

2022年度(令和4年度) 卒業論文

翻訳における文パターンの利用

指導教員

村上仁一  
村田真樹

鳥取大学工学部 電気情報系学科

自然言語処理研究室

B19T2092K 深谷 諒

## 概要

近年，機械翻訳の1つに Neural Machine Translation(NMT)がある．また，機械翻訳には様々な問題があると言われている．その中の一つに，要素合成法の問題というものがある．文全体を考えず，単語ごとに翻訳し，意味が通らなくなる，という問題だ．本研究で用いる NMT にも同様の問題があるとされている．ここで，要素合成法の問題を解決し翻訳精度を向上させるには，文全体の意味を考慮して翻訳することが必要だと考えた．

そこで本研究では，文全体を考えて翻訳するために NMT の学習データに文パターンを追加することを考えた．文パターンは単文対訳学習文から抽出した，単文で変数が一つのものを用いる．実験の結果，文パターンを追加したほうが，自動評価，人手評価，共に精度がわずかに向上した．

# 目次

第1章	はじめに	1
第2章	NMTについて	2
2.1	NMTとは	2
2.2	NMTの学習と翻訳	2
2.3	Encoder-Decoderモデル	3
2.4	Attentionモデル	3
第3章	要素合成法	4
3.1	要素合成法の問題	4
3.2	要素合成法を解決するために	5
第4章	文パターン	6
第5章	提案手法	9
第6章	実験	10
6.1	実験手順	10
6.2	実験条件	10
6.3	実験データ	11
第7章	実験結果	12
7.1	自動評価結果	12
7.2	人手評価結果	13
7.3	出力結果の例	14
7.3.1	提案手法○の例	14
7.3.2	ベースライン○の例1	18
7.3.3	互角○の例	20

7.4	互角×の例 . . . . .	21
<b>第8章</b>	<b>考察</b>	<b>23</b>
8.1	複数回の実験 . . . . .	23
8.2	要素合成法の問題について . . . . .	26
8.3	影響する文パターンについて . . . . .	26
<b>第9章</b>	<b>おわりに</b>	<b>27</b>

# 目 次

2.1	NMT の仕組み . . . . .	2
4.1	対訳単語の作り方 . . . . .	6
4.2	単語に基づいた文パターンの作り方 . . . . .	7
4.3	対訳句の作り方 . . . . .	7
4.4	対訳句に基づく文パターンの作り方 . . . . .	8
5.1	提案手法の流れ . . . . .	9

# 表 目 次

3.1	間違った出力の例 1 . . . . .	4
3.2	間違った出力の例 2 . . . . .	4
3.3	間違った出力の例 3 . . . . .	4
3.4	文パターンの例 . . . . .	5
4.1	文パターンの例 . . . . .	6
6.1	実験データ . . . . .	11
6.2	対訳学習文の例 . . . . .	11
7.1	自動評価の比較 1 回目 . . . . .	12
7.2	自動評価の比較 2 回目 . . . . .	12
7.3	自動評価の比較 3 回目 . . . . .	12
7.4	対比較評価の結果 . . . . .	13
7.5	提案手法○の例 1 . . . . .	14
7.6	提案手法○の例 2 . . . . .	15
7.7	提案手法○の例 3 . . . . .	15
7.8	提案手法○の例 4 . . . . .	16
7.9	提案手法○の例 5 . . . . .	16
7.10	提案手法○の例 6 . . . . .	17
7.11	ベースライン○例 1 . . . . .	18
7.12	ベースライン○の例 2 . . . . .	18
7.13	ベースライン○の例 3 . . . . .	19
7.14	ベースライン○の例 4 . . . . .	19
7.15	互角○の例 1 . . . . .	20
7.16	互角○の例 2 . . . . .	20
7.17	互角×の例 1 . . . . .	21

7.18	互角×の例2 . . . . .	21
7.19	互角×の例3 . . . . .	22
8.1	3回分の出力結果の例1 . . . . .	23
8.2	3回分の出力結果の例2 . . . . .	24
8.3	3回分の出力結果の例3 . . . . .	24
8.4	人手評価が割れる例1 . . . . .	25
8.5	人手評価が割れる例2 . . . . .	25
8.6	人手評価が割れる例3 . . . . .	25

# 第1章 はじめに

近年、機械翻訳の1つに Neural Machine Translation(NMT)がある。また、機械翻訳には様々な問題があると言われている。その中の一つに、入力文を単語ごとに翻訳をしまい、文全体としては意味が通じなくなったり、おかしい文が出力される、というものがある。それを要素合成法の問題という。本研究で用いる NMT にも同様の問題がある。ここで、要素合成法の問題を軽減し翻訳精度を向上させるには、文全体の意味を考慮して翻訳することが必要だと考えた。

そこで本研究では、要素合成法の軽減方法の一つとして、NMT の学習において対訳学習文に、対訳学習文から抽出した文パターンを追加することにより、翻訳精度の向上を試みる。文パターンを用いる理由は、文パターンを追加することによって字面が近いものが優先され、変数の部分に任意の出力をすることができるからである。結果として文パターンを追加することによって、単語ごとの部分的な翻訳ではなく、文全体を考えた翻訳を強化することができ、翻訳精度が向上すると考えられる。また、本研究では、単文対訳学習文から抽出した変数が1つの単文文パターンを用いる。

本論文の構成は以下の通りである。第2章では、NMT について述べる。第3章では、要素合成法について述べる。第4章では、文パターンについて述べる。第5章では、提案手法について述べる。第6章では、実験について述べる。第7章では、実験結果について述べる。第8章では、考察について述べる。第9章では、本実験の簡単なまとめについて述べる。

## 第2章 NMTについて

### 2.1 NMTとは

NMT[1]はニューラルネットワークを用いた機械翻訳である。NMTの学習には対訳学習文を用いる。NMTの手法には、Encoder-DecoderモデルとAttentionモデルがある。Encoder-Decoderモデルは入力を符号化し、この符号化された情報を復号化するモデルである。AttentionモデルはEncoder-Decoderモデルの出力を生成するときに、出力と入力の参照を行う手法である。

### 2.2 NMTの学習と翻訳

まず、学習の過程では、対訳学習文から日本語文と英語文を学習し、モデルを作成する。次に、翻訳の過程では、学習の過程でNMTが作成したモデルを用いて、入力文を翻訳、出力する。NMTの学習と翻訳の仕組みを以下の図2.1に簡単に示す。

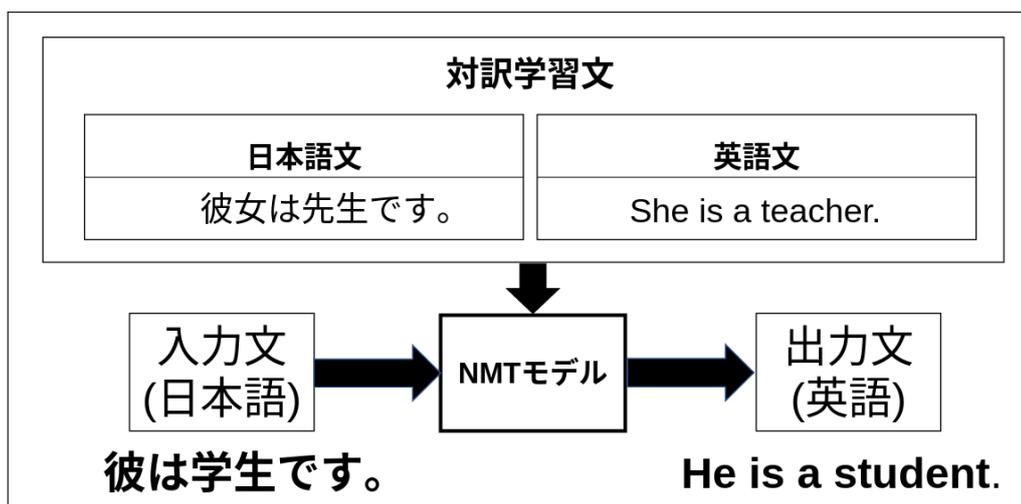


図 2.1: NMT の仕組み

## 2.3 Encoder-Decoder モデル

Encoder-Decoder モデルとは、2つ RNN により構成されるニューラル機械翻訳のモデルである。Encoder では入力データを処理して符号化し、Decoder ではこの符号化された情報を復号化する。出力も時系列となるために、Decoder は自身の出力を次に入力として受け取って処理する。

## 2.4 Attention モデル

Encoder-Decoder モデルの問題点として、長文を翻訳した場合に、翻訳精度が下がるどころが挙げられる。そこで Attention モデルは、この問題を解決するために提案されたモデルである。Attention モデルの Encoder は、入力を前後から RNN に渡す。Encoder は入力を前後から読み込んだものを組み合わせてそれぞれの単語を符号化する。Decoder は、モデルが出力文中の単語を生成するときに、生成した単語と対訳学習文から次の単語を予測して出力する。

## 第3章 要素合成法

### 3.1 要素合成法の問題

NMT を用いて文を翻訳する際、単語ごとに翻訳すると間違っただ訳になる場合がある。例を以下の表 3.1 ～表 3.3 に示す。

表 3.1: 間違っただ訳の例 1

入力文	彼は我を通した。
参照文	He had his own way.
出力文	He passed me.

この出力文は、”通した”を”passed”、”我”を”me”と単語ごとに訳しているため間違いが起きている。

表 3.2: 間違っただ訳の例 2

入力文	彼女は道草を食った。
参照文	She wasted time in the park.
出力文	She ate the road grass in the park.

この出力文は、”道草”を”road grass”、”食った”を”ate”と単語ごとに訳しているため間違いが起きている。

表 3.3: 間違っただ訳の例 3

入力文	信号が青より赤に変わった。
参照文	The signal changed from green to red .
出力文	The signal changed green than red .

この出力文は、”より”を”than”と単語ごとに訳しているため間違いが起きている。

要素合成法とは，上記のように入力文を単語ごとに翻訳し，その結果を組み合わせて出力を行う手法である．結果として，一般的な意味と異なった翻訳になることを要素合成法の問題 [2] という．

### 3.2 要素合成法を解決するために

要素合成法を解決するために，対訳学習文に文パターンを追加して，学習するという方法がある．文パターンを追加することにより，文全体を考えた翻訳が強化されると考える．上記の表 3.2 の例ならば，以下に示すような文パターンを追加する．

表 3.4: 文パターンの例

日本語文パターン	英語文パターン
昨日、X は道草を食った。	X wasted time in the park yesterday.

このような似た文パターンがあれば，文パターンに近い出力が優先され，間違った出力が出にくくなると考える．それにより，要素合成法の解決に近づき，翻訳精度が向上すると考える．

## 第4章 文パターン

文パターンとは、変数を用いてパターン化された文である。本研究は、単文対訳学習文から自動抽出した変数が1つの単文文パターンを用いる(森本 [3] 参照)。文パターンの精度は約 80 %であった。また、文パターンの例を表 4.1 に示す。

表 4.1: 文パターンの例

日本語文パターン	英語文パターン
来週 X に行きます。	I'm going to X next week .
X は先生です	X is a teacher.
連夜のマージャンで X は体をこわした。	Playing mah-jongg nightly ruined N health .
彼はその X に夢中になっていた。	He was engrossed in the X .
ここにあるなかで、一番大きな X はこれです。	This is the largest X that is being kept here .
X はこの点で有害な影響を及ぼす。	X exercises a pernicious

手順 1 GIZA++[4] を用いて対訳学習文から翻訳確率を得る。そして任意の対訳学習文と翻訳確率から、対訳単語を作成する。図 4.1 に対訳単語の作り方を示す。

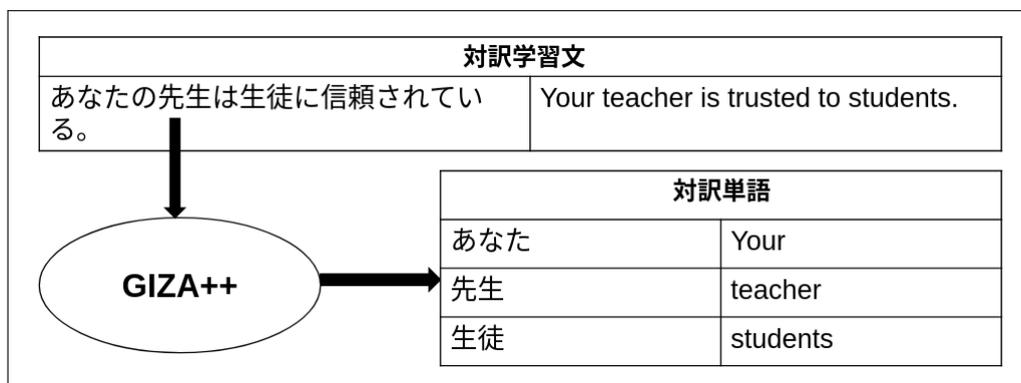


図 4.1: 対訳単語の作り方

手順 2 対訳単語と対訳学習文を用い、単語に基づいた文パターンを作成する。図 4.2 に単語に基づいた文パターンの作り方を示す

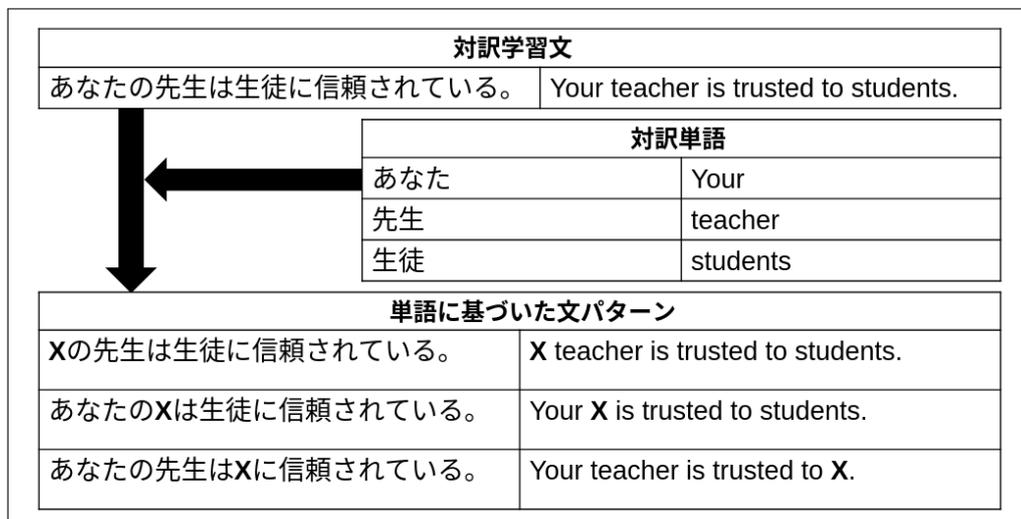


図 4.2: 単語に基づいた文パターンの作り方

手順 3 対訳学習文と単語に基づいた文パターンを照合し、適合した場合、単語に基づいた文パターンの変数部に対応した対訳句を抽出する。

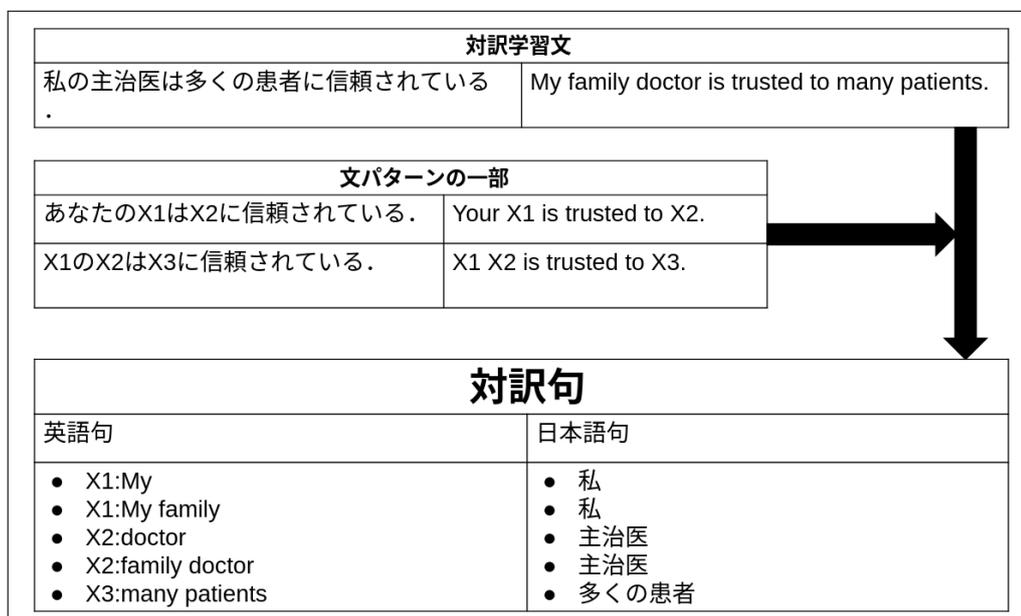


図 4.3: 対訳句の作り方

手順 4 対訳学習文と対訳句を照合し、適合した場合、該当箇所を変数化し、対訳句に基

づいた文パターンを生成する。

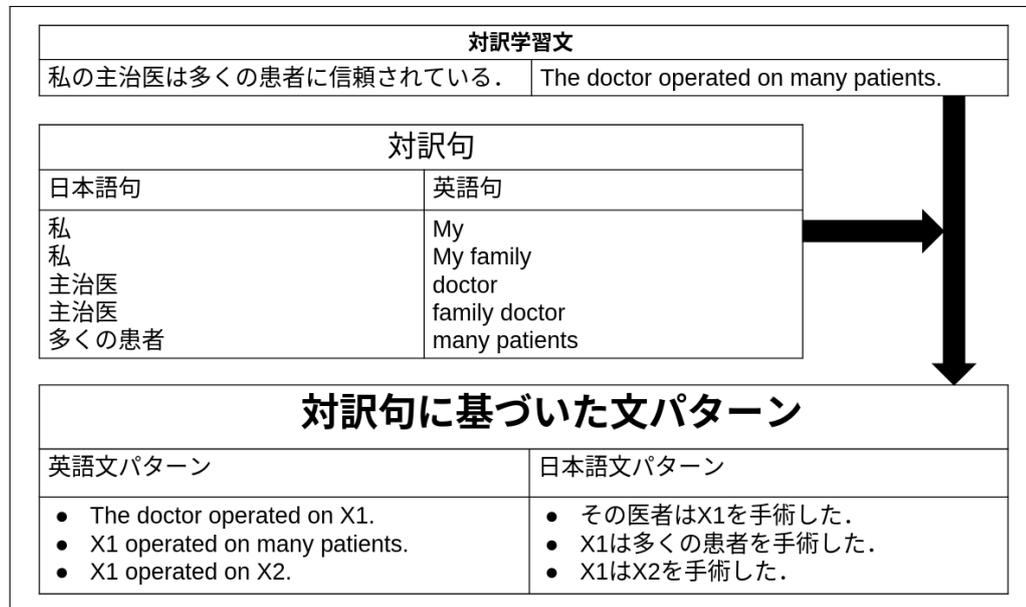


図 4.4: 対訳句に基づく文パターンの作り方

手順 5 手順 4 で生成した対訳句に基づいた文パターンを用い、手順 3,4 をもう一度繰り返す。最終的に、単語に基づいた文パターン、対訳句に基づいた文パターン 1、対訳句に基づいた文パターン 2 の 3 種類のパターンができる。本研究では、3 種類のパターンのうち、変数が 1 つのものを用いる。

## 第5章 提案手法

単語ごとの翻訳ではなく，文全体を考えて文パターンを追加することによって，要素合成法の問題が解決され，翻訳精度が向上すると考える．文パターンを追加する流れと手順を以下に示す．

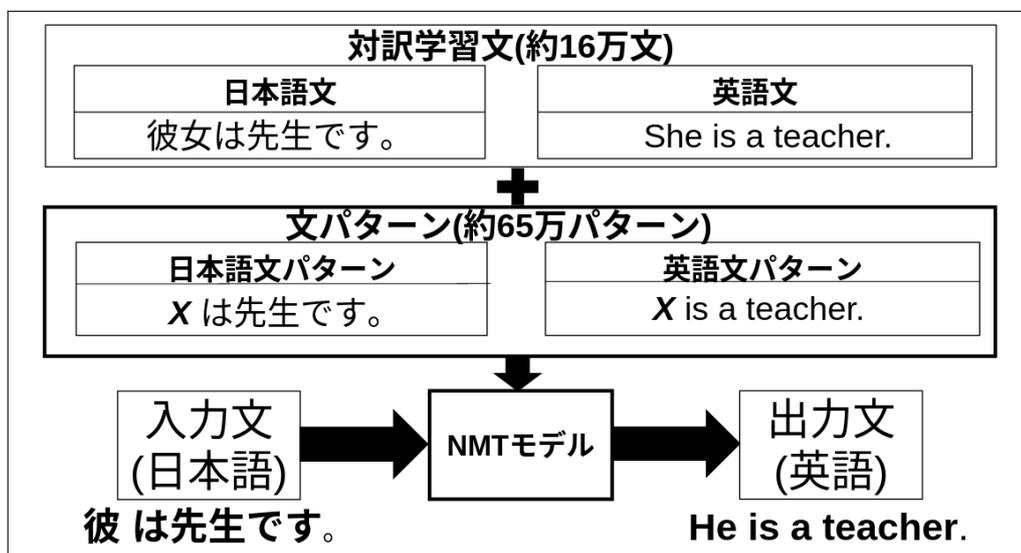


図 5.1: 提案手法の流れ

手順 1 対訳学習文に文パターンを追加して，新たな学習データを作る．

手順 2 手順 1 のデータを NMT で学習する．

手順 3 手順 2 で学習した NMT モデル を使って翻訳を行う．

## 第6章 実験

はじめに、本研究で扱う対訳学習文，文パターン，テスト文はいずれも単文である．

### 6.1 実験手順

手順 1 NMT に単文対訳学習文約 16 万文を学習する．

手順 2 1 で学習した NMT モデルを用い，テスト文を翻訳しベースラインとする．

手順 3 対訳学習文に変数が 1 つの文パターン 65 万パターンを追加し，新たな学習データを作る．

手順 4 3 の学習データを使って NMT で学習する．

手順 5 4 で学習させた NMT モデルを用い，テスト文を翻訳し提案手法とする．

手順 6 提案手法とベースラインを比較して自動評価 (BLEU, METEOR, RIBES, TER) を行う．

手順 7 出力文からランダムに選んだ 100 文について提案手法とベースラインを比較して人手評価を行う．

### 6.2 実験条件

本研究では，OpenNMT[5] を用いて日英ニューラル機械翻訳を行う．OpenNMT とは 2016 年 12 月にハーバード大学 NLP グループと SYSTRAN によって開始され，その後，いくつかの研究および産業用途で利用されている．現在，SYSTRAN と Ubiqus によってメンテナンスされている．

## 6.3 実験データ

実験に用いるデータの内訳を表 6.1 に示す。

また対訳学習文の例を表 6.2 に示す。

表 6.1: 実験データ

単文対訳学習文	163,188 文
単文テスト文	16,328 文
単文文パターン	647,488 パターン

表 6.2: 対訳学習文の例

日本語	英語
あなたの発音は不明瞭でよく聞きとれなかった。	I could not hear you well because your pronunciation is not clear .
きょうから向こう 1 か月間、図書館は閉館になる。	The library is closed for a month from today .
その証拠によって犯罪の動機が十分に証明された。	The evidence sufficiently established the motive for the crime .
雲の陰から丸い月が現われた。	The disk of the moon appeared from behind a cloud .

## 第7章 実験結果

### 7.1 自動評価結果

テスト文約1万6千文を入力文として翻訳実験を行い，出力文に対して自動評価を行った。表7.1～7.3に，自動評価 (BLEU, METEOR, RIBES, TER) の結果を示す。

表 7.1: 自動評価の比較1回目

評価方法	BLEU	METEOR	RIBES	TER
提案手法	0.1910	0.4614	0.7706	0.6185
ベースライン	0.1837	0.4539	0.7659	0.6254

表 7.2: 自動評価の比較2回目

評価方法	BLEU	METEOR	RIBES	TER
提案手法	0.1913	0.4630	0.7711	0.6142
ベースライン	0.1878	0.4604	0.7684	0.6180

表 7.3: 自動評価の比較3回目

評価方法	BLEU	METEOR	RIBES	TER
提案手法	0.1916	0.4582	0.7701	0.6228
ベースライン	0.1831	0.4544	0.7651	0.6245

表7.1～7.3の結果より，自動評価ではベースラインと比較して提案手法の精度が良いことがわかった。

## 7.2 人手評価結果

出力文からランダムに選んだ 100 文を用い，人手による対比較評価を行った．評価結果を表 5 に示す．表 7.4 の結果は，4 名が同じ 100 文を評価し，その結果をまとめた．以下の表 7.4 では 4 名の被験者をそれぞれ被験者 A，被験者 B，被験者 C，被験者 D とする．また，評価の説明を以下に示す．

### 提案手法○

提案手法とベースラインの比較した際，ベースラインの出力文の方が入力文の意味に近い．

### ベースライン○

提案手法とベースラインを比較した際，対訳句追加時の出力文の方が入力文の意味に近い．

### 互角

提案手法とベースラインを比較した際，両手法の出力文の評価が互角．

表 7.4: 対比較評価の結果

	被験者 A	被験者 B	被験者 C	被験者 D	total
提案手法○	27	27	18	27	99
ベースライン○	26	25	16	22	89
互角	47	48	66	51	212

表 7.4 の結果より，人手評価では提案手法がわずかに精度がいいことがわかった．

## 7.3 出力結果の例

ベースラインと提案手法との対比較評価において、提案手法○とした例を表 7.5～7.10 に示す。ベースライン○とした例を表 7.11～7.14 に示す。互角の例を表 7.15～7.19 に示す。また、提案手法の出力結果に影響したであろう文パターンをそれぞれ示す。

### 7.3.1 提案手法○の例

表 7.5: 提案手法○の例 1

テスト文	彼の論文が私の発見の手掛りとなりました。
提案手法	His article was a harbinger of my discovery .
ベースライン	His article was the understudy to my discovery .
参照文	I found a clue for my discovery in his paper .
日本語パターン	この X が、この新しい分野における発展の幕開けとなりました。
英語パターン	This X was a harbinger of the development in this new field .

表 7.5 の例において、”手掛かり”に対して”harbinger”という比較的近い訳が得られている提案手法を○とした。

表 7.6: 提案手法○の例 2

テスト文	彼はそこで戦闘中に殺された。
提案手法	He was killed there in the battle .
ベースライン	He was killed there .
参照文	He was killed in action there .
日本語パターン	戦闘で X の血が流された。
英語パターン	X blood was spilled in the battle .

表 7.6 の例において, ”戦闘中”に対して”in the battle”という比較的近い訳が得られている提案手法を○とした。

表 7.7: 提案手法○の例 3

テスト文	この証拠はその主張の正当性を裏づけていない。
提案手法	This evidence doesn't support the justice of the claim .
ベースライン	This evidence justifies the rationality of the claim .
参照文	The evidence does not bear out the claim .
日本語パターン	北部 X 以外では裏付けとなる遺跡、遺物はほとんどなかった。
英語パターン	However , only in the northern X has much been excavated to substantially support the descriptions .

表 7.7 の例において, ”裏付ける”に対して”support”という比較的近い訳が得られている提案手法を○とした。

表 7.8: 提案手法○の例 4

テスト文	お隣から、見事なりんごを頂いた。
提案手法	Trains are crowded all day , but especially around eight in the morning .
ベースライン	She received a fine apple following the door .
参照文	We received a fine apple from the next door .
日本語パターン	X 伝いに 隣家 へ 行った。
英語パターン	I went to the next door from garden to X.

表 7.8 の例において, ”お隣” に対して ”the next door” という比較的近い訳が得られている提案手法を○とした。

表 7.9: 提案手法○の例 5

テスト文	いくつかの前例がこの決定を正当化している。
提案手法	Several precedent justify this decision .
ベースライン	Several precedent have justice in this decision .
参照文	Several precedents justify this decision .
日本語パターン	目的は必ずしも X を正当化しない。
英語パターン	The end does not necessarily justify the X.

表 7.9 の例において, ”正当化” に対して ”justify” という比較的近い訳が得られている提案手法を○とした。

表 7.10: 提案手法○の例 6

テスト文	このごろでは読書の趣味もなくなった。
提案手法	Nowadays I have lost interest in reading .
ベースライン	Recently reading too much reading is the hobby of reading .
参照文	I have now lost interest in reading .
日本語パターン	Xには趣味がなくなった。
英語パターン	I have lost interest in X.

表 7.10 の例において, ”読書の趣味もなくなった” に対して ”lost interest in reading” という比較的近い訳が得られている提案手法を○とした.

### 7.3.2 ベースライン○の例1

表 7.11: ベースライン○例1

テスト文	ぶらんこが揺れている。
提案手法	The swing is flickering .
ベースライン	The swing is swinging .
参照文	The swing is swinging .
日本語パターン	ろうそく X 炎が揺れている。
英語パターン	X candle flame is flickering .

表 7.11 の例において, "flickering" より "swinging" のほうが適していると考え, ベースライン○とした。

表 7.12: ベースライン○の例2

テスト文	韓国にもまた、一層の努力を求めたい。
提案手法	We urge South Korea to make more effort to pursue its efforts .
ベースライン	We also urge South Korea to make greater efforts .
参照文	We exhort South Korea to make stronger efforts as well .
日本語パターン	患者団体や X などの一層の努力がのぞまれる。
英語パターン	Patient organizations and local X can play a crucial role in this respect .

表 7.12 の例において, 提案手法の後半が不要だと感じた。よって, ベースライン○とした。

表 7.13: ベースライン○の例 3

テスト文	この問題は近似的にしか解けません。
提案手法	This problem can only be solved amicably .
ベースライン	The problem can only be solved in approximate terms .
参照文	Only an approximate solution can be found for this problem .
日本語パターン	労使間の X が円満に解決した。
英語パターン	The X between management and labor was settled amicably .

表 7.13 の例において，”近似的”に対して”approximate”という比較的近い訳が得られているベースラインを○とした。

表 7.14: ベースライン○の例 4

テスト文	きょうは頭がなんとなくぼうっとしている。
提案手法	My head is fuzzy today .
ベースライン	My head is somewhat fuzzy today .
参照文	I feel somehow dazed today .
日本語パターン	睡眠不足で X がぼうっとしている。
英語パターン	My X is fuzzy from lack of sleep .

表 7.14 の例において，”なんとなく”に対して”somewhat”という比較的近い訳が得られているベースラインを○とした。

### 7.3.3 互角○の例

表 7.15: 互角○の例 1

テスト文	豊富な語彙をもっている。
提案手法	He has a large vocabulary .
ベースライン	He possesses a wide vocabulary .
参照文	He has a copious vocabulary .
日本語パターン	彼は X 単語をたくさん知っている。
英語パターン	He has a large vocabulary of X .

表 7.15 の例において、どちらも正しい訳と判断して、互角○とした。

表 7.16: 互角○の例 2

テスト文	この体育館にはいろいろな設備が整っている。
提案手法	This gymnasium is equipped with various facilities .
ベースライン	This gymnasium is equipped with all kinds of facilities .
参照文	This gymnasium is equipped with many facilities .
日本語パターン	あの X は設備が整っている。
英語パターン	That X is well equipped .

表 7.16 の例において、どちらも正しい訳と判断して、互角○とした。

## 7.4 互角×の例

表 7.17: 互角×の例 1

テスト文	英語をやり直さねばならない。
提案手法	We must overrate the English .
ベースライン	We must converse English .
参照文	I must brush up my English .
日本語パターン	X 彼らを買いかぶる。
英語パターン	We are X to overrate their importance .

表 7.17 の例において、どちらも”やり直す”の意味を翻訳できていないと判断して、互角×とした。

表 7.18: 互角×の例 2

テスト文	この男性の知人のパキスタン人男性も九月中旬、都内の部品製造会社から突然説明もなく解雇された。
提案手法	Men of men and acquaintances of the man were dismissed from the Tokyo parts in September in the middle of the city .
ベースライン	In the late September male of the male male who male men in the city were dismissed in the middle of September .
参照文	A 36-year-old Pakistani man , who is friends with the spice-shop owner , said he was suddenly fired from his job at a machinery-parts manufacturer in Tokyo in mid-September .

表 7.18 の例において、どちらも間違っただ訳と判断して、互角×とした。この例は類似パターンがなかった。

表 7.19: 互角×の例 3

テスト文	これと同時にアジア太平洋経済協力会議諸国は、太平洋自由貿易圏への長い道程に乗り出した。
提案手法	Men of men and acquaintances of the man were dismissed from the Tokyo parts in September in the middle of the city .
ベースライン	At the same time , the Asia-Pacific Economic Cooperation forum in Asia has started a long suspension in the Asia-Pacific market .
参照文	At the same time , the APEC countries launched out on the long journey to a Pacific free-trade zone .

表 7.19 の例において、どちらも間違っただ訳と判断して、互角×とした。この例は類似パターンがなかった。

## 第8章 考察

### 8.1 複数回の実験

出力結果がどれほど変わるのか調べるために、ベースラインと提案手法をそれぞれ3回実験を行った。3回分の出力結果の例を以下の表8.1～表8.3に示す。

表 8.1: 3回分の出力結果の例 1

テスト文	捕虜 たち が 逃亡 を 図った。
提案手法 1 回目	The prisoners attempted to escape .
提案手法 2 回目	The prisoners attempted to escape .
提案手法 3 回目	The captives tried to escape the escape .
ベースライン 1 回目	The prisoners of war prevented the escape .
ベースライン 2 回目	The prisoners tried to escape the escape .
ベースライン 3 回目	The prisoners fled to escape .
参照文	The captives attempted to make their escape .

表 8.2: 3 回分の出力結果の例 2

テスト文	彼はそこで戦闘中に殺された。 。
提案手法 1 回目	He was killed there .
提案手法 2 回目	He was killed in death there .
提案手法 3 回目	He was killed there by the end .
ベースライン 1 回目	He was killed there in the battle .
ベースライン 2 回目	He was killed in battle there .
ベースライン 3 回目	He was killed there during the fight .
参照文	He was killed in action there .

表 8.3: 3 回分の出力結果の例 3

テスト文	子供たちの旅行の話は食べ物 に集中した。
提案手法 1 回目	The conversation of the children concentrated on food .
提案手法 2 回目	The children's travel covered food .
提案手法 3 回目	The neighbors' accounts were cen- tered on food .
ベースライン 1 回目	The talk about the children's trip centered on food .
ベースライン 2 回目	The children's journey focused on food .
ベースライン 3 回目	The children of our children's travel concentrated on food .
参照文	The children's description of the trip centered on the food .

表のように、複数回実験を行っても、提案手法のほうが精度が良い傾向は変わらなかった。また、自動評価結果でも、表 7.1~7.3 に示すように、すべての結果で提案手法のほうが精度が良い、という結果になった。だが、一方で以下の表 8.4~表 8.6 のような例もある。

表 8.4: 人手評価が割れる例 1

テスト文	我々の大学は長い輝かしい歴史を誇っている。	評価
提案手法 1 回目	Our college prides itself on long historical history .	
ベースライン 1 回目	Our college has boasted long history .	○
参照文	Our university can boast a long and glorious history .	

表 8.5: 人手評価が割れる例 2

テスト文	我々の大学は長い輝かしい歴史を誇っている。	評価
提案手法 2 回目	Our college boasts a long history .	○
ベースライン 2 回目	Our university has a long history of history .	
参照文	Our university can boast a long and glorious history .	

表 8.6: 人手評価が割れる例 3

テスト文	我々の大学は長い輝かしい歴史を誇っている。	評価
提案手法 3 回目	Our college boasts a long glorious past .	○
ベースライン 3 回目	Our college boasts a long historical history .	
参照文	Our university can boast a long and glorious history .	

上記は、1 回目はベースラインのほうが良かったが、2,3 回目は提案手法のほうが精度が良い例だ。原因はやはり乱数だと考えられる。自動評価では提案手法のほうが数値が高いので問題はないと考える。ここからわかることは、NMT は乱数の問題により毎回出力結果が微妙に異なるが、大体の傾向は変わらないということだ。

## 8.2 要素合成法の問題について

本実験では、要素合成法の問題の解決するために文パターンを追加して実験を行った。両評価方法において提案手法がベースラインより精度が良いという結果になった。だが、要素合成法の問題を解決するために追加した文パターンにより、翻訳結果がおかしくなったと思われるものも存在した。今回は、それを差し引いても自動評価、人手評価共に精度が向上したので、私は少しは解決したと考える。だが勿論、今後の課題は存在するので、引き続き実験を行っていく予定である。

## 8.3 影響する文パターンについて

実験の出力結果を見てみると、提案手法のほうが精度が良いときは提案手法の出力結果に近い文パターンがデータに含まれている傾向にあることが分かった。今後の課題としてその数を調べていきたい。また、類似文検索を利用して出力結果に近い文パターンを調べる予定だ。

## 第9章 おわりに

本研究では，要素合成法の問題を解決するために，NMTの学習データが対訳学習文約16万のみだったところに，対訳学習文から抽出した変数が1つの文パターン約65万パターンを追加した．実験結果より，提案手法のBLEU値が約0.7ポイント，他の自動評価でもベースラインを上回った．また，人手評価では提案手法がベースラインの精度をわずかに上回った．結果的に，文パターンを追加すると僅かではあるが文全体を考えた翻訳になる，ということがわかった．

また今回の実験では，助詞を変数にしてしまっていて，文パターンとしておかしいものなども存在した．そのせいで精度が下がった可能性も考えられる．よって今後の課題として，文パターンとしておかしいものを排除するとどれくらい精度が上がるのか調査がある．また変数が2つ以上の文パターンを増やすと精度が上がるかの調査も行う予定だ．

## 謝辞

最後に、一年間に渡り、本研究のご指導をいただきました鳥取大学工学部知能情報工学科自然言語処理研究室の村上仁一准教授、村田真樹教授に深く感謝すると共に、厚く御礼申し上げます。そして、日常の議論を通じて多くの知識や示唆を頂いた同研究室の皆様に深謝いたします。

## 参考文献

- [1] Minh-Thang Luong Hieu Pham Christopher D. Manning:Effective Approaches to Attention-based Neural Machine Translation
- [2] 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複文の日英文型パターン化
- [3] 森本 世人 村上 仁一 : 類似度を利用した変換テーブルの精度向上
- [4] Franz Josef Och, Hermann Ney: “ A System- atic Comparison of Various Statistical Alignment Models ” , Computational Linguistics, pp.19-51, 2003.
- [5] OpenNMT : <https://pypi.org/project/OpenNMT-py/1.2.0/>
- [6] 江木孝史 村上仁一 徳久雅人 : 句に基づく対訳文パターンの自動作成と統計的手法を用いた英日パターン翻訳