

概要

近年、機械翻訳の方式として等価的類推思考の原理に基づく機械翻訳方式が提案されている。この方式の実現に向けて、日本語の重文・複文を対象とした文型パターンを大量に蓄積した文型パターン辞書の構築が進められている。現在、入力文と文型パターンを照合し、入力文に適合する文型パターンを抽出する文型パターンパーサが試作されている。この文型パターンパーサを使用すると、複数の文型パターンが入力文に適合する。入力文の意味を考えると適合した文型パターンの中には英文生成に不向きなパターンも含まれる。よって、入力文の英訳を出力するために適合パターンの選択が必要となる。

本研究では、入力文の英文生成が可能な適合パターンの選択手法を検討する。具体的には、入力文に適合した文型パターンを多変量解析によって分析し、評価関数を求める。評価関数を使用して適合パターンの得点を求め、英文生成に使用する適合パターンの選択を行う。適合パターンの選択実験を行った結果、クローズドテストでは72%、オープンテストでは83%の入力文において、正しい適合パターンの選択を行うことができた。また、正しい適合パターンを持つ全ての入力文において、上位8位までに正しい適合パターンが存在した。本研究で得られた評価関数は、入力文に多数の文型パターンが適合する場合において十分な適合パターン選択精度を得ることができなかった。今後は、多数の文型パターンが適合する入力文に対応する評価関数を求める必要があると考えられる。

目次

1	はじめに	1
2	文型パターン	2
2.1	文型パターンの記述形式	2
2.2	文型パターン照合	3
2.3	文型パターンを利用した英文生成	4
3	多変量解析による評価関数の作成	5
3.1	本研究の目的	5
3.2	意味属性大系	5
3.2.1	一般名詞意味属性大系	5
3.2.2	用言意味属性大系	6
3.3	評価関数	6
3.4	評価関数作成の実験条件	9
3.5	評価関数作成結果	9
3.6	各評価における作成訳の例	10
4	適合パターン選択実験	11
4.1	選択関数の評価	11
4.2	入力文の比較	11
4.3	適合パターン選択実験結果	12
5	考察	14
5.1	適合パターン選択失敗の原因	14
5.2	適合パターンの数と評価関数の値の関係	16
5.3	適合パターンの最大値と評価	18
5.4	評価関数の説明力	20
5.5	パターン元字面適合率	21
5.6	様々な評価関数	22
5.6.1	寄与率の比較	23
5.6.2	各条件下における適合パターン選択精度	23
5.6.3	各条件下における適合パターンの最大値と評価の関係	27
6	まとめ	31

目 次

1	文型パターン照合	3
2	一般名詞意味属性大系の一部	5
3	用言意味属性大系の一部	6
4	適合パターン数と評価関数の値の関係 (クローズド)	16
5	適合パターン数と評価関数の値の関係 (オープン)	16
6	評価関数の値と累積パターン数の関係 (クローズド)	17
7	評価関数の値と累積パターン数の関係 (オープン)	17
8	適合パターンの最大値と評価の関係 (クローズド)	18
9	適合パターンの最大値と評価の関係 (オープン)	18
10	適合パターンの最大値と評価の関係 (意味属性, クローズド)	28
11	適合パターンの最大値と評価の関係 (意味属性, オープン)	28
12	適合パターンの最大値と評価の関係 (P 元, クローズド)	29
13	適合パターンの最大値と評価の関係 (P 元, オープン)	29
14	適合パターンの最大値と評価の関係 (P 元+意味属性, クローズド)	30
15	適合パターンの最大値と評価の関係 (P 元+意味属性, オープン)	30

表 目 次

1	要素記号の一覧	2
2	各評価における作成訳の例	10
3	入力文の比較	11
4	実験結果	12
5	評価関数の信頼値	19
6	評価関数の信頼値	19
7	寄与率の比較	21
8	パターン元字面適合率のみによる実験結果	21
9	寄与率の比較	23
10	第8候補までの累積正解率(クローズドテスト)	24
11	第8候補までの累積正解率(オープンテスト)	24
12	各条件下における正解適合パターンの一意選択精度	25

1 はじめに

要素合成法を基本とする従来の機械翻訳システムでは、言語表現を分割する過程で全体の意味が失われることが問題であった。この問題を解決するには、文構造と意味を一体的に扱う仕組みが必要である。

近年提案された等価的類推思考の原理に基づく機械翻訳方式 [1] は、言語表現の意味類型化を基本とした翻訳方式である。翻訳対象となる両言語を「文型パターン」によって記述しておき、意味的に等価な文型パターンを対応付けることで、意味の失われぬ解析・生成を実現しようとしている。文型パターンは原言語と目的言語の線形要素を特定の記号に置き換えたものである。ここで線形要素とは、言語表現の構成要素と同等の意味を持つ他の要素に置き換えても、元の言語表現全体の概念が変化しない要素である。

この方式の実現に向けて、日本語の重文・複文を対象とした文型パターンを大量に蓄積した文型パターン辞書の構築が進められている [2]。現在、入力文と文型パターンを照合し、入力文に適合する文型パターンを抽出する文型パターンパーサが試作されている [3]。この文型パターンパーサを使用すると、複数の文型パターンが入力文に適合する。入力文の意味を考えると適合した文型パターンの中には英文生成に不向きなパターンも含まれる。よって、入力文の英訳を出力するために適合パターンの選択が必要となる。

そこで本研究では、入力文の英文生成が可能な適合パターンの選択手法を検討する。具体的には、入力文に適合した文型パターンを多変量解析によって分析し、評価関数を求める。得られた評価関数を使用して適合パターンの得点を求め、英文生成に使用する適合パターンの選択を行う。また、評価関数作成に用いたパラメータを分析することで、正しい文型パターンを選択するために必要な要素を明らかにする。

本研究では以下、第 2 章では文型パターンの説明、第 3 章では多変量解析による評価関数作成、第 4 章では適合パターン選択実験、第 5 章では考察、第 6 章にまとめを述べる。

2 文型パターン

2.1 文型パターンの記述形式

文型パターンは可読性と網羅性を意識して設計されており, 字面, 変数, 関数, 記号で記述されている [4]. 日英文対応の対訳コーパスの原文を, 単語レベル, 句レベル, 節レベルにパターン化した構造を持つ. 各レベルの粒度でアライメントが取れた部分は, 線形要素として変数化されている. また, 変数化すると対訳の訳出が困難になる部分は変数化されず, 非線形要素として字面, あるいは関数の形式で残されている.

以下に原文 (L) と単語レベルパターン (W) を示す. 文型パターンは, 日本語文型パターン (WJ) と, 対応する英語文型パターン (WE) で記述されており, 変数を介して両言語の要素対応付けができる.

LJ :将来は作家になりたいと思っている.

LE :I want to become a writer in the future.

WJ : $TIME1$ は/ $N2$ に/ $V3.tai$ / と思っている.

WE :I want to $V3$ $N2$ in $TIME1$.

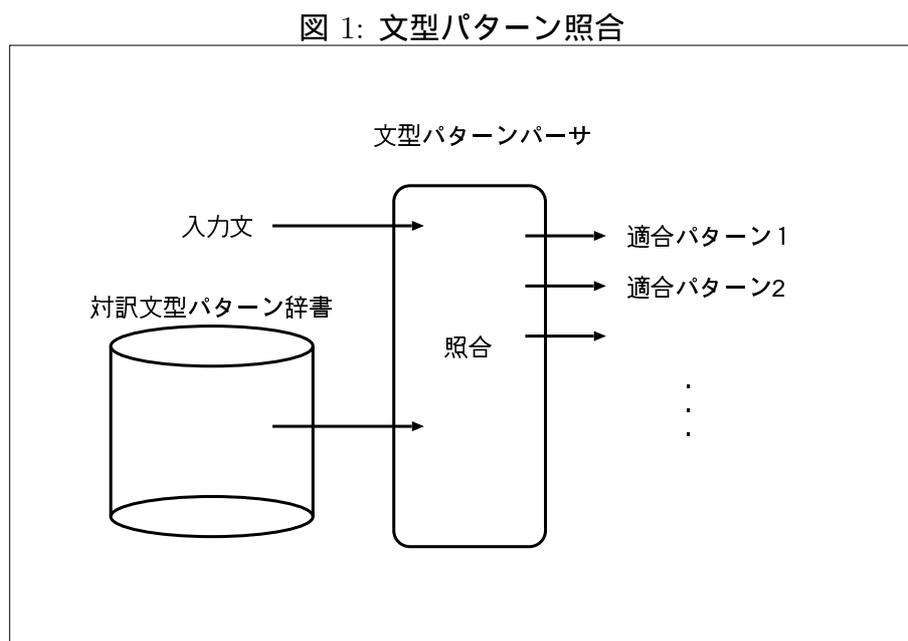
変数には名詞や動詞の単語を表す Nn や Vn など 8 種類がある. 関数には $.tai$ や $.kako$ などがあり, 字面の指定や表現の統括を行う. 記号はパターン記述要素の適合の仕方について, 任意化, 選択, 順序変更などの制御を行う (表 1).

表 1: 要素記号の一覧

記号名	表記	意味
選択要素記号	(... ...)	いずれかの要素列と適合
任意要素記号	[...]	文型選択上, 任意の要素
補間要素記号	< ... >	ゼロ代名詞等
順序任意要素 指定記号	{... ...}	順序入れ換え可能な範囲 (例 各要素の順序)
位置変更可能 要素指定記号	$\$n^{\{ \text{定義} \}}$ $\$n$	指定位置に入れ換え可能 (例 副詞の位置)

2.2 文型パターン照合

文型パターンの照合は、対訳文型パターン辞書から、入力文に適合する文型パターンを全て検索すると同時に、パターン記述要素の適合の仕方、すなわち、「適合解」を全て求めることである [3]。文型パターンが入力文に適合する条件は、文型パターンの全ての記述要素が、パターンの指定する順序通りに、入力文と対応することである。図 1 に文型パターン照合の様子を示す。



対応関係に曖昧性があれば、別の適合解として出力する。ここで対応関係の曖昧性とは、対応する入力文の部分が異なる場合である。例えば、任意要素の部分に適合する要素がある場合と、任意要素の部分に適合する要素がない場合を別々に適合解として出力する。以下に例を示す。

入力文

青い海に行きたいと思っている。

適合パターン 1

$WJ1 : \#1[AJ2]/N3$ に $/V4.tai$ と $/思っている$.

青い海に 行きたいと 思っている.

適合パターン 2

$WJ2 : \#1[AJ2]/N3$ に $/V4.tai$ と $/思っている$

海に 行きたいと 思っている.

2.3 文型パターンを利用した英文生成

文型パターンを利用した英文生成は, 適合パターンに対応する英語文型パターンを使用する [5]. 日本語文型パターンの変数と対応する入力文の箇所を翻訳し, 英語文型パターンに対応する箇所と置換することで英文を生成する. 置換の際, 英語パターンの変数に付随する関数に従い, 単語を変形させる. 単語レベルの文型パターンを利用した英文生成の例を示す.

入力文

将来は作家になりたいと思っている.

適合パターン 1

$WJ1$: $TIME1$ は $N2$ に $V3.tai$ / と思っている.

$WE1$: I want to $V3$ $N2$ in $TIME1$.

作成訳 1

I want to become a writer in the future

適合パターン 2

$WJ2$: $N1$ は $N2$ に $V3.tai$ / と思っている.

$WE2$: $N1$ be thinking of $V3.ing$ to $N2$.

作成訳 2

The future is thinking of becoming to a writer.

適合パターン 1 による入力文の英文生成を説明する. "将来" と " $TIME1$ ", "作家" と " $N2$ ", "なる" と " $V3$ " が対応する. よって "将来" の訳 "the future" を英語パターンの " $TIME1$ " と置換し, "作家" の訳 "a writer" を英語パターンの " $N2$ " と置換し, "なる" の訳 "become" を英語パターンの " $V3$ " と置換する. 作成した英文 (作成訳 1) は入力文の高品質の英訳と言える.

適合パターン 2 による入力文の英文生成を説明する. "将来" と " $N1$ ", "作家" と " $N2$ ", "なる" と " $V3$ " が対応する. よって "将来" の訳 "the future" を英語パターンの " $N1$ " と置換し, "作家" の訳 "a writer" を英語パターンの " $N2$ " と置換し, "なる" の訳 "become" を英語パターンの " $V3$ " と置換する. 作成した英文 (作成訳 2) は入力文の英訳として使用できない.

上記の例より, 適合パターンに対応する英語文型パターンを使用しても, 必ずしも良質の英文を生成できるとは限らないことがわかる. よって, 品質の良い英文生成が可能な適合パターンを選択する必要がある.

3 多変量解析による評価関数の作成

3.1 本研究の目的

本研究では、複数の適合パターンの中から、入力文の高品質な英文生成が可能な適合パターンを選択する。まず、テスト入力文の適合パターンを多変量解析によって分析し、評価関数を求める。次に、得られた評価関数で適合パターンの得点を求め、翻訳に使用する適合パターンの選択を行う。

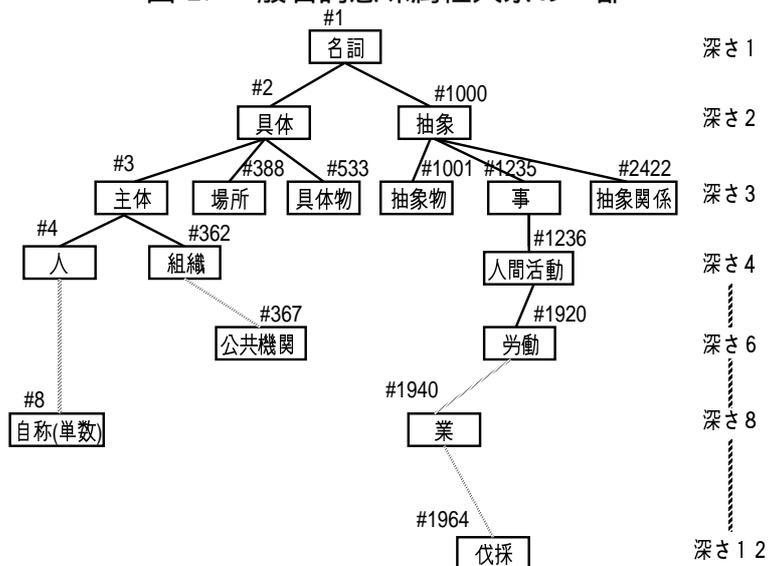
3.2 意味属性大系

適合パターン選択評価関数を求める際に名詞、動詞の意味属性距離を調べ、評価関数のパラメータとする。具体的には、入力文と、適合パターンの元となる原文との間で、変数を介して対応する箇所の名詞、動詞の意味属性距離を調べ、評価関数を求める際にパラメータとして使用する。意味属性は、日本語語彙大系 [6] に記載されている「一般名詞意味属性大系」および「用言意味属性大系」を使用する。

3.2.1 一般名詞意味属性大系

一般名詞意味属性大系は、名詞の意味的用法に着目してシソーラスとして大系化されている (図 2)。約 40 万語の名詞を、最大 12 段の木構造で 2710 の意味属性に分類している。また、一般名詞意味属性大系は、木構造を基本構成とし、各ノードにおいて上位の意味属性の性質を下位の意味属性が継承する。

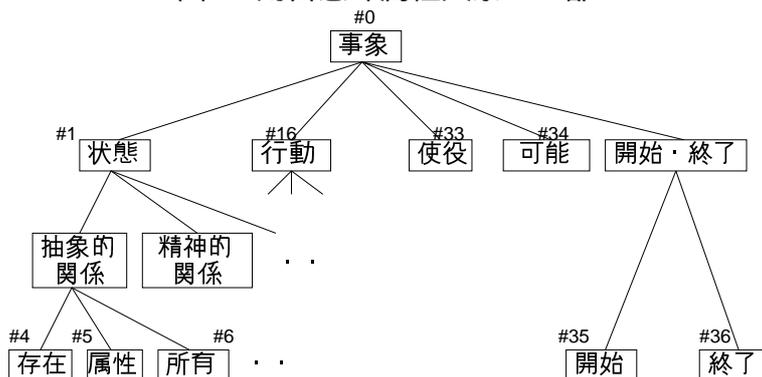
図 2: 一般名詞意味属性大系の一部



3.2.2 用言意味属性大系

用言意味属性大系 [7] は, 用言の意味的用法に着目してシソーラスとして大系化されている (図 3). 約 6000 語の用言を, 最大 4 段の木構造で, 36 の意味分類に分類されている. また, 用言意味属性大系は, 木構造を基本構成とし, 各ノードにおいて上位の意味属性の性質を下位の意味属性が継承する.

図 3: 用言意味属性大系の一部



3.3 評価関数

まず, 適合パターンを使用してテスト入力文を人手で英文を生成する. 次に, 作成した英文の品質を各適合パターンの評価値とし, 入力文と適合パターンの関係からパラメータを抽出する. 最後に, テスト入力文の適合パターンのパラメータと評価値を多変量解析によって分析し, 評価関数を求める.

適合パターンの以下のパラメータを評価パラメータとする. 評価パラメータを回帰分析することで評価関数を求める. 評価関数を \hat{y} (式 1) とし, 評価値 y との残差 $e(e = y - \hat{y})$ の 2 乗の総和を最小にする回帰係数 b_1, \dots, b_7 と切片 a の値を求める.

< 評価関数 >

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 \text{ (式 1)}$$

< 評価パラメータ >

- y : 評価値

適合パターンに対応する英語文型パターンを使用して, 入力文を人手で翻訳する. 作成した英文の品質で適合パターンを評価する. 評価は以下の A~D の 4 段階で行い, 評価関数作成の際には評価に応じた値を使用する.

評価 A : 1

品質の高い英文が生成できる

評価 B : 0.66

重要ではない要素の欠如はあるが簡単に修正可能

評価 C : 0.33

入力文を部分的に訳せている

評価 D : 0

入力文の訳としては使用不可能

- x_1 : パターン適合率

入力文と、適合パターンの文字単位の一一致する割合をパターン適合率とする。単語単位で計算し、パターンに適合している単語数と入力文の総単語数の除算で求める。以下の例では、”道路”と”N1”，”横断する”と”V2^{rentai}”，”注意しなさい”と”V3^{meirei}”が対応しており、入力文の全ての要素がパターンに適合している。よってパターン適合率は1.00(16文字/16文字)である。

入力文

道路を横断するときは注意しなさい。

適合パターン

N1 を V2^{rentai} ときは V3^{meirei}。

- x_2 : パターン字面適合率

入力文と、適合パターンに共通する字面の一一致する割合をパターン字面適合率とする。単語単位で計算し、一致単語数と入力文の総単語数の除算で求める。以下の例ではパターン字面適合率は0.43(3単語/7単語)である。

入力文

道路 を 横断する ときは 注意しなさい。

適合パターン

N1 を V2^{rentai} ときは V3^{meirei}。

- x_3 : パターン元字面適合率

入力文と、適合パターンを作成する際に用いた原文に共通する字面の一一致する割合をパターン元字面適合率とする。単語単位で計算し、一致単語数と入力文の総単語数の除算で求める。以下の例ではパターン字面適合率は0.71(5単語/7単語)である。

入力文

道路 を 横断する ときは 注意しなさい。

適合パターンの原文

線路 を 渡る ときは 注意しなさい。

- x_4 :記号の適合率

適合パターンに含まれる要素記号が使用される割合を記号の適合率とする。
表1の記号を対象にする。

- x_5 :変数の適合率

適合パターンに含まれる変数が、入力文との適合に使用される割合を変数の適合率とする。

- x_6 :名詞の平均意味属性距離の逆数

入力文と、適合パターンの元となった原文との間で、変数を介して対応する名詞箇所に関して意味属性距離を調べ、平均値の逆数を使用する。平均意味属性距離が0の場合は1とする。以下の例では、名詞の平均意味属性距離の逆数は0.25(1/4)である。

入力文

道路 を 横断するときは 注意しなさい。

適合パターン

$N1$ を $V2^{\wedge}rentai$ ときは $V3^{\wedge}meirei$ 。

適合パターンの原文

線路 を 渡る ときは 注意しなさい。

- x_7 :動詞の平均意味属性距離の逆数

入力文と、適合パターンの元となった原文との間で、変数を介して対応する動詞箇所に関して意味属性距離を調べ、平均値の逆数を使用する。平均意味属性距離が0の場合は1とする。以下の例では、動詞の平均意味属性距離の逆数は1(平均意味属性距離0のため)である。

入力文

道路 を 横断するときは 注意しなさい。

適合パターン

$N1$ を $V2^{\wedge}rentai$ ときは $V3^{\wedge}meirei$ 。

適合パターンの原文

線路 を 渡る ときは 注意しなさい。

3.4 評価関数作成の実験条件

入力文には, 対訳文型パターン集を作成した際に使用した原文約 12 万文からランダムに 200 文を選び, テスト入力文として評価関数作成に使用する. ただし, 入力文から作成した文型パターンが適合した場合は実験に使用しない.

文型パターンパーサは *wjpp.ver.2.4*[3] を使用する. 英文の生成は文型パターンパーサの出力を人手で修正する. 回帰分析には MicrosoftOffice の回帰分析ツールを使用する.

3.5 評価関数作成結果

テスト入力文 200 文のうち 72 文に適合パターンが存在した. 1 入力文に対して平均 26 パターンの文型パターンが適合した. 各入力文毎に最大 30 パターンまで調査し, 765 パターンを評価関数作成に使用した. 得られた評価関数を (式 2) に示す.

$$\hat{y} = -0.403 + 0.122x_1 - 0.194x_2 + 0.498x_3 + 0.027x_4 + 0.208x_5 + 0.195x_6 + 0.130x_7 \quad (\text{式 2})$$

式 2 より, 評価パラメータ x_3 (パターン元字面適合率) の回帰係数が最も高い値になった.

3.6 各評価における作成訳の例

評価 A から D の出力英文の例を表 2 に示す。各評価毎に、入力文、翻訳例、適合パターン、作成訳を示す。

表 2: 各評価における作成訳の例

評価 A	
入力文	愛情を持続させることは難しい
翻訳例	It is hard to keep love alive.
適合パターン (日)	$N1$ を / ($V2.sase^rentai$ $V2^sase^rentai$)! ことは / 難しい.
適合パターン (英)	It is quite impossible to $V2$ all $N1$.
作成訳	It is quite impossible to continue all love.
評価 B	
入力文	彼女は彼にすぐ行くよう命じた
翻訳例	She commanded him to go at once.
適合パターン (日)	$\$1^{\{N1\text{は}\}}/N2$ に $\$1/V3.suitei$ $\$1/V4.kako$.
適合パターン (英)	$N1 V4.past N2.obj$ to $V3$.
作成訳	She commanded him to go.
評価 C	
入力文	これは今まで使ったなかでいちばんおもしろい辞書だ
翻訳例	This is the most interesting dictionary I have ever used.
適合パターン (日)	$\#1$ [その/辺で]/いちばん/ $AJ2^rentai$! $N3.da$.
適合パターン (英)	It is $AJ2.st N3 \#1$ [thereabout].
作成訳	It is most interesting dictionary.
評価 D	
入力文	母は赤ん坊をあやして笑わせた
翻訳例	Mother played with the baby and got him to smile.
適合パターン (日)	$\$1^{\{N1\text{は}\}}/N2$ を / $V3$ (て で) $\$1 /V4.kako$.
適合パターン (英)	$N1 be.past V4.ed$ at $V3.ing N2$.
作成訳	Mother was made laugh at amusing baby.

4 適合パターン選択実験

4.1 選択関数の評価

得られた評価関数を使用して、品質の高い英文を生成できる適合パターンの選択を行う。得られた評価関数に適合パターンの評価パラメータの値を代入し、各適合パターンの評価関数の値 (\hat{y}) を求める。評価 A,B の適合パターンを正解適合パターンとし、各入力文毎に第 8 位までの累積正解率で関数を評価する。

テスト入力文 200 文を使用してクローズドテストを行う。オープンテストには、対訳文型パターン集を作成した際に使用した原文約 12 万文から、テスト入力文以外の 200 文をランダムに抽出して使用する。

4.2 入力文の比較

クローズドテスト、オープンテストで使用する入力文の比較を表 3 に示す。調査対象入力文は適合パターンを持つ入力文の数である。本研究では各入力文ごとに 30 パターンまで調査に使用したため、総適合パターン数と調査パターン数は異なる。ランダム選択における平均正解含有率は以下の式で求める。

平均正解含有率

$$= \frac{\sum \text{各入力文における正解パターンの割合} (\%)}{\text{調査対象入力文の数}} \quad (\text{式 3})$$

表 3 の調査対象入力文、総適合パターン数、平均適合パターン数を比較すると、クローズドテストに使用する入力文はオープンテストに使用する入力文より、多くの文型パターンに適合する入力文が多いとわかる。

表 3: 入力文の比較

	クローズド	オープン
入力文数	200 文	200 文
調査対象入力文	72 文	82 文
総適合パターン数	1847 パターン	1716 パターン
平均適合パターン数	26 パターン	21 パターン
調査パターン数	765 パターン	680 パターン
正解適合パターンを持つ入力文の数	29 文	29 文
正解パターンの割合	13%(96/765)	14%(95/680)
平均正解含有率	17%	23%

4.3 適合パターン選択実験結果

適合パターン選択実験を行った結果を表4に示す。表4より、クローズドテストでは72%、オープンテストでは83%の入力文において、第1候補に正解適合パターンが存在した。また、ほぼ全ての入力文において第8候補までに正解適合パターンが存在した。

表4: 実験結果

候補	クローズドテスト	オープンテスト
第1候補	72%(21/29)	83%(24/29)
~第2候補	86%(25/29)	90%(26/29)
~第4候補	90%(26/29)	100%(29/29)
~第8候補	97%(28/29)	100%(29/29)

以下に正解適合パターンが第1候補に存在した入力文の例を示す。評価が高い適合パターンが上位になっていた。

入力文 この対立には根の深い差別意識がある。

翻訳例 There is a deep-rooted sense of discrimination behind this confrontation.

第1候補 評価関数の値:0.598(評価A)

WJ1 GEN1/N2 には/根の/深い!N3N4 が/ある。

WE1 There is an ingrained N4 of N3 behind AJ1 N2.

作成訳1 There is an ingrained consciousness of discrimination behind this opposition.

第2候補 評価関数の値:0.236(評価A)

WJ2 \$1^{}/N1 には }/AJ2^rentai!N3 が\$1/ある。

WE2 N1 have AJ2 N3.

作成訳2 The opposition has deep discrimination consciousness.

第4候補 評価関数の値:0.192(評価D)

WJ4 \$1^{}/#1[N2 の]} /N3 には/AJ4^rentai!N5 が\$1/ある。

WE4 #1[N2.poss] N3 have AJ4 N5 to them.

作成訳4 The opposition has deep discrimination consiousness to them.

入力文 体裁のいい本だ
翻訳例 The book is elegantly got up.

第 1 候補 評価関数の値:0.561(評価 A)
WJ1 /体裁の/いい!N1.da.
WE1 N1 be smartly got up.
作成訳 1 This book is smartly got up.

第 4 候補 評価関数の値:0.227(評価 A)
WJ4 </N1 は>#2[/まことに]/体裁の/f いい!N3.da.
WE4 (N1|it) be a smart-looking N3.
作成訳 4 It is a smart-looking book.

第 5 候補 評価関数の値:0.199(評価 D)
WJ5 </N1 は>/N2 の/AJ3^rentai!N4.da.
WE5 N1 be N4 of AJ3 N2.
作成訳 5 It is this book of good appearance.

入力文 私は子どもの将来を思うと切ない。
翻訳例 I get distressed when I think of my children's future.

第 1 候補 評価関数の値:0.553(評価 A)
WJ1 N1 は#2[/N3 の]/TIME4 を/V5 と/切ない。
WE1 N1 get stressed out when N1 V5 of #2[N1.poss N3]'s N4 .
作成訳 1 I get stressed out when I think of my child's future.

第 1 候補 評価関数の値:0.553(評価 A)
WJ2 N1 は/N2 の/TIME3 を/V4 と/切ない
WE2 N1 feel disheartened when N1 V4 of N1.poss N2's N3 .
作成訳 2 I feel disheartened when I think of my children's future.

第 3 候補 評価関数の値:0.09(評価 D)
WJ3 N1 は/V2 と/AJ3.
WE3 N1's AJ3 to V2 to.
作成訳 3 I's distressing to think to.

5 考察

5.1 適合パターン選択失敗の原因

不正解適合パターンが第1候補にある入力文の大部分は、大量の文型パターンに適合していた。入力文に大量の文型パターンが適合するとき、入力文と適合パターンおよび原文から得られる表面的な情報に差が少なかった。不正解適合パターンの中にも、名詞・動詞の平均意味属性距離が小さい適合パターンがある。よって、正解適合パターンを第1候補にできなかったと考えられる。多くの文型パターンに適合する入力文がオープンテストには少なかったため、オープンテストの結果がクローズドテストの結果より良い値を示したと考えられる。

以下に大量の文型パターンが適合した入力文の例を示す。上記の理由により、正解文型パターンが第1候補に存在しなかった。各適合パターンの評価関数の値の差が少ない傾向があった。

以下に例を示す。

入力文 母は赤ん坊をあやして笑わせた (適合パターン数 204)

翻訳例 Mother played with the baby and got him to smile.

第1候補 評価関数の値:0.298(評価 D)

WJ1 $\$1^{\wedge}\{/N1\text{は}\}/N2\text{を}\$/V3(\text{て}|\text{で})\$/V4.kako|ND4\text{をした}.$

WE1 $N1(V4|V(ND4)).past, V3.ing\text{ behind }N2.$

作成訳1 Mother made laugh, amusing behind baby.

第2候補 評価関数の値:0.291(評価 D)

WJ2 $\$1^{\wedge}\{/N1\text{は}\}/N2\text{を}/V3(\text{て}|\text{で})\$/1\#4[!ADV5]/V6.kako.$

WE2 $N3\text{ be.}past(V4|V(ND4)).past\text{ to }V2\ N1.$

作成訳2 Mother was made laugh to amuse baby.

・
・
・

第5候補 評価関数の値:0.282(評価 A)

WJ5 $\$1^{\wedge}\{/tk\ N1\text{は}\}/N2\text{を}\$/V3(\text{て}|\text{で})\$/V4.kako.$

WE5 $N1\ V3.past\ N2\ \text{and}\ V4.past.$

作成訳5 Mother amused her baby and made laugh.

入力文 警官が来て騒ぎを鎮めた (適合パターン数 44)
翻訳例 The policemen came and got the things under control.

第 1 候補 評価関数の値:0.277(評価 D)
WJ1 *N1 が/V2(て | で)/N3 を/V4.kako.*
WE1 *N1 V2.past to V4 N3.*
作成訳 1 The policeman came to suppress ruckus.

第 1 候補 評価関数の値 0.277:(評価 D)
WJ1k *N1 が/V2(て | で)/N3 を/V4.kako.*
WE1 *N1 V2.past to V4 N3.*
作成訳 1 The policeman came to suppress ruckus.

第 3 候補 評価関数の値 0.276:(評価 A)
WJ3 *N1 が/V2(て | で)/N3 を/V4.kako.*
WE3 *N1 V2.past and V4.past N3.*
作成訳 3 The policeman came and suppressed ruckus.

入力文 親に隠れて煙草を吸った (適合パターン数 32)
翻訳例 I hid from my parents and smoked a cigarette.

第 1 候補 評価関数の値:0.288(評価 D)
WJ1 *</N1 は>#2[/今日は]/V3(て | で)/V4.kako.*
WE1 *N1 V4.past with V3.ing.*
作成訳 1 I smoked with hiding.

第 2 候補 評価関数の値:0.235(評価 D)
WJ2 *</N1 は>/N2 に/V3(て | で)#4[/窓の]/N5 を/V6.kako.*
WE2 *N1 V3.past into N2 and V6.past N5.*
作成訳 2 I hid into my parents and smoked a cigarette.

.
.
.

第 20 候補 評価関数の値:0.130(評価 A)
WJ20 *</N1 は>/N2 に/(V3|ND3をし)(て | で)/N4 を/V5.kako.*
WE20 *(N1|I) (V3|V(ND3)).past N2 , V5.ing (N1|I).pron.poss N4 .*
作成訳 20 I hid my parents, smoking my cigarette.

5.2 適合パターン数と評価関数の値の関係

正解・不正解適合パターンの頻度と評価関数の値の関係について調査した。クローズドテストの適合パターン数と評価関数の値の関係を図4に示す。オープンテストの適合パターン数と評価関数の値の関係を図5に示す。

クローズドテスト並びにオープンテストにおいて、正解適合パターンの数が最も多い評価関数の値は0.22であった。評価関数の値が0.2以下になると、クローズドテスト並びにオープンテストで不正解文型パターンの数が増大する傾向があった。

図4: 適合パターン数と評価関数の値の関係 (クローズド)

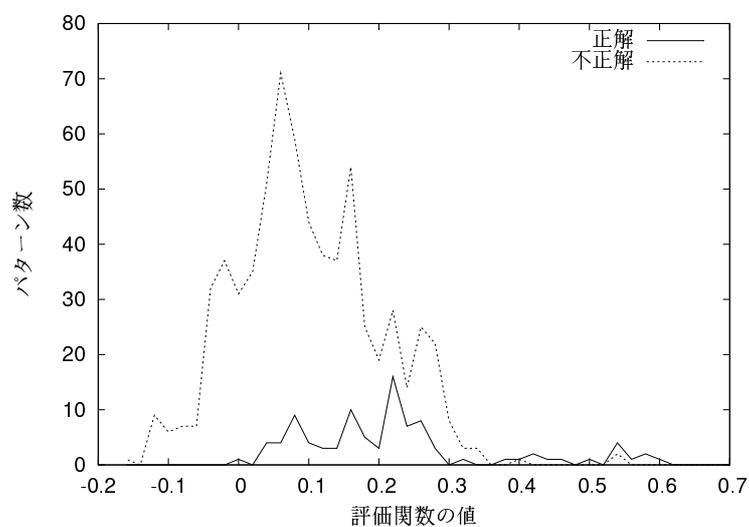


図5: 適合パターン数と評価関数の値の関係 (オープン)

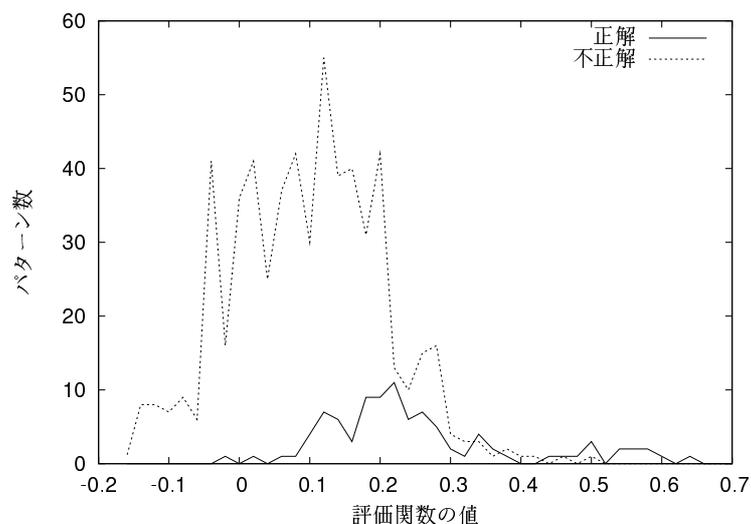


図4の累積表示を図6に, 図5 累積表示を図7に示す.

クローズドテストでは, 評価関数の値が0.3で正解適合パターン数と不正解適合パターン数が逆転した. オープンテストでは, 評価関数の値が0.28で正解適合パターン数と不正解適合パターン数が逆転した. よって閾値により適合パターンの正解・不正解を判断する場合, 評価関数の値が0.3以下に設定すると, 不正解適合パターンを正解適合パターンと判断する割合が増大すると推測できる.

図 6: 評価関数の値と累積パターン数の関係 (クローズド)

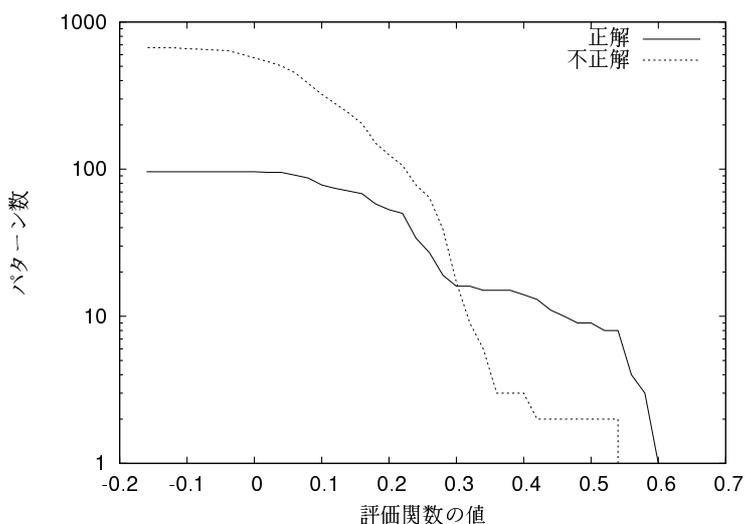
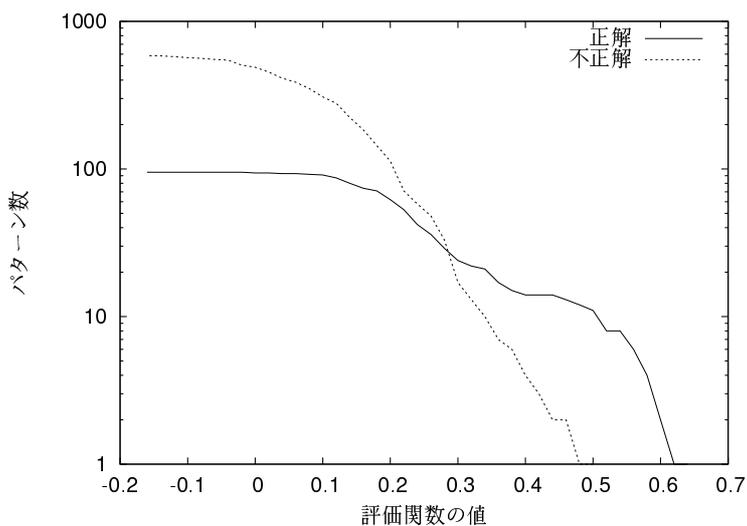


図 7: 評価関数の値と累積パターン数の関係 (オープン)



5.3 適合パターンの最大値と評価

各入力文ごとに、評価関数による値が最大の適合パターンを調べ、正解・不正解と評価関数の値の関係を調査した。クローズドテストの結果を図8、オープンテストの結果を図9に示す。図より、適合パターンの最大値が5.0以上の入力文は、最大値を持つ適合パターンで入力文を翻訳できると推測できる。適合パターンの最大値が0.1以下の入力文は、最大値を持つ適合パターンでは入力文の翻訳は困難であると推測できる。

図8: 適合パターンの最大値と評価の関係 (クローズド)

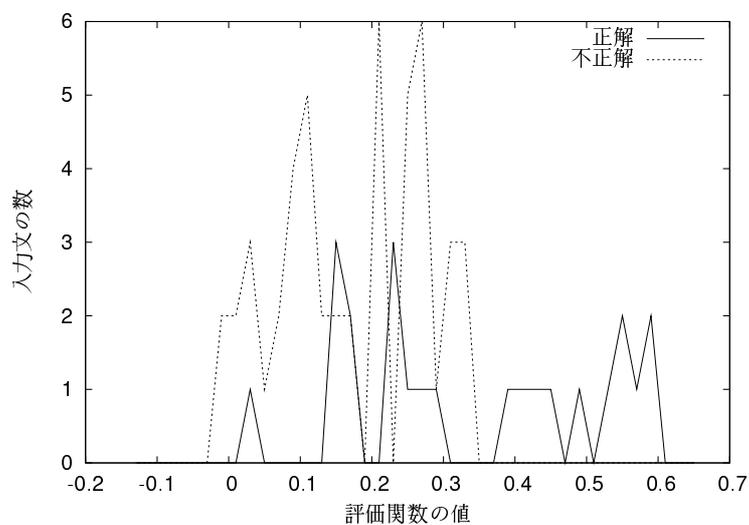
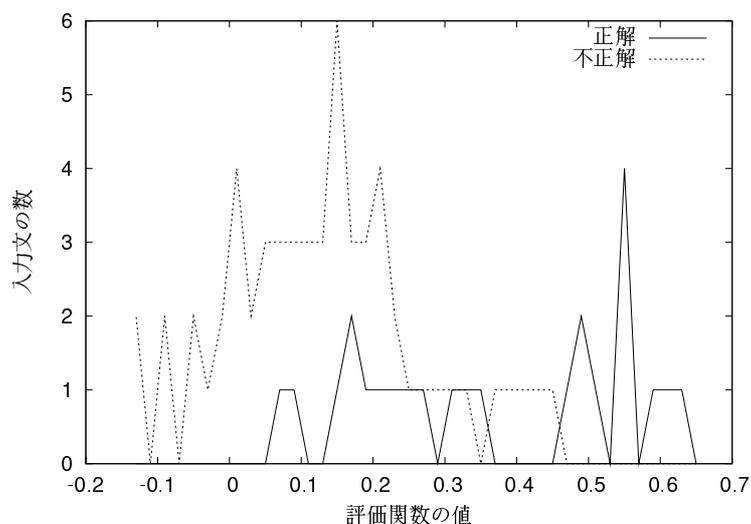


図9: 適合パターンの最大値と評価の関係 (オープン)



評価関数による値に閾値を設定し、第1候補の適合パターンが正解・不正解か判別できるか調査した。調査方法は、第1候補の適合パターンが正解・不正解か正しく判別できる入力文の数を調べ、適合パターンが存在した入力文の数で除算する。クローズドテストの結果を表5に、オープンテストの結果を表6に示す。

表5、表6より、閾値を下げれば、第1候補の正解適合パターンが正しいと判断できる入力文の数が増加する。しかし、第1候補が不正解適合パターンだと判断できる入力文の数が減少する。図8、図9より、閾値0.2付近で、第1候補に不正解適合パターンが存在する入力文の数が増大する。0.2以下を閾値とすると、第1候補に不正解適合パターンが存在する入力文の区別が困難になると予想できる。

閾値を0.4に設定すると、正解適合パターンを正しく判断できた入力文の割合は低い、不正解適合パターンを全て正しく判断できた入力文の割合は高かった。よって、評価関数の信頼値は最も高い値を示した。

表 5: 評価関数の信頼値 (クローズドテスト)

閾値	信頼値	正しく判断できた割合	
0.6	60%	正 0%(0/29)	不 100%(43/43)
0.5	68%	正 21%(6/29)	不 100%(43/43)
0.4	73%	正 34%(10/29)	不 100%(43/43)
0.3	66%	正 38%(11/29)	不 86%(37/43)
0.2	50%	正 59%(17/29)	不 44%(19/43)

表 6: 評価関数の信頼値 (オープンテスト)

閾値	評価値	正しく判断できた割合	
0.6	67%	正 7%(2/29)	不 100%(53/53)
0.5	74%	正 28%(8/29)	不 100%(53/53)
0.4	74%	正 38%(11/29)	不 94%(50/53)
0.3	74%	正 48%(14/29)	不 89%(47/53)
0.2	67%	正 62%(18/29)	不 70%(37/53)

本研究で調査した入力文は、正解適合パターンがある入力文の数が、不正解適合パターンしか無い入力文の数より少ない。よって、第1候補が不正解適合パターンだと正しく判断できる閾値で評価値が高い値を示す傾向が見られた。

5.4 評価関数の説明力

本研究では評価関数を評価値 y と評価パラメータ x の関係から求めている。しかし、データのばらつきが大きければ評価関数の信頼性が低くなる問題がある。そこで、得られた評価関数が全体のデータのばらつきをどの程度説明しているか調査した。

評価関数の信頼性は自由度修正済 R^2 値によって判断する。自由度修正済 R^2 値は、評価値のデータの変動が評価パラメータの変動でどの程度説明できるか表している。よって、自由度修正済 R^2 値 (R'^2) で評価関数の寄与率 (説明力) がわかる。寄与率は、以下の式で求めた。

$$\text{寄与率 (\%)} = R'^2 \times 100 \quad (\text{式 3})$$

$$R'^2 = 1 - \frac{S_E / (n - p - 1)}{S_T / (n - 1)} \quad (\text{式 5})$$

$$S_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\text{式 6})$$

$$S_T = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (\text{式 7})$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (\text{式 8})$$

p : 評価パラメータの数 (8)

n : 評価関数作成に使用した適合パターン数 (765)

\hat{y} : 評価関数の値

y : 評価値

本研究で求めた評価関数の寄与率と、各評価パラメータ単独で回帰分析し、寄与率を求めた結果を表 7 に示す。本研究で求めた評価関数は適合パターンの選択に効果があったが、寄与率は 16.5% と低い値であった。名詞の意味属性情報より動詞の意味属性情報は、適合パターン選択に必要な情報を多く持つとわかった。多くの適合パターンを持つ入力文において、パターン字面適合率に差が少ない場合が多く、パターン字面適合率の寄与率が低くなった原因だと推測できる。

寄与率の値から、適合パターン選択に必要な情報を最も持つ評価パラメータはパターン元字面適合率であると考えられる。次節では、パターン元字面適合率のみを評価パラメータとし、得られた評価関数の適合パターン選択精度を検証した。

表 7: 寄与率の比較

関数	寄与率
本研究作成評価関数	16.5%
パターン適合率のみ	7.1%
パターン字面適合率のみ	5.7%
パターン元字面適合率のみ	11.4%
変数の適合率のみ	2.7%
記号の適合率のみ	1.8%
名詞意味属性のみ	3.5%
動詞意味属性のみ	4.7%

5.5 パターン元字面適合率

評価パラメータ単独で回帰分析を行った結果, パターン元字面適合率のみで求めた評価関数の寄与率が最も高かった. パターン元字面適合率と評価値の関係から得られた評価関数で, 適合パターン選択実験を行った. 評価パラメータが単独のため, 複数の適合パターンが第 1 候補になる場合が多かった. クローズドテストでは 9 文, オープンテストでは 7 文の入力文において, 正解・不正解適合パターンが第 1 候補に混在していた. そこで, 第 1 候補に不正解適合パターンが含まれる入力文を除外し, 本研究で求めた評価関数と比較した. 結果を表 8 に示す.

表 8 より, パターン元字面適合率単独で求めた評価関数は, 本研究で求めた評価関数より選択精度は劣るが, 多くの入力文において正解適合パターンを一意に選択できた. よって, パターン元字面適合率は適合パターン選択に必要な情報を多く持つと言える.

表 8: パターン元字面適合率のみによる実験結果

評価関数	クローズド	オープン
本研究作成評価関数	69%(20/29)	83%(24/29)
パターン元字面適合率	59%(17/29)	66%(19/29)

5.6 様々な評価関数

適合パターン選択において、本研究で評価関数を求める際に使用した評価パラメータ全てが必ずしも必要ではない可能性がある。評価パラメータの数を減らすことができれば、評価関数を求める時間の減少が見込める。評価パラメータの削減を目指し、様々な条件下で評価関数を求め、適合パターン選択精度を調査した。

以下の条件下で実験を行った。

- 名詞意味属性評価関数
評価パラメータに x_6 (名詞意味属性距離の平均の逆数) のみを使用
- 動詞意味属性評価関数
評価パラメータに x_7 (動詞意味属性距離の平均の逆数) のみを使用
- 意味属性評価関数
評価パラメータに x_6 (名詞意味属性距離の平均の逆数), x_7 (動詞意味属性距離の平均の逆数) を使用
- パターン元字面適合率+意味属性評価関数
評価パラメータに x_3 (パターン元字面適合率), x_6 (名詞意味属性距離の平均の逆数), x_7 (動詞意味属性距離の平均の逆数) を使用

以下の項目を、本研究で求めた評価関数、パターン元字面適合率のみで求めた評価関数と比較した。

- 寄与率
- 第8候補までの累積正解率
- 正解適合パターンの一意選択精度
- 適合パターンの最大値と評価の関係

5.6.1 寄与率の比較

各条件下で評価関数を求め、寄与率を比較した。結果を表9に示す。

パターン元字面適合率と意味属性で評価関数を求めたところ、本研究で作成した評価関数と同じ寄与率になった。よって、パターン元字面適合率、意味属性を評価パラメータとし、求めた評価関数は、本研究で求めた評価関数と同程度の適合パターン選択能力を持つと考えられる。

表 9: 寄与率の比較

関数	寄与率
本研究作成評価関数	16.5%
名詞意味属性のみ	3.5%
動詞意味属性のみ	4.7%
名詞・動詞意味属性のみ	7.4%
パターン元字面適合率のみ	11.4%
パターン元字面適合率 +名詞・動詞意味属性	16.5%

5.6.2 各条件下における適合パターン選択精度

各条件下で求めた評価関数で適合パターン選択実験を行い、結果を比較した。クローズドテストの結果を表10に示す。オープンテストの結果を表11に示す。

どの条件で求めた評価関数を使用しても、適合パターン選択実験を行った結果、第8候補までに正解文型パターンが存在した。パターン元字面適合率のみ、名詞意味属性距離の平均の逆数のみ、動詞意味属性の平均の逆数のみを評価パラメータとして求めた評価関数は、一見適合パターン選択精度が高く見える。しかし、評価パラメータが単独のため、第1候補に複数の適合パターンがあり、正解・不正解適合パターンが混在していた。よって、一意に正解適合パターンを選択するのは困難であると予想できる。そこで、第1候補に不正解適合パターンが含まれる入力文を除外し、正解適合パターンの一意選択精度を求めた。結果を表12に示す。

表 10: 第 8 候補までの累積正解率 (クローズドテスト)

	本研究作成	名詞意味属性	動詞意味属性	意味属性
第 1 候補	72%(21/29)	72%(21/29)	90%(26/29)	65%(19/29)
第 2 候補	86%(25/29)	76%(22/29)	93%(27/29)	76%(22/29)
第 4 候補	90%(26/29)	86%(25/29)	97%(28/29)	86%(25/29)
第 8 候補	97%(28/29)	97%(28/29)	97%(28/29)	93%(27/29)

	パターン元字面適合率	パターン元字面適合率 +意味属性
第 1 候補	90%(26/29)	76%(22/29)
第 2 候補	93%(27/29)	86%(25/29)
第 4 候補	100%(29/29)	90%(26/29)
第 8 候補	100%(29/29)	97%(28/29)

表 11: 第 8 候補までの累積正解率 (オープンテスト)

	本研究作成	名詞意味属性	動詞意味属性	意味属性
第 1 候補	83%(24/29)	66%(19/29)	93%(27/29)	83%(24/29)
第 2 候補	90%(26/29)	86%(25/29)	97%(28/29)	86%(25/29)
第 4 候補	100%(29/29)	97%(28/29)	100%(29/29)	97%(28/29)
第 8 候補	100%(29/29)	100%(29/29)	100%(29/29)	100%(29/29)

	パターン元字面適合率	パターン元字面適合率 +意味属性
第 1 候補	90%(26/29)	83%(24/29)
第 2 候補	97%(28/29)	90%(26/29)
第 4 候補	100%(29/29)	100%(29/29)
第 8 候補	100%(29/29)	100%(29/29)

表 12: 各条件下における正解適合パターンの一意選択精度

関数	クローズドテスト	オープンテスト
本研究作成評価関数	69%(20/29)	83%(24/29)
名詞意味属性のみ	52%(15/29)	55%(16/29)
動詞意味属性のみ	28%(8/29)	48%(14/29)
意味属性のみ	55%(16/29)	72%(21/29)
パターン元字面適合率	59%(17/29)	66%(19/29)
パターン元字面適合率 +意味属性	69%(20/29)	79%(23/29)

パターン元字面適合率と意味属性で求めた評価関数で、第1候補に不正解適合パターンが混在した入力文を調べた。以下に例を示す。

入力文 双方の言い分を聞かないと真相はわからない

翻訳例 You cannot learn the truth unless you hear what both parties have to say.

評価 C

WJ </N1 は>/N2 を/V3.hitei と</N4 は>/V5.hitei.

WE If N1 don't V3 any N2, N4 can't V5.ing on this.

原文(日) 金を貰わないと動けない。

原文(英) If I don't get any money, I can't get moving on this.

作成訳 If I don't hear any claim, the truth can't understanding on this.

評価 B

WJ <N1 は>/N2 を/V3.hitei と</N4 は>/V5.hitei.

WE If N1 don't V3 any N2, N4 can't V5.ing on this.

原文(日) 金を貰わないと動けない。

原文(英) If I don't get any money, I can't get moving on this.

作成訳 If I don't hear any claim, I can't understanding on this.

評価 C の候補では、”真相” が”N4” に適合している。従って、英語文型パターンの”N4” と”the truth” を置換させている。評価 B の候補では、”真相” が”N4” に適合していないと判断している。よって、英語文型パターンの”N4” に置換する要素がない。本研究では、英文生成の際に補間要素の不一致による要素の欠如がある場合、補間要素の中の変数が対応する箇所に”I” を挿入している。日本語では、聞き手もしくは読み手が文脈から用意に推測できる場合には、主語や目的語等の各要素が省略される場合が多いからである [9]。名詞、動詞の意味属性距離は対応関係が取れる箇所しか調べない。よって、パターン元字面適合率と意味属性から求めた評価関数では、2つの適合パターンは同じ値を持った。

パターン元字面適合率から求めた評価関数で, 第 1 候補に不正解適合パターンが混在した入力文を調べた. 以下に例を示す.

入力文 彼は頭を打って記憶を喪失した

翻訳例 He hit his head and lost his memory.

評価 A

WJ $\$1^{\wedge}\{/N1\text{は}\}/N2\text{を}/V3(\text{て}|\text{で})\$1/N4\text{を}/V5.kako.$

WE $N1\ V5.past\ N4\ \text{by}\ V3.ing\ N1.poss\ N2.$

原文 彼は私財をなげうって難民を救った.

作成訳 He lost his memory by hitting his head.

評価 D

WJ $\$1^{\wedge}\{/N1\text{は}\}/N2\text{を}/V3(\text{て}|\text{で})\$1/N4\text{を}/V5.kako.$

WE $N1\ V5.past\ N4\ V3.past\ N2.$

原文 彼は手紙を同封して物を送った.

作成訳 He lost his memory hit his head.

評価 D

WJ $\$1^{\wedge}\{/N1\text{は}\}/\text{頭を}/V2(\text{て}|\text{で})\$1\#3[!REN4]/N5\text{を}/V6.kako.$

WE $N1\ V2.past\ N1.poss\ \text{thinking and}\ V6.past\#3[AJ4]\ N5.$

原文 彼は頭を切り替えて別の方法を試した.

評価 A, 評価 D の適合パターンは, どれもパターン元字面一致数が同じである. よって, パターン元字面適合率から作成した評価関数ではどちらの適合パターンも同じ値になるため, 正解適合パターンを一意に選択できない.

大量に文型パターンが適合する入力文では, 適合した文型パターンのパターン元字面適合率に差が無い傾向があった. よって, パターン元字面適合率のみで求めた評価関数では, 大量に文型パターンが適合する入力文の正解適合パターンの選択は困難であると予想される.

パターン元字面適合率のみ, 名詞意味属性距離の平均の逆数のみ, 動詞意味属性距離の平均の逆数のみを評価パラメータとして求めた評価関数は, 第1候補に正解適合パターンが存在した入力文の数が多かった。しかし, 多くの入力文において第1候補に正解・不正解適合パターンが混在しており, 正解適合パターンを一意に選択することは困難であった。特に動詞の意味属性距離の平均の逆数のみを使用して求めた評価関数では, 第1候補に正解適合パターンが存在するが, 不正解適合パターンも混在する入力文が多く, 動詞の意味属性距離のみで適合パターンの選択は難しいと言える。

評価パラメータにパターン元字面適合率と意味属性を使用して求めた評価関数は, 本研究で求めた評価関数と同程度の累積正解率, 正解適合パターンの一意選択精度を持っていた。

5.6.3 各条件下における適合パターンの最大値と評価の関係

正解適合パターンの一意選択精度において, 比較的高い選択精度を持っていた以下の3種類の評価関数について, 各入力文の評価関数による値が最大の適合パターンを調べ, 正解・不正解と評価関数の値の関係を調査した。

意味属性のみで求めた評価関数

パターン元字面適合率のみで求めた評価関数

パターン元字面適合率と意味属性から求めた評価関数

- 意味属性のみで求めた評価関数

意味属性のみで求めた評価関数を使用して、各入力文において適合パターン
の評価関数の値を求め、値が最大の適合パターンの正解・不正解と評価関数
による値の関係を調べた。クローズドテストの結果を図 10、オープンテスト
の結果を図 11 に示す。

評価関数の値が最大の正解・不正解の適合パターンの頻度が重なっていた。
閾値を設定しても、評価関数の値が最大の適合パターンの正解・不正解の正
確な判断は困難であると予想できる。

図 10: 適合パターンの最大値と評価の関係 (意味属性, クローズド)

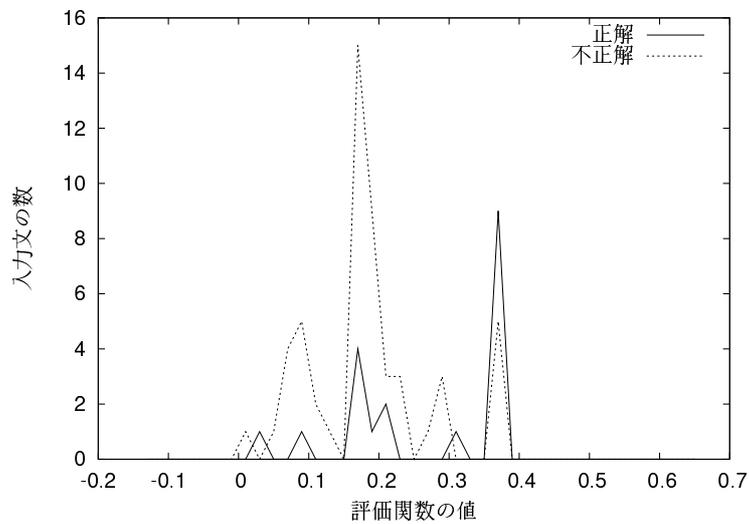
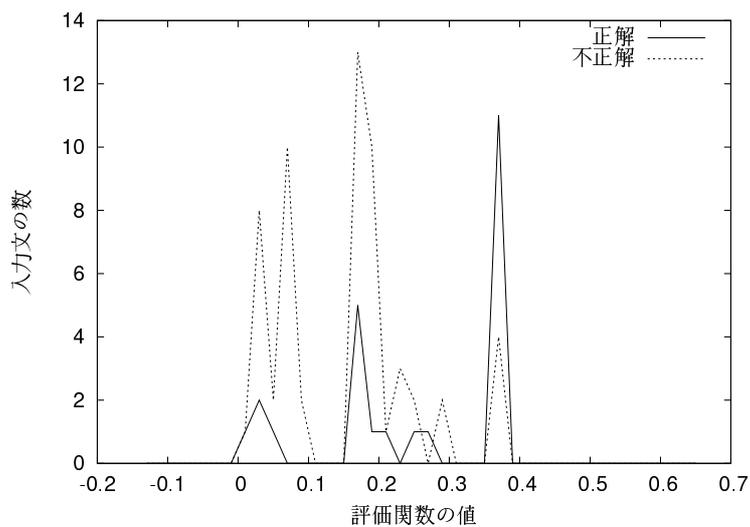


図 11: 適合パターンの最大値と評価の関係 (意味属性, オープン)



- パターン元字面適合率のみで求めた評価関数

パターン元字面適合率のみで求めた評価関数を使用して、各入力文において適合パターンの評価関数の値を求め、値が最大の適合パターンの正解・不正解と評価関数による値の関係を調べた。クローズドテストの結果を図 12、オープンテストの結果を図 13 に示す。

一部の評価関数の値が高い適合パターンを持つ入力文があるが、評価関数の値が 0.5 以下になると、正解・不正解の頻度が重なっていた。よって、評価関数による値が 0.5 以下では、評価関数の値が最大の適合パターンの正解・不正解の正確な判断は困難であると予想できる。

図 12: 適合パターンの最大値と評価の関係 (P 元, クローズド)

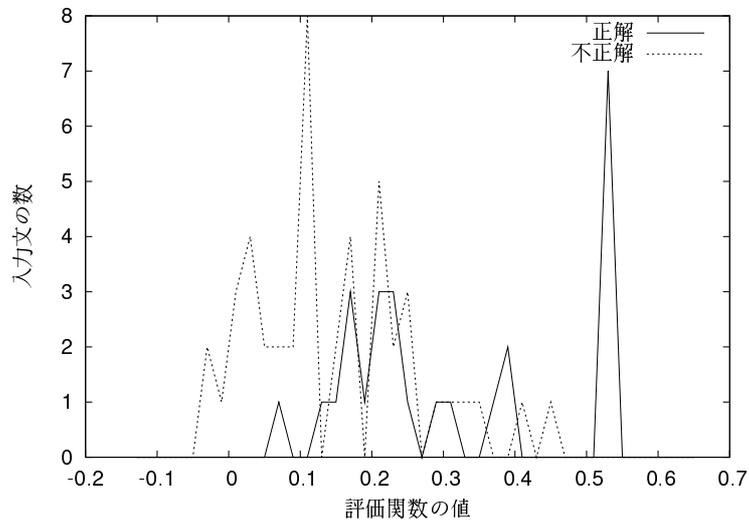
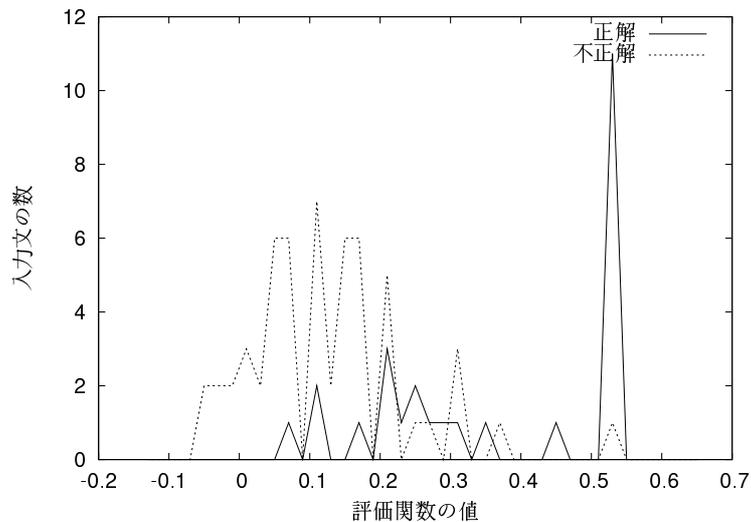


図 13: 適合パターンの最大値と評価の関係 (P 元, オープン)



- パターン元字面適合率と意味属性から求めた評価関数

パターン元字面適合率と意味属性から求めた評価関数を使用して、各入力文において適合パターンの評価関数の値を求め、値が最大の適合パターンの正解・不正解と評価関数による値の関係を調べた。クローズドテストの結果を図 14、オープンテストの結果を図 15 に示す。

評価関数の値の最大が正解適合パターンである入力文の頻度と、評価関数の値の最大が不正解適合パターンである入力文の頻度が、評価関数の値によってある程度分離できていた。よって、本研究で作成した評価関数と同程度の信頼値が期待できる。

図 14: 適合パターンの最大値と評価の関係 (P 元+意味属性, クローズド)

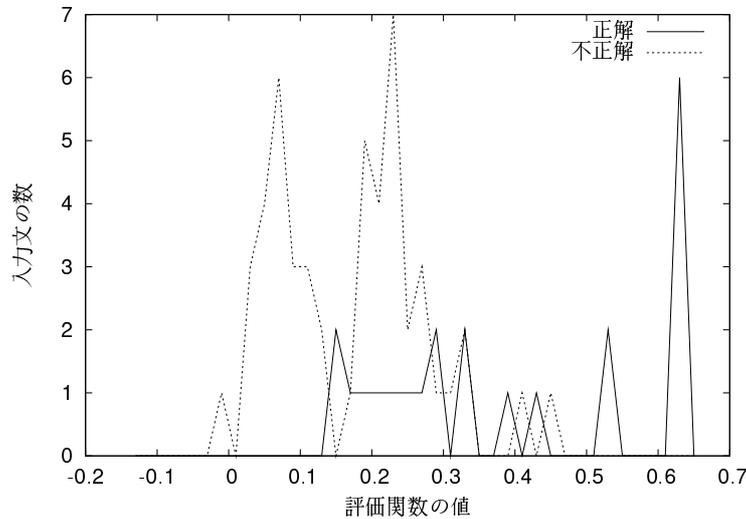
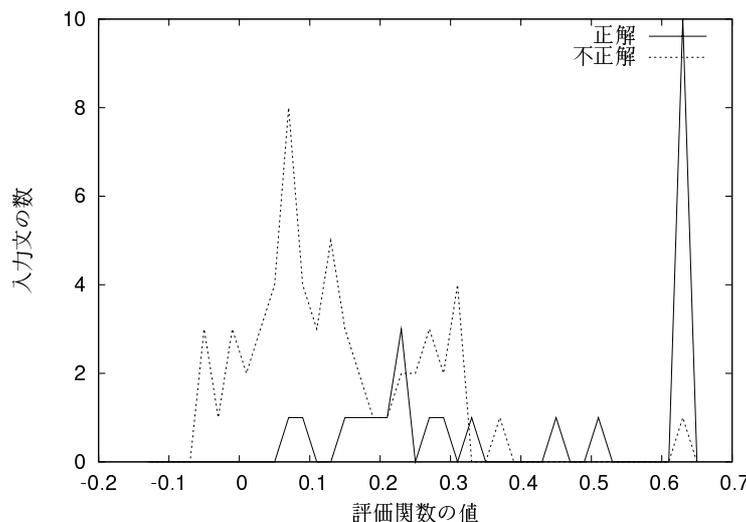


図 15: 適合パターンの最大値と評価の関係 (P 元+意味属性, オープン)



6 まとめ

本研究では, 等価的類推思考の原理に基づく機械翻訳方式の実現に向け, 文型パターンパーサが出力する複数の適合結果から, 入力文の英文生成に適した適合パターンの選択を行った. 具体的には, テスト入力文の適合パターンの情報を多変量解析によって分析し, 適合パターン選択評価関数を求めた. そして, 得られた評価関数で適合パターンの選択を行った.

実験の結果, クローズドテストでは 72%, オープンテストでは 83% の入力文において, 第 1 候補に正解適合パターンが存在した. また, ほぼ全ての正解適合パターンを持つ入力文において, 上位 8 位までに正解適合パターンが存在した. よって, 本研究で求めた評価関数は高品質の適合パターン選択精度を持っていた.

適合パターン選択に必要な要素を調べるため, 評価関数の計算に使用する評価パラメータを変更して適合パターン選択実験を行った. 評価パラメータに適合パターンの元となる原文, 名詞意味属性距離の平均の逆数, 動詞意味属性距離の平均の逆数を使用すれば, 本研究で求めた評価関数と同程度の品質を持つ評価関数を作成できることがわかった. よって, 評価パラメータ数の削減が期待できることがわかった.

今後は, 多くの文型パターンが適合する入力文において選択精度を上げる必要があると考えられる.

謝辞

最後に, 本研究において御指導頂きました鳥取大学工学部知能情報工学科池原教授, 村上助教授, 徳久助手に厚く御礼申し上げます. また, 本研究で使用させて頂いた論文, 本の著者に御礼申し上げます.

参考文献

- [1] 池原悟ほか：等価的類推思考の原理による機械翻訳方式，信学技報，TL2002-34，pp.7-12，2002.
- [2] 池原悟ほか：非線型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，自然言語処理，Vol.11，No.3，pp.69-95，2004.
- [3] 徳久雅人，池原悟，村上仁一：文型パターンパーサの試作，言語処理学会第10回年次大会発表論文集，pp.608-611，2004.
- [4] 池原悟ほか：機械翻訳のための日英文型パターン記述言語，信学技報，TL2002-48/NLC2002-90，pp.1-6，2003.
- [5] 前田春奈，池原悟，村上仁一：パターンを使用した重文複文の日英翻訳の精度，言語処理学会第10回年次大会発表論文集，pp.237-240，2004.
- [6] 池原，宮崎，白井 他：日本語語彙大系，岩波書店，1997.
- [7] 中岩浩巳，池原悟：日英の構文的対応関係に着目した日本語用言意味属性の分類，情報処理学会論文誌，Vol.38，No.2，pp.215-225.1997.
- [8] 渡辺美智子，小山斉：Excel 徹底活用統計データ分析，秀和システム，2003.
- [9] 中岩浩巳，山田節夫：日英対訳コーパスからのゼロ代名詞とその指示対象の自動抽出，言語処理学会第3回年次大会発表論文集，pp.409-412，1997.