から，1,000 文を Open テスト用のデータとしてランダムに抽出する．残りの 121,719 文は学習に用いる．単文コーパスと同様に 121,719 文から 1,000 文，5,000 文，10,000 文，50,000 文，100,000 文をランダムに抽出する．各学習データ量における，日本語と英語の総単語数を表 4 に示す．

表 3 と表 4 を比較すると，重文複文の総単語数は単文

表 4　重文複文コーパスの総単語数

学習データ (文)	日本語の総単語数	英語の総単語数
1,000	13,663	11,050
5,000	69,107	56,132
10,000	138,109	112,136
50,000	691,893	560,389
100,000	1,381,961	1,119,533
121,719	1,711,869	1,378,791

の単語数より 3 割ほど多いことがわかる．

4.2 max-phrase-length の値と翻訳精度の関係

英仏翻訳のように似た言語間では，翻訳前と翻訳後を比較して，単語の位置に大きな変更はない．しかし，日英翻訳のように動詞の位置が大きく異なる言語間では，日本語と英語のフレーズの長さは長いほうが、翻訳精度は高くなると考えられる．そこで，フレーズテーブルの学習に用いる max-phrase-length の値と翻訳精度の関係を調べる．

実験は，学習データは重文複文コーパス 121,719 文を用い，テストデータは重文複文を用いる．Closed テストと Open テストの結果を表 5 に示す．

表 5　max-phrase-length と翻訳精度の関係

max-phrase-length	Closed		Open	
	BLEU	METEOR	BLEU	METEOR
5	0.235	0.450	0.082	0.315
10	0.492	0.645	0.092	0.323
20	0.895	0.930	0.101	0.332
100	0.901	0.934	0.101	0.332

結果から，max-phrase-length の値を大きくすることで，翻訳精度が向上していることがわかる．しかし，Open テストにおいて，max-phrase-length の値が 20 の場合と，100 の間で精度の向上がない．このことから，以降の実験では max-phrase-length の値に 20 を用いる．

4.3 N-gram の次数と翻訳精度の関係

学習データが大量にある場合，高次の N-gram モデルほど翻訳精度が高くなると考えられる．しかし，一般に，学習データ量は限られるため，高次の N-gram モデルはパラメータの数が多くなり，信頼性が低くなる．そこで，N-gram の次数と翻訳精度の関係を調べる．

実験は，学習データは単文コーパス 181,899 文を用い，テストデータは単文を用いる．Closed テストと Open テストの結果を表 6 に示す．

表 6　N-gram の次数と翻訳精度の関係

$N-$gram の次数	Closed		Open	
	BLEU	METEOR	BLEU	METEOR
1-gram	0.699	0.839	0.059	0.293
2-gram	0.789	0.856	0.113	0.366
3-gram	0.847	0.895	0.135	0.386
4-gram	0.852	0.890	0.139	0.388
5-gram	0.853	0.901	0.139	0.389
6-gram	0.853	0.901	0.139	0.388
7-gram	0.853	0.901	0.139	0.388

結果から，$N-$gram の次数を大きくすることで，5-gram までは翻訳精度が向上していることがわかる．しかし，6-gram 以上では翻訳精度は向上していない．このことから，以降の実験では 5-gram モデルを用いる．

4.4 $N-$gram モデルの学習データ量と翻訳精度の関係

統計翻訳では様々な研究が行なわれているが，日英翻訳において，$N-$gram モデルの学習データ量と翻訳精度の関係の報告は見あたらなかった．そこで，フレーズテーブルの学習データ量を一定にしたときの，$N-$gram モデルの学習データ量と翻訳精度の関係を調査する．

実験は，テストデータと学習データに単文を用いて行う．フレーズテーブルの学習データ量は常に 181,899 文とし，$N-$gram モデルの学習データ量を 1,000 から 181,899 文に変えて翻訳実験を行う．Closed テストと Open テストの結果を表 7 に示す．

表 7　$N-$gram モデルの学習データ量と翻訳精度の関係

学習データ (文)	Closed		Open	
	BLEU	METEOR	BLEU	METEOR
1,000	0.979	0.985	0.053	0.264
5,000	0.954	0.967	0.077	0.305
10,000	0.935	0.953	0.085	0.316
50,000	0.887	0.923	0.116	0.363
100,000	0.870	0.911	0.130	0.379
181,899	0.853	0.901	0.139	0.389

結果から，フレーズテーブルの学習データ量を一定にし，$N-$gram モデルの学習データ量を増加させた場合に，学習データ量に対して，翻訳精度はほぼ線形に変化することがわかる．

4.5 フレーズテーブルの学習データ量と翻訳精度の関係

日英翻訳において，フレーズテーブルの学習データ量と翻訳精度の関係の報告は見当たらなかった．そこで，$N-$gram モデルの学習データ量を一定にしたときの，フレーズテーブルの学習データ量と翻訳精度の関係を調査する．

実験は，テストデータと学習データに単文を用いて行う．$N-$gram モデルの学習データ量は常に 181,899 文とし，フレーズテーブルの学習データ量を 1,000 から 181,899 文に変えて翻訳実験を行う．Closed テストと Open テストの結果を表 8 に示す．

表 8　フレーズテーブルの学習データ量と翻訳精度の関係

学習データ (文)	Closed		Open	
	BLEU	METEOR	BLEU	METEOR
1,000	0.988	0.992	0.012	0.126
5,000	0.980	0.986	0.032	0.200
10,000	0.973	0.983	0.044	0.226
50,000	0.914	0.943	0.093	0.321
100,000	0.882	0.920	0.113	0.353
181,899	0.853	0.901	0.139	0.389

結果から，言語モデルを一定にし，翻訳モデルの学習データ量を増加した場合，翻訳精度が非線形に変化することがわかる．特に，学習データ量が少ないときに，精度の変化が大きい．

5. 単文と重文複文の翻訳精度の評価

5.1 目　　的

日英統計翻訳において，旅行会話タスク[3] と特許翻訳タスク[4] といったドメインの違いによる翻訳精度が報告れている．しかし，単文と重文複文といった文の構造の違いによる翻訳精度の比較は報告はされていない．

重文複文には単文にない表現が存在することから，重文複文の翻訳は学習データとして重文複文を用いることが有効であると考えられる．しかし，重文複文は文の構造が複雑であるため，適切なフレーズ対は多く得られない可能性がある．

そこで，単文と重文複文コーパス各 100,000 文を学習データに，10,000 文をテストデータに用いて，それぞれの翻訳精度の比較を行う．

5.2 実験データ

5.2.1 単文コーパス

4 章と同様に，実験には辞書の例文から抽出した単文コーパス 182,899 文を用いる．単文コーパスから，10,000 文を Open テストデータとしてランダムに抽出する．学習データは 100,000 文を残りの 172,988 文からランダムに抽出する．テストデータと学習データの日本語と英語の総単語数を表 9 に示す．

表 9 単文コーパスから抽出した各データの総単語数

データの種類	日本語の総単語数	英語の総単語数
テストデータ (10,000 文)	105,706	90,621
学習データ (100,000 文)	999,020	831,391

5.2.2 重文複文コーパス

4 章と同様に，実験には辞書の例文から抽出した重文複文コーパス 122,719 文を用いる．重文複文コーパスから，10,000 文を Open テストデータとしてランダムに抽出する．学習データは 100,000 文を残りの 112,719 文からランダムに抽出する．テストデータと学習データの日本語と英語の総単語数を表 10 に示す．

表 10 重文複文コーパスから抽出した各データの総単語数

データの種類	日本語の総単語数	英語の総単語数
テストデータ (10,000 文)	140,696	112,966
学習データ (100,000 文)	1,405,780	1,132,331

5.3 翻 訳 精 度

5.3.1 単文の翻訳精度の評価

テストデータに単文を用いて翻訳実験を行う．また，学習データは単文と重文複文をそれぞれ用いる．各学習データにおける，単文の翻訳精度を表 11 に示す．

表 11 単文の翻訳精度 (テストデータ:10,000 文)

学習データ	BLEU	METEOR
単文	0.1159	**0.3468**
重文複文	**0.1250**	0.3304

結果から，単文の翻訳において，BLEU は学習データに重文複文を用いた時の方がスコアが高い．しかし，METEOR は学習データに単文を用いた時にスコアが高くなることがわかる．

5.3.2 重文複文の翻訳精度の評価

テストデータに重文複文を用いて翻訳実験を行う．また，学習データは単文と重文複文をそれぞれ用いる．各学習データにおける，重文複文の翻訳精度を表 12 に示す．

表 12 重文複文の翻訳精度 (テストデータ:10,000 文)

学習データ	BLEU	METEOR
単文	**0.0947**	0.2960
重文複文	0.0933	**0.3195**

結果から，BLEU は学習データに単文を用いた時にスコアが高い．しかし，METEOR は学習データに重文複文を用いた時にスコアが高くなることがわかる．

5.4 人 手 評 価

5.4.1 評 価 基 準

表 11 と表 12 の翻訳結果について，人手による 4 段階の評価を行う．評価基準と評価例を以下に示す．なお，固有名詞の未知語はローマ字変換して評価する．

評価 A　出力文は入力文の訳文として問題がない．

入力文　私 が たいへん 驚い た こと に 彼 は その 試験 に 合格 し た 。
正解文　To my great surprise , he passed the examination .
出力文　I was surprised that he passed the examination .

評価 B　出力文は文法的に間違っていたり，情報が欠けていたりと，不自然さを伴うが，入力文で伝えたい情報が容易に理解できる．

入力文　彼 は 偏見 が あり その ため 信頼 でき なかっ た 。
正解文　He was biased , and so unreliable .
出力文　He has a bias It was unreliable .

評価 C　出力文は多くの情報が抜けているが重要な情報は含まれており，文脈や断片的な情報から入力文で伝えたい情報が理解できる．

入力文　彼 が 生まれ た の は 第 2 次 大戦 たけなわ の 頃 の こと だっ た 。
正解文　It was at the height of World War II that he was born .
出力文　He was born in World War II たけなわ when it was .

評価 D　出力文では伝えたい情報が理解できない．または，異なって理解される．

入力文　注文 を 決める 前 に 念入り に メニュー に 目 を 通し なさい 。
正解文　Look through the menu carefully before you decide .
出力文　A menu his carefully before you order to the end .

5.4.2 評 価 結 果

人手評価の対象文として，各翻訳結果から 100 文ずつランダムに抽出した．評価結果を表 13 に示す．

表 13 人手評価

入力文	単文の翻訳		重文複文の翻訳	
学習データ	単文	重文複文	単文	重文複文
評価 A	6/100	9/100	3/100	3/100
評価 B	23/100	12/100	19/100	15/100
評価 C	33/100	39/100	32/100	33/100
評価 D	38/100	40/100	46/100	49/100

結果から，単文の翻訳においても，重文複文の翻訳においても，単文を学習データとして用いたとき，評価 D の出力が少ないことがわかる．このことから，人手評価においても，重文複文の翻訳において，学習データとして単文が有効であることがわかる．

6. 考　　察

6.1 単文と重文複文の単語数の違い

表 9 と表 10 を比較すると，文の数が同じであっても，単文の総単語数と重文複文の総単語数に大きな違いがある．そこで，単文と重文複文の学習データの総単語数を等しくするために，単文の学習データに，学習データとテストデータに使用していない 40,000 文を追加し，翻訳実験を行った．その結果，学習データに単文 140,000 文を用いたとき，単文の翻訳では，BLEU スコアで 0.1298 を，重文複文の翻訳では，BLEU スコアで 0.1177 という結果を得た．つまり，総単語数が等しいとき，重文複文の翻訳において，学習データとして単文が有効であることが明確にわかった．

6.2 フレーズテーブルと N−gram モデルにおける学習データの分類

5 章では，フレーズテーブルと N−gram モデルの学習データに同じデータを用いた．この節では，フレーズテーブルと N−gram モデルの学習データに単文と重文複文をそれぞれ用いて翻訳実験を行う．

6.2.1 各学習データにおける翻訳精度の評価

テストデータは単文と重文複文を用いる．また，学習データは表 9 の学習データと表 10 の学習データを用いる．単文と重文複文の翻訳精度を表 14 に示す．

表 14 各テスト文の翻訳精度 (テストデータ 10,000 文)

単文の翻訳			
フレーズテーブル	N−gram モデル	BLEU	METEOR
単文	単文	0.1159	**0.3468**
	重文複文	0.1003	0.3307
重文複文	重文複文	**0.1250**	0.3304
	単文	0.1223	0.3343
重文複文の翻訳			
フレーズテーブル	N−gram モデル	BLEU	METEOR
単文	単文	**0.0947**	0.2960
	重文複文	0.0933	0.2991
重文複文	重文複文	0.0933	**0.3195**
	単文	0.0772	0.2997

結果から，単文の翻訳はフレーズテーブルの学習データに重文複文を用いたとき，翻訳精度が高いことがわかる．また，重文複文の翻訳はフレーズテーブルの学習データに単文を用いたとき，翻訳精度が高いことがわかる．しかし，単文の翻訳においても，重文複文の翻訳においても，フレーズテーブルの学習データと N−gram モデルの学習データが異なるとき，翻訳精度が低い．特に，N−gram モデルの学習データがテストデータと異なる構造の文であるとき，翻訳精度は極めて低い．この原因について次のように考えている．

統計翻訳において，フレーズテーブルは単語列の翻訳の正しさを，N−gram モデルは生成された文の正しさを決める．しかし，フレーズテーブルの学習データと N−gram モデルの学習データが異なるとき，フレーズテーブル中に存在する英単語が N−gram モデル中にない可能性がある．そのため，N−gram モデルは文の正しさを適切に判断できず，翻訳精度が低い．

6.3 重文複文コーパスの意訳の問題点

単文の翻訳においても，重文複文の翻訳においても，学習データに重文複文を用いたとき，意味が理解できる出力文 (評価 A，B，C) は少なかった．この原因について，次のように考えている．

本研究では単文と重文複文を分類する際，日本語文のみを用いて分類を行った．そのため，重文複文の対訳データにおいて，日本語文が重文複文であっても，英語文は意訳され単文となっていることが多い．そのため，不適切なフレーズ対が多く生成され，意味が理解できない文が多く生成された．

重文複文コーパス中の意訳の例を表 15 に示す．

表 15 重文複文コーパス中の意訳の例

彼は愛車を駆って横浜へ行った。
He drove his car to Yokohama.
ウオツカはちびりちびり飲まないで一気にあおるんだ。
Do not sip your vodka.
ぼくは女の子の前に出るとすぐあがってしまう。
I am very nervous around girls.
彼は悪意があって言ったのではない。
He did not mean any harm.
うそをつくことはいやしむべき悪徳である。
Lying is a despicable vice.

6.4 不適切なフレーズ対の問題点

学習データに重文複文を用いたとき，訳出文には誤訳が多い．例を以下に示す．

入力文	この 商品 は 値段 を 1 0 パーセント 下げ た 。
正解文	The price of this product was lowered by 10 percent .
出力文	This product is a reduced price to 50 .

出力文には，入力文に対して明らかに不適切な単語 “50” が出力されている．出力文に用いられたフレーズ対を表 16 に示す．

表 16 不適切なフレーズ対を含むフレーズテーブル

日本語のフレーズ ||| 英語のフレーズ ||| 各種確率
この 商品 は ||| This product is ||| 0.1111 0.0397 0.125 0.0165
値段 ||| price ||| 0.3310 0.2259 0.6076 0.3931
を ||| to ||| 0.2878 0.2656 0.2601 0.1782
1 ||| a ||| 0.0115 0.0163 0.3483 0.1518
0 パーセント ||| **50** ||| 0.0417 0.0056 0.5 0.0180
下げ た ||| reduced ||| 0.1667 0.0021 1 0.0212
。 ||| . |||0.9973 0.9849 0.9921 0.9922

表から，“50” が “0 パーセント” に対する誤訳であることがわかる．そこで，フレーズテーブルから不適切なフレーズ対 “0 パーセント ||| 50” を削除すると以下の訳出文が得られる．

This product is reduced price of 10 percent .

重文複文は，単文に比べて 1 文中の単語数が多い．さらに，文の構造の複雑さと 6.3 節の意訳の問題もある．そのため，不適切なフレーズ対も多く生成される．そして，出力文のように誤訳を含む文が出力されたと考えている．

このことから，学習データとして重文複文を用いたフレーズテーブルは，クリーニングを行い，不適切なフレーズ対を削除する必要があると考えている．

6.5 実験データの問題点

単文の翻訳においても，重文複文の翻訳のおいても，テストデータと同じ構造の文を学習データとして用いたとき，METEOR スコアは高い．この原因について，次のように考えている．

本研究に使用した実験データは辞書から抽出した対訳データである．辞書の例文には，似た表現が多く存在する．そのため，テストデータと同じ構造の文を学習データに用いたとき，未知語が少なくなり，METEOR スコアが高くなった．

単文コーパスと重文複文コーパス中の似た表現の例を表 17 に示す．

表 17 各コーパス中の似た表現の例

単文コーパス	
水溜まりが出来る。	A puddle is made.
行列が出来る。	A line is formed.
運が傾く。	One's luck is on the wane.
運が傾く。	One's luck is on the decline.
夜が明ける。	The day breaks.
年が明ける。	A new years begins.
重文複文コーパス	
嘔吐を催すような匂いだ。	It is a sickly smell.
嘔吐を催すような光景だ。	It is a disgusting sight.
彼は来るだろうと思った。	I thought he would come.
君は来るだろうと思った。	I knew you would come.
彼が走っているのを見た。	I saw him run away.
彼が走っているのを見た。	I saw him running.

6.6 未知語の問題

表 11 と表 12 の訳出文には，未知語が多く存在し，翻訳精度を下げる原因となっている．例を以下に示す．また，訳出文の未知語の数を表 18 に示す．

入力文 1	ご 子息 の こと で 内談 が あっ て 参り まし た 。
出力文 1	As a matter of 子息 内談 I have come .
入力文 2	彼 は 9つ だ が すぐ 1 0 に なる 。
出力文 2	The 9つ He is quick to be ten .

表 18 各訳出文の未知語の総数

単文の翻訳		
学習データ	総数	未知語を含む文数
単文	7,319	4,844
重文複文	8,190	5,282
重文複文の翻訳		
学習データ	総数	未知語を含む文数
単文	11,353	6,559
重文複文	8,048	5,250

表 18 から，単文の翻訳には単文を，重文複文の翻訳には重文複文を学習データに用いた方が未知語が少ないことがわかる．これは，文の構造が同じであることだけでなく，6.5 節の実験データの問題が原因であると考えている．

また，単文と重文複文のどちらを学習データとして用いた場合も，未知語は多く存在する．特に重文複文の翻訳では，テストデータの単語数が多いため，単文と比較して多くの未知語が出力されている．このことから，より大規模なコーパスを用いて，翻訳実験を行う必要がある．もしくは，未知語処理が必要であると考えている．

7. おわりに

本研究では，単文 181,899 文，重文複文 121,719 文を用いて，単文と重文複文の翻訳精度の比較を行なった．

実験の結果，重文複文の翻訳において，学習データには重文複文を用いるよりも，単文を用いるほうが有効であるということが確認できた．また，重文複文コーパスは意訳を含んでおり，学習データとして重文複文を用いたとき，生成されるフレーズテーブルには誤訳が多いため，入力文に対して適切な翻訳が得られないことがわかった．

今後は，フレーズテーブルのクリーニングを行ない，有効性を調査することを考えている．

参考文献

1) Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer, "The Mathematics of Statistical Machine Translation, Parameter Estimation", Computational Linguistics, 19(2), 1993.
2) Jin'ichi Murakami, Masato Tokuhisa, Satoru Ikehara, "Statistical Machine Translation using Large J/E Parallel Corpus and Long Phrase Tables", International Workshop on Spoken Language Translation 2007, pp.151-155, 2007.
3) Yasuhiro Akiba, Marcello Federico, Noriko Kando, Hiromi Nakaiwa, Michael Paul, and Jun'ichi Tsujii, "Overview of the IWSLT04 evaluation campaign", In Proceedings of the International Workshop on Spoken Language Translation (IWSLT), pp. 1-12, 2004.
4) 内山将夫, 山本幹雄, 藤井敦, 宇津呂武仁, "特許情報を対象とした機械翻訳 -共通基盤による評価タスクを目指して-", 情報処理学会研究報告, Vol. 2007, No.(2007-NL-180), pp.133-138, 2007.
5) Philipp Koehn, Franz J. Och, and Daniel Marcu, "Statistical phrase-based translation", In Proceedings of HLT-NAACL 2003, pp. 127-133, 2003.
6) chasen, http://chasen-legacy.sourceforge.jp/
7) 西山七絵, 村上仁一, 徳久雅人, 池原悟, "単文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.
8) 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.
9) GIZA++, http://www.fjoch.com/GIZA++
10) training-release-1.3.tgz, http://www.statmt.org/wmt06/shared-task/baseline.html
11) SRILM, The SRI Language Modeling Toolkit, http://www.speech.sri.com/projects/srilm
12) Moses, moses.2007-05-29.tgz, http://www.statmt.org/moses/
13) NIST Open Machine Translation, http://www.nist.gov/speech/tests/mt
14) The METEOR Automatic Machine Translation Evaluation System, http://www.cs.cmu.edu/alavie/METEOR/