

ルールベース翻訳を前処理に用いた統計翻訳

福田智大 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

{s062047,tokuhisa,murakami,ikehara}@ike.tottori-u.ac.jp

1 はじめに

近年の機械翻訳では、統計翻訳が主流となっている。統計翻訳において、言語間の文法構造が大きく異なる日英翻訳と比較して、言語間の文法構造が類似している伊英翻訳では翻訳精度が高くなる傾向がある [1]。

そこで本研究では、言語間の文法構造が類似しているときには翻訳精度が高くなる統計翻訳の長所を効果的に用いる。そのために、まず前処理として、文法構造を大きく変換する必要がある文に対しても、英語の文法構造に類似した文を出力することができるルールベース翻訳で日英翻訳を行なう。次に、英英統計翻訳を行なうことで翻訳精度の向上を試みる。

2 日英統計翻訳システム

2.1 概要

日英統計翻訳では、日本語文のフレーズを英語文のフレーズに翻訳し、翻訳されたフレーズを並び替えて翻訳候補を生成する。本研究では、統計翻訳のみで日英翻訳を行なう手法をベースラインとする。日英統計翻訳システムの枠組を図 1 に示す。

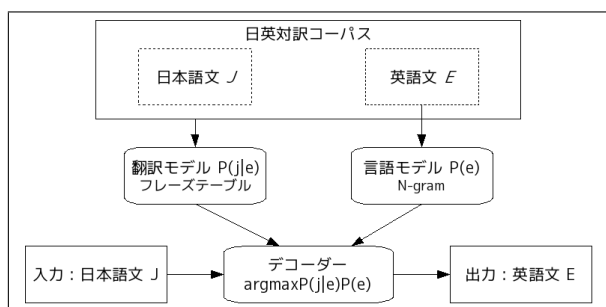


図 1 日英統計翻訳の枠組

2.2 言語モデル

言語モデルは、単語列に対して、それらの単語列が起こる確率を付与するモデルである。日英統計翻訳では、言語モデルを使用して、英語として自然な文を選出する。また言語モデルには、 N -gram モデルが一般的に使用される。

2.3 翻訳モデル

翻訳モデルは、日本語から英語へ確率的に翻訳を行なうためのモデルである。翻訳モデルでは、フレーズテーブルと呼ばれる表によって管理される。以下にその例を示す。

お婆さん		Grandmother		0.25	1	1	1
この 果物		This fruit		1	0.10	1	0.20

左から日本語フレーズ、英語フレーズ、フレーズの英日方向の翻訳確率 $P(j|e)$ 、英日方向の単語の翻訳確率の積、フレーズの日英方向の翻訳確率 $P(e|j)$ 、日英方向の単語の統計確率の積である。

2.4 デコーダ

デコーダでは、日本語文 J が与えられた場合に、翻訳モデルと言語モデルの組み合わせの中から確率が最大となる英語文 E を探索することで翻訳を行なう。

$$E = \operatorname{argmax}_e P(e|j) \\ \simeq \operatorname{argmax}_e P(j|e)P(e)$$

ここで $P(j|e)$ は翻訳モデル、 $P(e)$ は言語モデルを示す。

3 関連研究

ルールベース翻訳と統計翻訳を組み合わせる手法としては、L.Dugast ら [2] や M.Simard ら [3] の研究がある。彼らは、ルールベース翻訳として SYSTRAN を用いて翻訳を行ない、統計翻訳を後処理として用いた。その結果、SYSTRAN のみや統計翻訳のみのときと比較して、BLEU スコアが向上しており、彼らの手法の有効性が確認できた。

彼らの手法では、統計翻訳をルールベース翻訳の後処理として考えている。しかしながら、本研究では、ルールベース翻訳を統計翻訳の前処理として考える。

4 提案手法

4.1 提案手法の枠組

提案手法では、前処理として文法構造を目的言語に類似させるために、はじめに、ルールベース翻訳を使用して、日本語文を英語文に翻訳する。次に、翻訳された英語文に対して、英英統計翻訳を行なう。本研究において、この英英統計翻訳を英' 英統計翻訳と定義する。枠組を図 2 に示す。

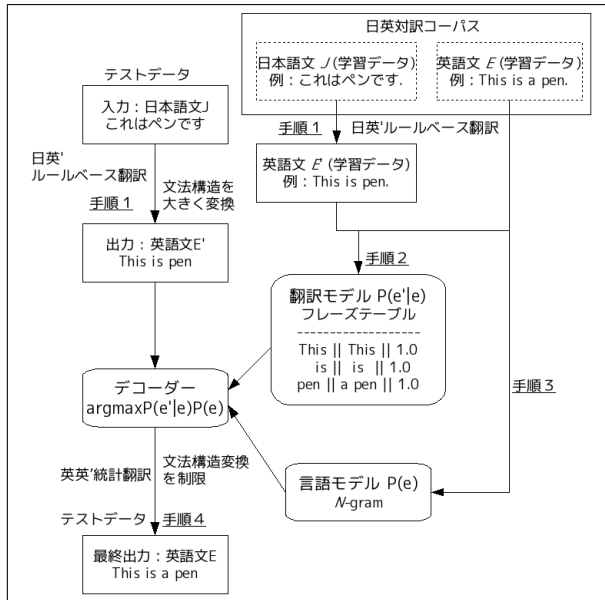


図2 提案手法の枠組

4.2 提案手法の手順

手順1 前処理としての日英'ルールベース翻訳

前処理として、日英対訳コーパス(学習データ、テストデータ、ディベロップメントデータ)の日本語文に対して、ルールベース翻訳でそれぞれ日英'翻訳を行なう。入力例、出力例、正解例を以下に示す。

入力例(日本語文)

免許をとった。
この箱はどうしても持ち上がらない。

出力例(ルールベース翻訳)

It got the license.
This box doesn't lift.

正解例(英語文)

I obtained a license.
This box won't lift.

手順2 翻訳モデルの作成

前処理を行なった学習データの英語文と日英対訳コーパスの英語文を用いて、英'英翻訳モデルを作成する。英'英フレーズテーブルの例を以下に示す。

英'英フレーズテーブルの例

It got the license ||| I obtained a license ||| (0)
(1) (2) (3) ||| (0) (1) (2) (3) ||| 1.000 0.000
1.000 0.000
This box ||| This box ||| (0) (1) ||| (0) (1) |||
0.625 0.512 0.833 0.421

手順3 言語モデルの作成

日英対訳コーパスの英語文から言語モデルを作成する。

手順4 英'英統計翻訳

手順1で作成した言語モデルと手順3で作成した翻訳モデルを用いて、統計翻訳で英'英翻訳を行なう。

5 実験環境

5.1 言語モデルの学習

言語モデルの学習には、“SRILM[4]”の“ngram-count”を用いる。本研究では、 N -gramモデルは5-gramとする。またスムージングには、“Kneser-Ney discount”を用いる。

5.2 翻訳モデルの学習

翻訳モデルの学習には、“train-factored-phrase-model.perl[5]”を用いる。

5.3 デコーダのパラメータ

デコーダには、“moses[5]”を用いる。またmosesのパラメータはパラメータチューニング[6]によって最適化する。ただし、パラメータチューニングで変更されない値である、“ttable-limit”と“distortion-limit”に関しては手作業で変更する。本研究では、“ttable-limit”の値は80とする。また“distortion-limit”の値は、ベースラインでは過去の研究により-1(無限大)とし、提案手法では文法構造の変換を制限するためにデフォルト値の6とする。

5.4 実験データ

5.4.1 単文コーパス

実験には、辞書の例文より抽出した単文コーパス182,899文[7]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、“chasen[8]”を用いて形態素解析を行なう。また英語文に対して、“tokenizer.perl[5]”を用いて、わかち書きを行なう。

5.4.2 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[9]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、“chasen[8]”を用いて形態素解析を行なう。また英語文に対して、“tokenizer.perl[5]”を用いて、わかち書きを行なう。

5.5 ルールベース翻訳

本研究では、ルールベース翻訳として、市販されている翻訳ソフトであるATLAS V12(富士通)と翻訳の王様Version5(IBM)を使用する。

5.6 評価方法

出力文の評価には自動評価法であるBLEU[10]とMETEOR[11]を使用する。また人手評価として、対比較評価を行なう。

6 実験結果

6.1 自動評価

ベースライン、ルールベース翻訳 (ATLAS) のみ、提案手法 1(ATLAS+ 統計翻訳)、ルールベース翻訳 (翻訳の王様) のみ、提案手法 2(翻訳の王様 + 統計翻訳) の自動評価法による評価結果を表 1 と表 2 に示す。

表 1 単文における評価

	BLEU	METEOR
ベースライン	0.1397	0.3844
ATLAS のみ	0.1552	0.4216
提案手法 1	0.1877	0.4418
翻訳の王様のみ	0.1206	0.3888
提案手法 2	0.1721	0.4227

表 2 重文複文における評価

	BLEU	METEOR
ベースライン	0.1175	0.3531
ATLAS のみ	0.0942	0.3623
提案手法 1	0.1364	0.3806
翻訳の王様のみ	0.0842	0.3522
提案手法 2	0.1346	0.3737

結果より、ベースラインと提案手法、ルールベース翻訳 (ATLAS) のみと提案手法 1、ルールベース翻訳 (翻訳の王様) のみと提案手法 2 を比較すると、すべて翻訳精度が向上していることが確認できる。

6.2 対比較評価

単文と重文複文において、提案手法 1 の出力文とベースラインの出力文からランダムに抽出した 100 文を用いて、対比較評価を行なった。評価基準は以下に示す。

提案手法 : 提案手法の方が良い

提案手法 × : 提案手法の方が悪い

提案手法 = : 提案手法と比較対象の文の単語対応や伝わる意味に差がない

提案手法 = : 提案手法と比較対象の文がまったく同じ
評価結果を表 3 に示す。

表 3 対比較評価

	単文	重文複文
提案手法	79 /100	48 /100
提案手法 ×	6 /100	15 /100
提案手法	12 /100	32 /100
提案手法 =	3 /100	5 /100

結果より、ベースラインと提案手法を比較したときでは、単文と重文複文の両方において、“提案手法の方が良い” が選択された数が最も多く、提案手法の効果が確認できた。

6.3 提案手法の翻訳例

提案手法によって、翻訳精度が向上した例として、単文の例を表 4 に、重文複文の例を表 5 に示す。

表 4 提案手法における単文の翻訳例

日本語文	電気コンロのコイルが焼き切れた。
正解文	The heater coil is burnt out.
ベースライン	The coil the cooking stove electricity.
提案手法 1	The coil of the electric heater has burned out.

表 5 提案手法における重文複文の翻訳例

日本語文	我々は地元チームが勝つとかけた。
正解文	We bet that the home team would win.
ベースライン	The team will win us.
提案手法 1	We are digits as the home team will win.

7 考察

7.1 自動評価と対比較評価の不整合

単文と重文複文において、提案手法 1 の出力文とルールベース翻訳 (ATLAS) のみの出力文からランダムに抽出した 100 文を用いて、対比較評価を行なった。評価基準は 6.2 章において示したものと同様とする。評価結果を表 6 に示す。

表 6 対比較評価

	単文	重文複文
提案手法	17 /100	8 /100
提案手法 ×	9 /100	42 /100
提案手法	54 /100	48 /100
提案手法 =	20 /100	2 /100

結果より、単文における対比較評価では、“提案手法の方が良い” と選択された数が、“提案手法の方が悪い” と選択された数よりも、わずかに多い。しかしながら、“差がない” とされた数が最も多く、提案手法の効果はあまり確認できなかった。

重文複文における対比較評価では、“提案手法の方が悪い” と選択された数が、“提案手法の方が良い” と選択された数よりも多い。この原因として、重文複文では文法構造が複雑なため、ルールベース翻訳での前処理において、英語の文法構造に類似させることができなかったと考えている。このため、統計翻訳において、低精度に翻訳され、提案手法の翻訳精度が低下したと考えている。

しかしながら、提案手法における BLEU スコアでは、4.22% の向上が確認されている。この結果は人手による対比較評価の結果と整合してしない。この原因として、学習データとテストデータが同じ分野から抽出した文のため、統計翻訳によって、分野に適応し、BLEU スコアが向上したと考えている。なお、BLEU スコアでは 4-gram を用いて評価を行なっているため、分野に適応した文では、分野に適応していない文と比較して、BLEU スコアが高くなる傾向がある。

したがって、自動評価と対比較評価の不整合は、統計翻訳によって分野に適応した文が、BLEU スコアにおいて有利に評価されるといった問題から生じたと考えている。

7.2 未知語の影響

本研究では、市販されている翻訳ソフトを用いるため、翻訳ソフト固有の辞書データを利用している。したがって、統計翻訳のみのときと比較して、未知語の大幅な減少が考えられる。よって、未知語数の調査を行なった。結果を表7に示す。

表7 未知語数の比較

	単文		重文複文	
	総数	ユニーク数	総数	ユニーク数
ベースライン	3914	3480	4020	3576
ATLASのみ	379	346	518	443
提案手法1	310	297	370	348
翻訳の王様のみ	665	588	770	642
提案手法2	446	429	492	461

結果より、ベースラインと比較して、提案手法1と提案手法2の両方において、未知語の大幅な減少が確認できた。また、ルールベース翻訳のみと比較しても、提案手法1と提案手法2の両方において、未知語の減少が確認できた。よって、未知語の減少が翻訳精度の向上に影響していると考えられる。

しかしながら、ルールベース翻訳のみと提案手法の未知語数では大きな差はない。それにも関わらず、提案手法では、ルールベース翻訳のみと比較して、BLEUスコアが大きく向上していることから、BLEUスコアに対する未知語の影響は小さいものであると考えられる。

7.3 distortion-limit の影響

英' 統計翻訳での、文法構造変換の必要性を調査するために、提案手法1(ATLAS+ 統計翻訳)において、単語移動距離制限である“distortion-limit”の値を変化させて実験を行なった。結果を表8に示す。

表8 ベースラインと提案手法の評価の比較

	単文		重文複文	
	BLEU	METEOR	BLEU	METEOR
ベースライン	0.1397	0.3845	0.1175	0.3531
ATLASのみ	0.1552	0.4216	0.0942	0.3623
提案手法				
distortion=1	0.1878	0.4417	0.1389	0.3854
distortion=3	0.1870	0.4378	0.1356	0.3846
distortion=6	0.1877	0.4418	0.1364	0.3806
distortion=12	0.1871	0.4404	0.1398	0.3793
distortion=-1	0.1870	0.4404	0.1418	0.3865

結果より、単文では“distortion-limit”の値を変化させても、翻訳精度にあまり影響しないことが確認された。しかしながら、重文複文では“distortion-limit”の値が大きいために、翻訳精度が高くなった。これは、単文ではルールベース翻訳による文法構造変換が高精度で行なっていたため、統計翻訳において、文法構造変換の必要性が低く、“distortion-limit”の影響が小さかったことが原因であると考えている。また、重文複文ではルールベース翻訳による文法構造変換が低精度で行なわれていたため、統計翻訳において、文法構造変換の必要性が高かったことが原因であると考えている。

このことから、前処理による文法構造変換を高精度に行なう必要があると考えている。また、対象文に応じ

て“distortion-limit”を変化させる必要もあると考えている。

8 おわりに

本研究では、文法構造が類似しているときには翻訳精度が高くなる統計翻訳の長所を用いるために、統計翻訳に対して、前処理としてルールベース翻訳を組み合わせた。結果として、単文において、ベースラインと比較して、BLEUスコアが4.80%向上し、METEORスコアが5.73%向上した。また重文複文において、ベースラインと比較して、BLEUスコアが1.89%向上し、METEORスコアが2.75%向上し、本研究の有効性が確認できた。

しかしながら、重文複文において、ルールベース翻訳(ATLAS)のみの出力文と提案手法の出力文に対して、対比較評価したときに、提案手法の方が悪い結果となった。これは前処理において、英語の文法構造に類似させることができなかつたことが原因であると考えている。よって、今後の研究では、重文複文に対しても、英語の文法構造に類似させることができる前処理の方法を調査することを考えている。

参考文献

- [1] H.Schwenk, Marta R.Costa-jussa, and Jose A.R.Fonollosa, “Continuous space language models for the IWSLT 2006 Task”, in IWSLT, 2006, pages.166-173
- [2] L.Dugast, J.Senellart, and P.Koehn, “Statistical postediting on SYSTRAN’s rule-based translation system”, in Second Workshop on SMT, 2007, pages.179-182
- [3] M.Simard, N.Ueffing, P.Isabelle, and R.Kuhn, “Rule-based translation with statistical phrase-based post-editing”, in Second Workshop on SMT, 2007, pages.203-206
- [4] SRILM(The SRI Language Modeling Toolkit) : srilm.tgz
<http://www.speech.sri.com/projects/srilm/>
- [5] Moses : moses.2007-05-29.tgz
<http://www.statmt.org/moses/>
- [6] Franz Josef Och, “Minimum error rate training for statistical machine translation”, Proceedings of the ACL, 2003.
- [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] ChaSen, <http://chasen-legacy.sourceforge.jp/>
- [9] 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
- [10] NIST Open Machine Translation (OpenMT) Evaluation
<http://www.itl.nist.gov/iad/mig/tests/mt/>
- [11] The METEOR Automatic Machine Translation Evaluation System
<http://www.cs.cmu.edu/~alavie/METEOR/>