

概要

本研究では、非線形言語モデルに基づく日英機械翻訳における格要素の翻訳について議論する。このモデルによる日英翻訳では、入力文の表現構造の意味を日文パターンで捉えるのだが、日文パターンから見て任意の格要素と判断された格要素は、対訳英文パターンでは翻訳されないという問題がある。そこで、本稿では、格要素と前置詞句のパターン対で構成する「格要素パターン辞書」を構築することを目的とする。既に構築された日文英文パターン対には格要素と前置詞句の対応が多く含まれていることに着目し、格要素と前置詞句のパターン対を抽出する。約22万件の重文複文文型パターン辞書から抽出した結果、925件のパターン対が得られた。日本語語彙大系と組み合わせて、単文の日英翻訳実験を行ったところ、日本語格要素パターンによる表現に対する網羅性は問題が無く、英語前置詞句パターンによる意味的なカバー率が79%であることが確認された。

目次

第1章	はじめに	1
第2章	研究の背景	3
2.1	パターン辞書	3
2.1.1	鳥バンク	3
2.1.2	日本語語彙大系	4
2.2	パターン照合システム SPM	4
2.3	パターン翻訳システム ITM	5
第3章	格要素パターン辞書の作成	7
3.1	パターンと辞書の名称	7
3.2	扱う格助詞と前置詞	7
3.3	作成方法	8
3.4	抽出例	9
3.5	作成結果	11
第4章	格要素パターン辞書を用いた機械翻訳	13
4.1	文パターンの拡張	13
4.2	翻訳方法	14
4.3	翻訳の様子	15
第5章	評価実験	19
5.1	評価方法	19
5.1.1	実験文の条件	19
5.1.2	評価の分類	20
5.2	実験結果	21

第6章 考察	23
6.1 事例の分析	23
6.1.1 出力文の候補において正解の前置詞が存在したもの（ の場合）	23
6.1.2 正解の前置詞句パターンが出力文候補に含まれていないもの（ の場合）	23
6.2 今後の対策	26
6.2.1 第6.1.1節に関する対策	26
6.2.2 第6.1.2節に関する対策	27
第7章 おわりに	28

目 次

2.1	日本語表現意味辞書における日英パターン対の例	3
2.2	日本語語彙大系データの 1 レコード	4
2.3	形態素解析の例	5
2.4	図 2.3 に対する SPM の出力における適合パターンの一例	5
2.5	ITM による翻訳例	6
3.1	格助詞と前置詞の関係	8
3.2	格要素パターン対のレベル	9
3.3	格助詞と前置詞の対が存在するパターン	9
3.4	格助詞相当句と前置詞の対が存在するパターン	10
3.5	格要素パターン辞書の一部	10
3.6	格助詞相当句	12
4.1	文パターンの拡張	13
4.2	翻訳処理の流れ	14
4.3	適合パターン情報 1	17
4.4	「適合パターン情報 2」の元になる格要素パターン候補群	17
4.5	新規英語パターンの例	17
4.6	翻訳例	18
5.1	任意格要素の翻訳が必要でない例	21
5.2	正解の表現で翻訳できた例	22
6.1	正解格要素パターンが選ばれなかった例	24
6.2	正解前置詞句が出力文候補に含まれていない例	25
6.3	新規格要素パターン対の例	26

表 目 次

3.1	抽出結果	11
5.1	入力文における任意格要素翻訳必要文数	22
5.2	5.1.2 節に基づく出力文の評価の内訳	22

第1章 はじめに

本研究は、非線形言語モデルに基づく日英機械翻訳 [1] における格要素の翻訳について現状を議論する。

非線形言語モデルは、まず大局的に、入力文の表現構造の意味を文単位のパターンで捉え、次に局所的に、そのパターンで定められる線形部分の意味を節、句、単語等を単位とするパターンで捉えるというモデルである。単文のパターンについては、日本語語彙大系で示されており [2]、重文・複文のパターンについては「鳥バンク」で公開されている [3]。

ここで、格要素は大きく2種類がある。1つは文構造の意味を定める上で必須のもの、もう1つは任意のものである。必須のものとは格要素の内容が変わると文全体の解釈が変わるもので、たとえば「小遣いを貰う」と「嫁を貰う」ではヲ格の内容の違いにより、「所有的移動」と「相対関係」という意味で違いが生じる。任意のものとは、その格要素の解釈とその格要素以外の解釈を合成して文全体の解釈とできるもので、たとえば「バスで学校へ行く」では「学校へ行く」という意味に「バスで」という「手段」の意味を線形的に加えることで、文全体が解釈できる。

日本語語彙大系においては、必須の格要素が洗練されて示されているが、任意の格要素が体系的には示されていない。鳥バンクにおいては、約12万件の日英文対について、線形部分と非線形部分の識別によりパターン化が行われ、任意の格要素の存在可能な位置が離散記号「/c」で示されたが、やはり、任意の格要素を陽に示すことはできていない。

こうした状況では、パターン翻訳を行う際、入力文に文パターンが適合しても、任意の格要素の翻訳が出来ないという問題がある。その解決には、任意の格要素を解釈するためのパターン辞書を新たに作成する必要がある。

一般的に、日英翻訳において任意の(=必須でない)日本語格要素は英語前置詞句で翻訳することが適当であるとすると、先行研究に見られるアライメント手法で格要素と前置詞句の対を抽出することが、解決方法の1つと言える [5], [6]。しかし、本稿では、鳥バンクにおける重文・複文の文型パターン辞書において、既に単語・句のアライメントがマークアップされていることに着目し、この辞書から格要素と前置詞句の対を抽出す

る方法を試みる。

非線形言語モデルに基づく格要素の扱いにおいて、(1) 格要素と前置詞句の対が任意のものとして使用可能かどうか、(2) 任意であるならばその解釈をどうするか（たとえば、「バスで」が「手段」と解釈するように）という点を考えなければならない。このうち本稿では、格要素と前置詞句の対の収集の試行、および、収集した全てを任意格要素と仮定して使用した翻訳の問題分析までを行うことを目的とする。

第2章 研究の背景

本章は，研究において必要な辞書，システムについて説明する．まず，機械翻訳に必要なパターン辞書，日本語の表現の解析器について説明し，それらを利用したパターン照合器について説明する．そして照合した結果を用いた翻訳器について説明する．

2.1 パターン辞書

2.1.1 鳥バンク

「鳥バンク」の「日本語表現意味辞書（重文複文編）」[3]には，日本語の重文・複文とその対訳英文の対を約12万文対，および，その文対でから作成された「意味類型パターン（22.7万件）」¹が収録されている．また本研究でのパターンは，[3]の記述仕様に従い，変数，字面，関数，および，記号で記述する．そのパターンの例を図2.1に示す．

日文: 列車で行くほうが安全だろう。

英文: It would be safer to go by train.

日文パターン:

/ytcfkN1(OR:列車,NI:988,IM:13690)で/cfV2(OR:行く,NY:0506,
NY:1500,NY:1801,NY:1804,NY:2001,NY:2301,NY:2302,NY:2303,
NY:2903,NY:3201,IY:6970,IY:8410)^rentai/fほうが/cfAJV3
(OR:安全だろ,NY:0506,IY:A610).you。

英文パターン:

It would be AJ3(OR:safer)^er to V2(OR:go)^base by N1(OR:train).

図 2.1: 日本語表現意味辞書における日英パターン対の例

図2.1では，日文の「列車」，「行く」，「安全」という単語が日文パターンにおいてそれぞれ $N1$ ， $V2$ ， $AJV3$ という変数に変換されているのがわかる．また，英文の「safer」，

¹これらをまとめて「鳥バンク」や「重文・複文文型パターン辞書」と呼ぶこともある．

「go」、「train」が同様に $AJ3$, $V2$, $N1$ に変換されているのがわかる。さらに、それらの変数の直後に記述されている「(...)」は、用例、もしくは意味属性によるその変数の制約を示している。

2.1.2 日本語語彙大系

日本語語彙大系には、結合価文法に基づく文型パターン(以下、文パターンと呼ぶ)が定義されている。パターンに記述された変数には意味属性による制約条件が付けられている。また、パターンには英文パターンも記述されている。具体例を図 2.2 に示す。

日文パターン:
$N1$ が $N2$ に/へ/まで $N3$ から/より 行く
制約条件:
$N1$ (3 主体 986 乗り物 535 動物) $N2$ (-418 道路 -419 鉄道 1057 創作物 (その他) 388 場所 2610 場) $N3$ (-418 道路 -419 鉄道 388 場所 2610 場) 英文パターン:
$N1$ go from $N3$ to $N2$

図 2.2: 日本語語彙大系データの 1 レコード

2.2 パターン照合システム SPM

SPM は、日本語文型パターン辞書を用いてパターン照合を行うシステムである [7]。SPM を使用するにあたって、まず、日本語文に対して解析器を用い、形態素解析を行う。入力文を単語ごとに分割し、それぞれに対して、品詞、意味属性など付与した情報を入力する。例を図 2.3 に示す。形態素解析された日本語文に対して、適合する文型パターンを検索する。形態素解析情報と適合したパターンの情報を入力する。図 2.3 に対する SPM の出力における適合パターンの例を図 2.4 に示す。

PATTERN 行を「=」で区切り、その 3 つ目の部分は、パターンが適合した形態素の番号を示している。また、変数行を「=」で区切り、その 1 つ目は、それに対応する形態素の番号を示している。

入力文	学生が学校に行く。
出力	1. /学生 (1100,NI:238,KR:1806s13,IM:11260) 2. +が (7410) 3. /学校 (1100,NI:405,NI:367,NI:863,KR:0101u04,KR:1806u00,...) 4. +に (7430) 5. /行く (2116,NY:18,NY:23,NY:15,NY:29,NY:32,NY:20,NY:5,...) 6. +。 (0110) 7. /nil

図 2.3: 形態素解析の例

$PATTERN=GT000768=[N1, が, N2, に, '行く']= [1,2,3,4,5]=5$
 $N1=[1]=1=1$
 $N2=[3]=3=1$

図 2.4: 図 2.3 に対する SPM の出力における適合パターンの一例

2.3 パターン翻訳システム ITM

ITM は、文型パターン辞書の文パターンを用いて日英パターン翻訳を行うシステムである。翻訳する文に対する形態素解析結果、それに対する SPM による日文パターンの照合結果、および、日文パターン対応する英文パターンを用いて翻訳を行う。英文パターンの変数の翻訳には、日英辞書引きを行い、尤度を用いて変数に対する複数の翻訳候補を決定し、翻訳文を出力する。この例を図 2.5 に示す。

入力文	学生が学校に行く。
形態素解析結果	1. /学生 (1100,NI:238,KR:1806s13,IM:11260) 2. +が (7410) 3. /学校 (1100,NI:405,NI:367,NI:863,KR:0101u04,KR:1806u00,...) 4. +に (7430) 5. /行く (2116,NY:18,NY:23,NY:15,NY:29,NY:32,NY:20,NY:5,...) 6. +。 (0110) 7. /nil
適合パターン	PATTERN=GT000768=[N1, が, N2, に, '行く']= [1,2,3,4,5]=5 N1=[1]=1=1 N2=[3]=3=1
対応する英文パターン	N1 go to N2
N1の翻訳候補	a student student students
N2の翻訳候補	a school school schooled schooling schools
翻訳結果	Students go to school

図 2.5: ITM による翻訳例

第3章 格要素パターン辞書の作成

本章では，格要素パターン辞書の作成方法を示す．

3.1 パターンと辞書の名称

日本語の格要素のパターンを「格要素パターン」，英語の前置詞句のパターンを「前置詞句パターン」と呼ぶことにする．さらに，格要素パターンと前置詞句パターンの対のことを単に「格要素パターン対」と呼び，格要素パターン対の集合を「格要素パターン辞書」と呼ぶことにする．

格要素パターン対には，3つのレベルがある．用例における制約付きの格要素パターン対のことを「用例レベル格要素パターン対」と呼ぶ．意味のレベルでの制約付きの格要素パターン対のことを「意味レベル格要素パターン対」と呼ぶ．制約を取り除いた格要素パターン対のことを「文法レベル格要素パターン対」と呼ぶ．

3.2 扱う格助詞と前置詞

今回使用する日本語格助詞は，[9]より，「が」を除く「を，に，と，で，へ，まで，より，から」とする．また，これらを含み，格助詞と同じような働きをする語句を格助詞相当句とする．格助詞相当句については，格要素パターン抽出時に作られるので，あらかじめ準備はしない．一方，英語前置詞に関しては，一般的な前置詞90件，および，その相当句45件を使用する．ただし「between $N1$ and $N2$ 」のような変数と字面が交互に出現するような複雑なパターンをもつ前置詞は除いている．また，これらは，付録Aに記載する．

格助詞と前置詞のパターンの関係を図1に示す．点線で囲まれた部分は今回使用する知識であり，囲まれていない格助詞相当句は，抽出時に作られることを示している．

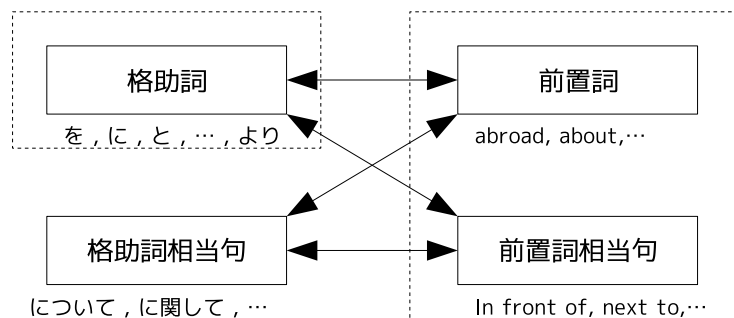


図 3.1: 格助詞と前置詞の関係

3.3 作成方法

「日本語表現意味辞書(重文複文編)」の日文英文パターン対から格要素パターン対を機械的に抽出する手順は次のとおりである。

1. 日文パターンより格助詞を検索する。
2. 格助詞の前の名詞変数からの表現, 後の文節境界までの表現がある場合, それを格助詞相当句とみなして記憶する。
3. 格助詞または格助詞相当句の直前に名詞変数があるならば, その変数を記憶する。
4. (2) と (3) より変数から格助詞もしくは格助詞相当句の終わりまでを格要素パターンとみなす。
5. 英文パターンより, 同一変数を検索する。
6. (5) の変数の直前に前置詞または前置詞相当句があるならば, それと変数をあわせて前置詞句パターンとみなす。
7. (4) の格要素パターンと (6) の前置詞句パターンを対にすることで, 格要素パターン対を得る。

ここで, 検索に失敗したり, 変数や前置詞等が存在しない場合, 格要素パターン対は得られない。また, 変数を扱う際, 変数の直後の制約を変数とともに抽出し, パターンに組み込むこととする。制約は「OR:...」より字面レベルでの条件記述がなされている。格要素パターン辞書には, 用例レベル格要素パターン対を登録する。しかし, 表現の広がり調べるために, 文法レベル格要素パターン対も集計する。その例を図 3.2 に示す。

用言レベル格要素パターン対:

N_k (OR:列車,NI:988,IM:13690) で by N_k (OR:train)

文法レベル格要素パターン対:

N_k で by N_k

図 3.2: 格要素パターン対のレベル

日文: 列車で行くほうが安全だろう。

英文: It would be safer to go by train.

日文パターン:

/ytcfk_N1(OR:列車,NI:988,IM:13690) で/cfV2(OR:行く,NY:0506,
NY:1500,NY:1801,NY:1804,NY:2001,NY:2301,NY:2302,NY:2303,
NY:2903,NY:3201,IY:6970,IY:8410)^rentai/f ほうが/cfAJV3
(OR:安全だろ,NY:0506,IY:A610).you.

英文パターン:

It would be AJ3(OR:safer)^er to V2(OR:go)^base by N1(OR:train).

図 3.3: 格助詞と前置詞の対が存在するパターン

3.4 抽出例

まず、鳥バンクのデータの例を図 3.3, 3.4 に示す。たとえば、図 3.3 の日文パターンにおいて、「で」の直前に変数 $N1$ が存在し、英文パターンの変数 $N1$ の直前に「by」が存在することから、これを格要素パターン対として取得する。また、図 3.4 の日文パターンにおいて、格助詞「に」を含む「において」の直前に変数 $N2$ が存在し、英文パターンの変数 $N2$ の直前に「about」が存在することから、同じく、これを格要素パターン対として取得する。この結果、下記の格要素パターン対が得られる。

格要素パターン対:

N_k (OR:列車,NI:988,IM:13690) で by N_k (OR:train)

N_k (OR:教育,NI:1808,IM:15265) について about N_k (OR:education)

同様にして得られる格要素パターン対のいくつかを図 3.5 に示す。ここでは用例レベルの格要素パターン対が列挙されている。同一の格助詞と同一の前置詞から成る対がみられる。したがって文法レベルのパターン対は用例レベルに比べて種類数が少ない。

日文: 親の中には 教育について独特の観念をもっている人が多い。
英文: Many parents have their own peculiar ideas about education.

日文パターン:

/ytcfkN1(OR:親,NI:1679,...) の/k(中—なか) には/tcfk_N2(OR: 教育,NI:1808,IM:15265)
について/cfREN3(OR:独特の)/kN4(OR:観念, NI:1027,...) を/cfV5(OR:もつ,NY:0506,
...).teiru^rentai/f(人—ひと—ヒト) が/cf 多い。

英文パターン:

Many N1(OR:parents) V5(OR:have) N1(OR:their)^pron^poss own AJ3(OR:peculiar)
N4(OR:ideas) about N2(OR:education)

図 3.4: 格助詞相当句と前置詞の対が存在するパターン

Nk (OR:教会,NI:377,NI:455,IM:11420,IM:12150) で at *Nk* (OR:a church)
Nk (OR:金利,NI:1198,NI:2596,IM:14A50,IM:16740) で at *Nk* (OR:rates)
Nk (OR:浜辺,NI:2667,NI:490,IM:12340,IM:168C0) で at *Nk* (OR:the beach)

Nk (OR:バス,NI:1056,NI:1349,NI:2354,NI:868,NI:988,...) で by *Nk* (OR:bus)
Nk (OR:電話,NI:1147,NI:1548,NI:970,IM:13680,...) で by *Nk* (OR:phone)

Nk (OR:雪,NI:2365,NI:744,IM:13540,IM:15424) で because of *Nk* (OR:the snow)
Nk (OR:病気,NI:2416,NI:2419,IM:15450) で because of *Nk* (OR:illness)

図 3.5: 格要素パターン辞書の一部

3.5 作成結果

格要素パターン対，および，格助詞相当句を抽出した結果を表に示す．また，抽出した格要素パターン対を文法レベルで付録Bに記載する．表において，用例レベルでの数から文法レベルでの数の圧縮率を見ると，1つの文法レベル格要素パターン対につき，平均12パターンの用例があることがわかる．

表 3.1: 抽出結果

抽出結果	抽出件数
用例レベルでの格要素パターン対総数	10,783
文法レベルでの格要素パターン対総数	925
格助詞相当句数	267

また，得られた格助詞相当句の例を図3.6に示す．今回，前置詞句との対応が見られた格助詞を含む表現は格助詞相当句と見なしている．したがって，得られた格助詞相当句の中には，「駅を」のような自立語を含む表現も格助詞相当句としているが，今後見直す必要がある．

あてに, かどうかについて, かに, からでない, からの, からは, からも, かを, がわに, ごとに, だけで, だけでは, だけに, だけを, づいているので, づたいに, でこそ, でしか, でだって, での, では, でも, でもある, という, ということにして, といっても, とか, とが, ときたら, として, としての, としては, としても, となつて, との, とは, とも, ともに, とを, と言う, と同時に, どちらに, なしで, なしでも, なしに, なしには, などから, などで, などへ, にあたって, にあたっては, において, にかけては, にさえ, にしか, にして, にしては, にだって, について, についての, については, についても, につき, につきまして, につきましては, につれて, にて, にでも, にと, にとつて, にとつての, にとつては, にとつても, にとり, には, にも, にもかかわらず, により, によりまして, による, によると, によれば, にわたり, にわたる, に沿つて, に関し, に関して, に関しては, に関しまして, に関しましては, に関する, に基づいて, に限り, に向かつて, に向け, に向けて, に向けての, に際して, に際しては, に至る, に至るまで, に従つて, に従つての, に対し, に対して, に対しては, に対しても, に対する, に伴い, に備えて, のくせに, のつもりで, のと, のところで, のに, のより, の為に, へと, への, へは, へも, へ向けて, へ向けての, みたいに, めに, もなしで, もなしに, よりも, をはじめ, をめぐつて, をめぐる, をもつて, を介して, を介しての, を含め, を通して, を通しての, を通じて, を通つて, テストを, 宛てに, 宛に, 案を, 以下で, 以外に, 以外には, 以外にも, 以後に, 以上であり, 以上に, 以内に, 一同を, 駅を, 沿いに, 下で, 下では, 下に, 過ぎに, 解析よりも, 開始してから, 間に, 間近になると, 期間内であれば, 気味で, 客に, 業務では, 近くに, 型に, 経営から, 経路で, 建設を, 権を, 個人にとつても, 故に, 後で, 後に, 向きでは, 向きに, 向けに, 攻めに, 降りに, 沙汰を, 姿であつた, 誌上で, 事故現場から, 時に, 時代から, 時代からの, 時代に, 自身に, 自体には, 社に, 社員として, 取引銀行に, 住まいを, 上からも, 上で, 上に, 上には, 上への, 上を, 場に, 状に, 進化を, 政体では, 前から, 前に, 前には, 前後であつた, 全土で, 全土にわたり, 全土にわたる, 増加を, 側から, 速度を, 代金を, 大学に, 大学を, 達を, 中で, 中であつたり, 中である, 中であるが, 中です, 中での, 中でも, 中に, 中には, 中を, 直後で, 通りに, 程度に, 的に, 的には, 店に, 伝いに, 同士で, 内で, 内に, 内を, 内蔵用として, 内部から, 内部で, 半分に, 版を, 比で, 品を, 付きで, 物を, 方しだいで, 方向に, 方面に, 方面へ, 法を, 無しに, 名から, 面での, 役として, 余りに, 用と, 用として, 用に, 用を, 用地として, 陽画にも, 留学を, 列を

図 3.6: 格助詞相当句

第4章 格要素パターン辞書を用いた機械 翻訳

本章では，日本語語彙大系の文パターン辞書と本研究の格要素パターン辞書を用いた翻訳を行う。

4.1 文パターンの拡張

日本語語彙大系の文パターンには，任意格要素の位置が示されておらず，任意格要素を含む入力文に適合させるためには，それを示す必要がある。また，文パターンが入力文に対して，任意的に適合したり，要素が前後入れ替わっている場合などに対応する必要がある。ゆえに，日本語語彙大系の文パターンを図 4.1 のように拡張する。

新たに追加されている $YRSNBF_k$ という変数は，任意格要素である部分を示している。[] で示している要素は，任意的に適合する部分であることを示しており，() 内では，| に区切られたいずれかが適合するようになっている。また，{} で示された部分は「,(コンマ)」ごとの要素が入れ替わって適合してもよいことを示している。

拡張前:

N1 が N2 に/へ/まで N3 から/より 行く

拡張後:

/cnj/y{[YRSNBF7][NP1(が|は)[、]],[YRSNBF8][NP2(に|へ|ま
で|は)[、]],[YRSNBF9][NP3(から|より|は)[、]][YRSNBF10]' 行
く'

図 4.1: 文パターンの拡張

4.2 翻訳方法

図 4.2 を参照しながら，日英翻訳の手順を説明する．

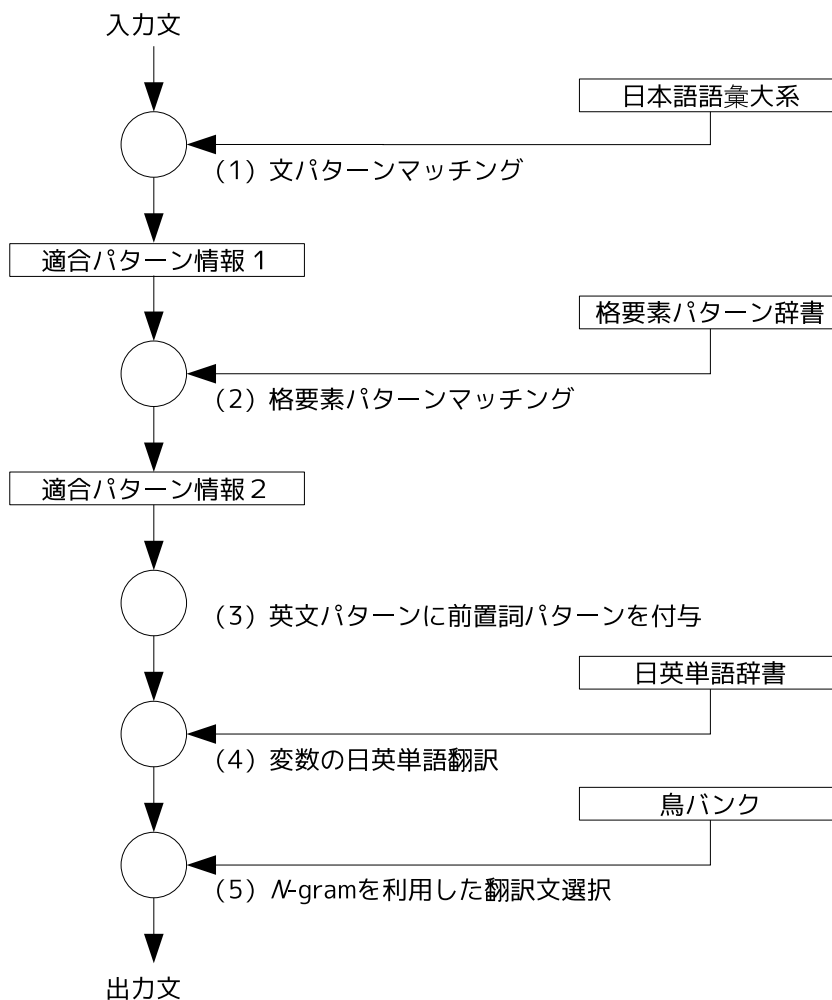


図 4.2: 翻訳処理の流れ

1. 日本語の入力文に対して形態素解析した後に，照合器 SPM により日文パターンと照合を行う．それにより適合した情報が「適合パターン情報 1」である．これは，入力文に対する形態素解析情報，それに適合した日文パターン情報と変数情報，および，英文パターンで構成する．
2. (1) で日文パターンの任意格要素部分である変数に適合した格要素がある場合，格

要素パターン辞書の格助詞パターンと照合し，ここでも変数の意味属性制約を検査する．この結果（適合した格要素パターン，それに対応する前置詞句パターン，および，変数の情報）を(1)の出力に付与したものが「適合パターン情報 2」である．

3. (1)の英文パターンに，(2)の前置詞句パターンを付与する．このときの前置詞句パターンで，格要素パターン辞書作成時に，抽出元となった英文パターンの先頭で使われていたものは(1)の英文パターンの先頭に，それ以外で使われていたものは後方にそれぞれ付与する．それらの判断は前置詞句パターンの先頭が大文字か小文字で行う．なお，英文パターンで入力文に対して必要のない前置詞句部分は削除する．
4. (3)によって作られた英文パターンの変数に対応する日本語を翻訳器 ITM により日英単語翻訳する．ここで，変数は名詞や形容詞であるので，単なる辞書引きで実現できる．このとき，複数の候補がある場合，全てを翻訳候補として登録する．
5. 入力文に対する複数の翻訳候補（パターンと英単語の両方に曖昧性がある）に対して，英単語の tri-gram で文のスコアを算出する．このスコアが最も高いものを出力とする．

4.3 翻訳の様子

翻訳の具体例を示す．「学生がバスで学校に行く。」という入力文を文パターンに照合した様子を以下に示す．

複数の適合パターンの中の 1 つを図 4.3 に示している．この情報は図 4.2 の「適合パターン情報 1」に対応している．

任意格要素部分である変数 $YRSNBF8$ の情報を格要素パターン辞書と照合する例を図 4.4 に示す．この情報を「適合パターン情報と変数情報 1」に付与したものが，図 4.2 の「適合パターン情報と変数情報 2」に対応している．図 4.4 の情報をもとにして，前置詞句を英文パターンに付与した結果を図 4.5 に示す．図 4.4 では，「at N_k 」，「by N_k 」のパターンの前置詞の先頭は小文字なので，図 4.5 の新規英文パターンでは後方に付与されている．また，図 4.3 の英文パターンで存在した「from $N3$ 」は，入力文に対して必要のない前置詞部分であるので削除している．

これらによって新しく作成された英文パターン，形態素解析情報，日文パターンと変数情報で翻訳器 ITM を用いて日英単語翻訳を行う．そして，複数の翻訳候補の中から

tri-gram によって選出を行う。日英翻訳した例を図 4.6 に示す。任意格要素に対する前置
詞句パターン候補が複数ある中から、tri-gram により、「by」がもっとも高いスコアと
なり、出力文のパターンに選ばれている。

入力文	学生がバスで学校に行く。
形態素解析情報	1. /学生 (1100,NI:238,KR:1806s13,IM:11260) 2. +が (7410) 3. /バス (1100,NI:2669,NI:988,NI:868,NI:2354,...) 4. +で (7410) 5. /学校 (1100,NI:405,NI:367,NI:863,KR:0101u04,...) 6. +に (7430) 7. /行く (2116,NY:18,NY:23,NY:15,NY:29,NY:32,...) 8. +。 (0110) 9. /nil
適合パターン情報	[NP1, が,YRSNBF8,NP2, に,'行く']= [1,2,3,4,5,6,7]=7 NP1=[1]=1=1 YRSNBF8=[3,4]==2 NP2=[5]=5=1 EPATTERN=NP1 go from NP3 to NP2

図 4.3: 適合パターン情報 1

適合パターン情報 :

N_k で in N_k (OR:自動車,NI:988,IM:13690)

N_k で on N_k (OR:電車,NI:988,IM:13690)

N_k で by N_k (OR:バス,NI:1056,NI:1349,NI:2354,NI:868,NI:988,...)

...

図 4.4: 「適合パターン情報 2」の元になる格要素パターン候補群

新規英語パターン:

$N1$ go to $N2$ in $N3$

$N1$ go to $N2$ on $N3$

$N1$ go to $N2$ by $N3$

...

図 4.5: 新規英語パターンの例

入力文：学生がバスで学校に行く。

適合パターン情報:

N_1 が N_2 に/へ/まで N_3 から/より 行く。

N_1 go from N_3 to N_2

N_1 (3 主体 986 乗り物 535 動物)

N_2 (-418 道路 -419 鉄道 1057 創作物 (その他) 388 場所 2610 場)

N_3 (-418 道路 -419 鉄道 388 場所 2610 場)

任意格要素:バスで

前置詞句の候補と新英文パターンによる翻訳時の文の N -gram の値:

by N_k : -48.690

of N_k : -52.434

for N_k : -58.266

to N_k : -58.500

in N_k : -58.736

on N_k : -58.959

with N_k : -59.365

at N_k : -59.652

but N_k : -61.444

After N_k : -68.572

With N_k : -68.863

like N_k : -69.354

出力された翻訳文:Students go to school by bus

図 4.6: 翻訳例

第5章 評価実験

入力文の任意格要素に対する格要素パターン辞書の効果を評価する．

5.1 評価方法

5.1.1 実験文の条件

実験には，辞書の例文より抽出した単文コーパス [4] から，格助詞を含む簡単な単文 100 文に対して翻訳を行う．具体的な条件は以下の通りである．

- 能動態である文
- 連用修飾語を含まない文
- 形態素解析に成功した文
- 日本語語彙大系の結合価パターンと適合する文

本稿で評価の対象とする文は，正解英語文が前置詞を使用した文であることと，入力文に対して日本語語彙大系の文パターンだけでは補えず，任意格要素の翻訳が必要である文である．以下では，このような文を任意格要素翻訳必要文とする．なお，以上のように，はじめから評価対象の文が決まらないのは，任意格要素翻訳必要文の判定基準が，日本語語彙大系のパターンにあるためである．

5.1.2 評価の分類

次に、評価する基準を4つ説明する。

...出力文において、正解文と同じ前置詞が存在したもの

の例

入力文：学生がバスで学校に行く。
任意格要素：バスで
前置詞句候補：by Nk, in Nk, on Nk, ...
出力文：Students go to school by bus
正解文：A student goes to school by bus.

... 以外で、出力文の候補において、正解の前置詞が存在したもの

の例

入力文：学生がバスで学校に行く。
任意格要素：バスで
前置詞句候補：by Nk, in Nk, on Nk, ...
出力文：Students go to school on bus
正解文：A student goes to school by bus.

...入力文の任意格要素に対して、格要素パターンが適合したが、正解の前置詞句パターンが出力文候補に含まれていないもの

の例

入力文：学生がバスで学校に行く。
任意格要素：バスで
前置詞句候補：with Nk, in Nk, on Nk, ...
出力文：Students go to school on bus
正解文：A student goes to school by bus.

× ...入力文の任意格要素に対して，格要素パターンが使われていないもの

× の例

入力文：学生がバスで学校に行く。
任意格要素：バスで
前置詞句候補：なし
出力文：Students go to school
正解文：A student goes to school by bus.

5.2 実験結果

実験を行った結果を表 5.1, 5.2 に示す。まず，表 5.1 より，任意格要素の翻訳が必要な入力文は，47 件であった。任意格要素の翻訳が必要でない場合の例を図 5.1 に示す。この場合では，入力文のすべての要素に対してパターンが適合しているので，任意格要素の翻訳が必要でなかったことがわかる。また，任意格要素に対して，正解の表現で翻訳できた例を図 5.2 に示す。

入力文:彼は魚を釣針から外した。
適合パターン情報:
N1 は N2 を N3 から 外す
N1 take N2 off N3

図 5.1: 任意格要素の翻訳が必要でない例

次に，任意格要素翻訳必要文 47 件に対して評価を行った結果が表 5.2 である。また，出力の評価を付録 C に記載する。正解の前置詞句パターンが翻訳候補に含まれたものは，と を合計して，37 件である。これは，任意格要素に対する格要素パターン辞書の意味的なカバー率が，79% (37/47) であると解釈できる。一方，× が 0 件であることについては，日本語の表現レベルでのカバー率が，100%であると解釈できる。したがって，格要素パターン辞書は，基本的には成功理に作成できたと言える。

入力文：地震で壁が崩れ落ちた。

適合パターン情報:

N_1 が 崩れる

N_1 crumble

N_1 (2 具体 1001 抽象物 1236 人間活動 2499 力・能力等)

任意格要素:地震で

前置詞句の候補と新英文パターンによる翻訳時の文の N -gram の値:

In N_k : -59.719

With N_k : -60.427

During N_k : -63.973

of N_k : -68.746

with N_k : -70.529

for N_k : -70.795

like N_k : -70.929

in N_k : -71.040

but N_k : -72.337

出力された翻訳文:In an earthquake a wall crumble

正解文:The wall crumbled to the ground in the earthquake .

図 5.2: 正解の表現で翻訳できた例

表 5.1: 入力文における任意格要素翻訳必要文数

全入力文数	任意格要素翻訳必要文数
100	47

表 5.2: 5.1.2 節に基づく出力文の評価の内訳

			×
16 件	21 件	10 件	0 件

第6章 考察

任意格要素翻訳必要文 47 件に対して、 は 16 件であることから、正解率は 34%である。これは、前置詞句の翻訳候補の選択能力が原因である。すなわち、格要素パターン辞書の運用上の問題である。運用をよりよくするために、 と の事例を分析する。

6.1 事例の分析

6.1.1 出力文の候補において正解の前置詞が存在したもの（ の場合）

の例を図 6.1 に示す。前置詞句パターンの候補には、正解文の前置詞「with」が含まれているが、翻訳時の文の tri-gram が「in」や「of」より低いことから、正しく前置詞句パターンが選ばれていなかった。これは、「in」や「of」が前置詞句の候補としてそもそも正しいのか、 N -gram がスコア付けとして正しいものだったか、ということが考えられる。

6.1.2 正解の前置詞句パターンが出力文候補に含まれていないもの（ の場合）

の例を図 6.2 に示す。補うべき前置詞句の候補に正しい前置詞句パターンがなかった。単に、格要素パターン辞書の用例が不足していたことが原因として考えられるが、一方で、候補の過度の削減も原因として考えられる。

入力文：彼は優秀な成績で大学を卒業しました。

適合パターン情報:

N_1 が N_2 を 卒業する

N_1 graduate from N_2

N_1 (4人) N_2 (362組織 1002抽象物(精神) 1237精神)

任意格要素:優秀な成績で

前置詞句の候補と新英文パターンによる翻訳時の文の N -gram の値:

in N_k : -58.809

of N_k : -60.119

with N_k : -60.555

for N_k : -60.691

on N_k : -61.087

but N_k : -63.332

like N_k : -73.121

at N_k : -73.390

With N_k : -76.320

出力された翻訳文:He graduate from university in excellent results

正解文:He graduated from college with excellent records .

図 6.1: 正解格要素パターンが選ばれなかった例

入力文：ノートが消しゴムで破れた。

適合パターン情報:

N_1 が 破れる

N_1 be torn

N_1 (533 具体物)

任意格要素:消しゴムで

前置詞句の候補と新英文パターンによる翻訳時の文の N -gram の値:

With N_k : -75.260

with N_k : -81.118

for N_k : -81.384

like N_k : -81.517

in N_k : -81.629

but N_k : -82.926

出力された翻訳文:With an eraser notebook be torn

正解文:A page of the notebook was torn by an eraser .

図 6.2: 正解前置詞句が出力文候補に含まれていない例

6.2 今後の対策

6.2.1 第 6.1.1 節に関する対策

第 6.1.1 節の の事例の、任意格要素に対する翻訳候補として挙げた前置詞句が正しいのかという点に関して、日本語の用言に共起することが適当な任意格要素があると考えた。例えば、「バスで学校に行く」については、「行く」という用言に対して「バスで」という格要素が伴うこと適当である。しかし、「バスで学校に連絡する」という表現では、「バスで」と「連絡する」の意味のつながりが不自然である。このように考えると格要素パターンには、用言意味属性制約が必要である考える。また、 N -gram がスコア付けとして適当でなかったと考えた点に関しては、日本語の任意格要素と用言の関係が英語の表現において同じように言えるからである。ゆえに、翻訳候補に対して、動詞、前置詞、および、それに伴う名詞の共起確率でのスコア付けが必要があると考える。以下に、それらの方法について考える。

格要素パターンの用言意味属性制約

まず、任意格要素と格要素パターン辞書の格要素パターンと照合する時に、名詞変数の意味属性だけでなく、格要素パターンに共起した用言意味属性も制約に加えることを考える。例えば、第 3.4 節において、図 3.3 の情報から格要素パターン対を抽出する際、用言の情報を同時に取得すると、図 3.6 のような情報が得られる。この情報を用いることによって、入力文の任意格要素に対して、より適当なものだけが適合すると考えられる。すなわち、格要素パターンには、動詞変数 V 、および、 V に対する制約を新たに追加する。

格要素パターン	前置詞句パターン
/N1(OR:列車,NI:988,...) で/V2(OR:行く,NY:0506,...)	by N1(OR:train)

図 6.3: 新規格要素パターン対の例

動詞、前置詞、および、それに伴う名詞の共起確率によるスコア付け
入力文に対する翻訳候補に対して、以下のような共起確率を取る。

翻訳候補: He graduate from university in excellent results
共起確率を取る部分: graduate, in, results

この方法で共起確率をとれば、動詞対して共起するのが適当な前置詞である翻訳文のスコアが上昇することによって、より正解の表現で出力できると考える。

6.2.2 第6.1.2節に関する対策

正解の前置詞句パターンが、任意格要素に対する翻訳候補に挙げられなかったことに関して、変数の意味属性制約による候補の過度の削減があったのではないかと考えられる。今回、鳥バンクより格要素パターン対の抽出を行ったが、抽出の元となったパターンの変数の意味属性は、その例文の用例に基づくものである。したがって、用例レベルでの厳しい制約となる。しかし、日本語格要素と英語前置句が同じパターンで複数存在する場合、それぞれの変数の意味属性が異なっても、兄弟関係にある属性であるならば、親の属性まで緩和してもよい場合がある。こうすることによって、用例レベル以上の範囲の制約で、任意格要素と格要素パターンとの適合ができる。この方法は、「*A*の*B*」型名詞句の翻訳に有効であった [8] ことから、試みる価値がある。

第7章 おわりに

本研究では，非線形言語モデルに基づく日英機械翻訳における任意の（必須でない）格要素の翻訳のための格要素パターン辞書の構築を行ない，その翻訳方法を示した．まず，格要素と前置詞句の対の収集を約 22 万件の重文複文文型パターン辞書から行った結果，用例レベルでの格要素パターン対を 10,783 件，文法レベルでの格要素パターン対を 925 件が得られた．同時に，格助詞相当句を 266 件が得られた．次に，格要素パターン辞書を用いて単文の日英機械翻訳を実験したところ，格要素パターン辞書の任意格要素に対する意味的なカバー率が 79%，および，日本語の表現レベルでのカバー率が 100%という結果を得た．したがって，格要素パターン辞書は，基本的には成功裡に作成できたと言える．今後の課題は，格要素パターン辞書の運用の方法を精緻することで翻訳性能を高めること，一方で，任意格要素に対して意味分類を付与し，意味解析に役立てることである．

謝辞

本研究を進めるに当たり，種々の御助言を頂きました池原悟教授，菅原一孔教授，および，村上仁一准教授に心から御礼申し上げます。

また，徳久雅人講師には，終始に渡り研究の進め方や本論文の書き方など，細部に渡る御指導を頂きました。ここに深く感謝いたします。

その他様々な場面で御助力をいただいた計算機工学講座 C 池原研究室の皆様には感謝の意を表します。

参考文献

- [1] 池原悟: 非線形言語モデルによる自然言語処理, 岩波書店, 2009 .
- [2] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語語彙大系, 岩波書店, 1997.
- [3] 池原悟: 鳥バンク, 日本語表現意味辞書 - 重文複文編 -, 2007.
<http://unicorn.iike.tottori-u.ac.jp/toribank>
- [4] 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.
- [5] Yamamoto, K., Matsumoto, Y.: Extracting Translation Knowledge from Paralel Corpora, M.Carl and A.Way(eds.), Recent Advances in Example-based Machine Translation, pp.365-395, Kluwer, 2003.
- [6] Och, F.J, Ney, H.: A Systematic Comparison of Various Statistical Alignment Models, Computational Linguistics, Vol.29, No.1, pp.19-51, 2003.
- [7] 徳久雅人, 村上仁一, 池原悟: 重文・複文文型パターン辞書からの構造照合型文型パターン検索, 情報処理学会研究報告, 自然言語処理, 2006-NL-176, pp.9-16, 2006.
- [8] 徳久雅人, 守谷有司, 村上仁一, 池原悟: 意味属性の共起による「AのB」型名詞句の翻訳規則, 情報技術フォーラム, 情報技術レターズ, vol.2, pp.87-88, 2003.
- [9] 益岡隆志, 田窪行則: 基礎日本語文法, くろしお出版, 1989.

付録 A 扱う前置詞と前置詞相当句

付録B 格要素パターン辞書

付録C 出力文の評価

付録D 学外発表