

概要

日本語の単文の解析に日本語語彙大系の結合価パターンが使用されている。このパターンは命題表現をカバーするためのものであり、基本的には能動態で定義されている。そのため、受動態や使役などの構文に対する格助詞の変化には対応していないという問題がある。そこで、本研究では、日本語語彙大系における結合価パターンを能動態から受動態へと変換する方法を示す。

寺村によると、日本語では能動態から受動態へと変換する際、動詞ごとに助詞の変化内容が異なるが、ある程度の規則性があると言われている。そこで、大局的に能動態から受動態へと変換する規則を日本語語彙大系の結合価パターンに適用して受動態のパターンを作成し、局所的に動詞ごとの差異を吸収するためにパターンを修正するという方針でこの問題に取り組む。

本研究ではまず、寺村・庭により示された変換規則9件を適用して、11,310件の結合価パターンを受動態のパターンに変換した。次に、寺村が考察した受動態の例文に対して受動パターンで意味解析を行った。その結果、変換規則に助詞の不足があったため、変換規則に改良を加え受動パターンを大局的に作り直した。

評価実験において、受動パターンの網羅性を検証するため、学習辞書などで使用された単文の受動文に対して受動パターンで意味解析を行った。結果は、受動パターンのカバー率は約5割となった。誤り分析によると、動作主、対象物、あるいは手段を表すための多様な格助詞を網羅するという問題、および、能動態では必須ではなくとも受動態では必須となる格要素を追加するという問題があるとわかった。

目次

第1章	はじめに	1
第2章	関連研究	2
2.1	日本語語彙大系	2
2.2	日本語の受動態	3
第3章	受動パターンの作成	6
3.1	能動パターンと受動パターン	6
3.2	受動パターンの作成手順	6
3.3	受動パターンの作成例	7
3.4	受動パターンの作成結果	7
第4章	受動文の意味解析	8
4.1	パターンを用いた意味解析	8
4.2	意味解析の手順	8
4.3	解析の様子	9
第5章	予備実験	11
5.1	評価方法	11
5.2	入力文	12
5.3	実験結果	12
5.4	誤り分析	12
5.4.1	分類1「形態素解析の誤り」	13
5.4.2	分類2「受動パターンの助詞が不足」	13
5.4.3	分類3「格要素の意味属性制約が満たされていない」	14
5.4.4	分類4「格要素の個数に変化を要する」	15

第 6 章	変換規則の改良	17
6.1	改良の方針	17
6.2	改良を行った変換規則	18
6.3	改良の効果	18
第 7 章	評価実験	19
7.1	評価方法	19
7.1.1	カバー率	19
7.1.2	精度	19
7.2	入力文	20
7.3	実験結果	20
7.4	誤り分析	21
7.4.1	分類 1 「入力文の形態素解析の誤り」	21
7.4.2	分類 2 「適切な用言受動パターンの作成不足」	22
7.4.3	分類 5 「複数の候補からの選択」	25
7.4.4	分類 7 「文の曖昧性」	25
7.4.5	分類 8 日本語語彙大系の網羅性	26
第 8 章	おわりに	27

目 次

2.1	結合価パターンの例	3
4.1	形態素解析結果	9
4.2	パターンの検索結果	9
7.1	入力文の形態素解析結果	22

表 目 次

2.1	変換規則集	5
5.1	予備実験における変換失敗の内訳	12
6.1	改良後の変換規則集	18
6.2	改良後の変換失敗の内訳	18
7.1	評価実験における変換失敗の内訳	21

第1章 はじめに

文型パターンを用いた言語表現の解析において、パターンの網羅性を高めるために、パターンで使用される格助詞や様相表現に自由度を持たせる必要がある [1]。とりわけ、受動態と使役への対策には格助詞の変形が伴うため、単なる過去形や推量の表現に対応するよりも対策が複雑である。

ここで、日本語の単文についての代表的な文型パターン辞書として日本語語彙大系が挙げられる [2]。これは多段翻訳方式において日本語の命題表現をカバーするためのものであり、原則としてパターンは能動態で記述されている。そのため、受動態の文に直接的には対応しておらず、パターン外の処理において対処する必要がある。ところが、受動態に対処する方法は、[2] において具体的に示されておらず、この辞書の利用者が各自で対処しなければならないという問題がある。

受動文を結合価パターンと適合させる方法として、受動文の格助詞に対応した結合価パターンを用意する方法と、パターン照合の際に入力文を書き換えて結合価パターンの格助詞と対応させる方法の2つがある。

ここで、寺村の考察 [3] を参照すると、『日本語において、格の移動により能動文と受動文が対応する以上、一定の規則で表すことのできる変化であるが、動詞の性質や名詞の性質によって規則の適用される範囲は限定的である』と述べている。この考察は、動詞ごとに格助詞変化があるのだが、ある程度の規則性があると解釈できる。つまり、動詞ごとに受動態のパターンを用意するべきであるが、その多くは規則的に能動態のパターンから作成できるかもしれないといえる。そこで、本研究では、受動文の格助詞に対応した結合価パターンを能動態のパターンごとに作成するという前者の方法を試みる。

第2章 関連研究

本章では，本研究において使用する日本語語彙大系，および，受動態の一般論を解説する．

2.1 日本語語彙大系

日本語語彙大系 [2] には，日本語の用言（約 6,000 語）の結合価パターン（約 14,000 件）が定義されている．結合価パターンは，用言と格要素の意味的共起関係を文型化したものである．結合価パターンの格要素は，一般名詞意味属性により制限を設けられた名詞と格助詞を組み合わせて構成している．格要素の使い方により用言の意味に違いが生じるので，一つの用言に対し複数の結合価パターンが存在している．そのため，格要素を構成する名詞は，変数化されており，一般名詞意味属性による制約条件が付与されている．

一般名詞意味属性は，単語を意味的用法により階層的に分類，体系化したものである．約 40 万の単語を 12 段の木構造により 2,710 の意味属性に分類している．一般名詞意味属性では，上位の意味属性の性質を下位の意味属性に継承している．したがって，制約条件として中位の意味属性が与えられている場合でも，その意味属性の下位にあたる名詞は制約を満たす．

一方，結合価パターンは，用言としての意味的用法が用言意味属性により示されている．用言意味属性は「用言が持つ動的属性の種類」の観点から分類した 36 種類の属性体系である．このことから，パターンによる意味解析ができるといえる．

例えば，用言「読む」を検索すると，図 2.1 に示される結合価パターンが得られる．以下の結合価パターン (1) と (2) は文法レベルでは同じ記述であるが，名詞変数 $N2$ には意味属性制約により異なる名詞が適合する．その結果，用言に対する意味が異なることを表している．

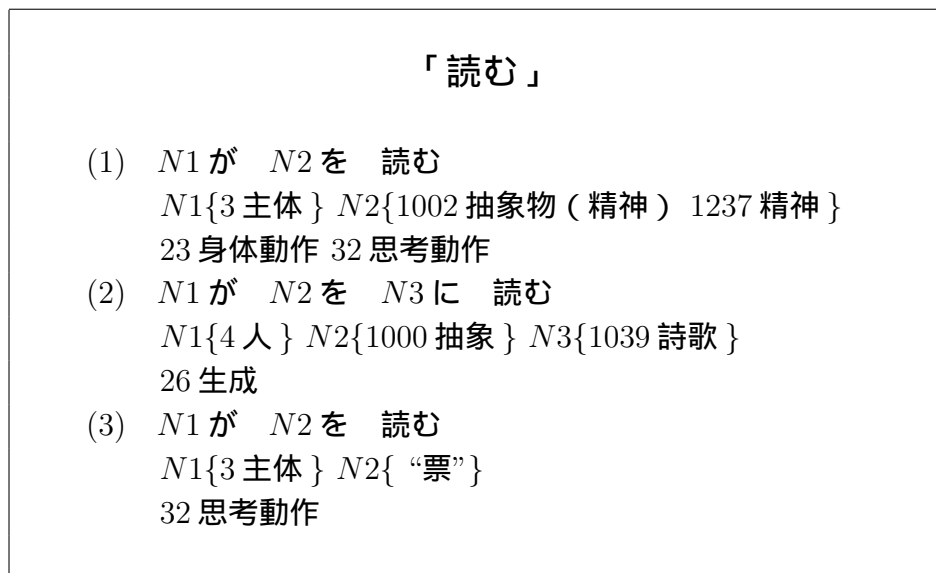


図 2.1: 結合価パターンの例

2.2 日本語の受動態

寺村の考察 [3] を受け，庭は，格助詞「ガ，ニ，ヲ」を中心に，能動態から受動態への変形規則を述べた [4]．[4] は受動態を「直接受身」と「 N の受身」と「間接受身」の3つに分類し，能動文と受動文の対応を規則的に述べている．以下に解説する．

直接受身

「直接受身」は，動作を受ける対象の立場にたって，その動作を受けることを表す．基本は「 N が」が人や動物など有情の主体をとる場合を基本とし，5つの変形パターンがある．

a : A が B を V する B が A に V される

猫が魚を食べる 魚が猫に食べられる

動詞の対象が「 B を」で示されるような動詞での受身である．受身形での動作の主体は通常「 A に」で表される．

b : A が B に V する B が A に V される

犬が私にかみつ 私が犬にかみつかる

aにおける「Bを」が「Bに」となっている場合について、動詞は通常、自動詞とされるが、直接的な対象を取るものはそのまま直接の受身となる。

c: AがBにCをVする BがAにCをVされる
友達が私に仕事を頼む 私が友達に仕事を頼まれる

補語が「に・を」の型で、対象の「Bに」が受身文において「Bが」となる。このパターンでは受身での「Aに」の代わりとして、「Aから」も用いることができる。能動文の「Bに」が「到達点」のような意味合いを持っており、Bに対する「出発点」としての「Aから」となる。

d: AがBにCをVする CがAからBにVされる
太郎が皆に花子を紹介する 花子が太郎から皆に紹介される。

能動文における「BにCを」の「Cを」が、受身文の「Cが」に変化するものである。「Bに」が変化しないため、「Aが」が分かりにくくなるのを避け、「Aに」ではなく「Aから」になる。口語表現では「Aから」「Aに」が省略されることが多い。

e: AがBをCとVする BがAにCとVされる
太郎が花子の行為を悪影響と見なす 花子の行為が太郎に悪影響とみなされる。

能動文の「BをCと」の補語をとる動詞となっているものである。また、Bが省略される場合もある

例) 太郎が花子と絶交する 花子が太郎に絶交される

Nの受身

直接受身では、受身の「Nが」となるものが能動文の補語であったが、そうではない場合の受身として、「Nの受身」がある。「Nの受身」は、受動態では動作の対象となる補語を「Nの」の形で修飾している要素が、受身で「Nが」になるものとしている。

f: AがBのCをVする BがAにCをVされる
誰かが私の足を踏んだ 私は誰かに足を踏まれた

間接受身

「間接受身」は元の文には、受身文の「Nが」が直接出ていないものとされている。何らかの被害・影響を受ける立場にある。そして、この受身文の特徴は、必ず「迷惑」を

受けることを表す．以下に例を示す．

1. 子どもたちは、学校の帰りに雨に降られた。(雨が降る)
彼女は若くして夫に死なれた。(夫が死ぬ)
2. 弟にケーキを全部食べられてしまった。(弟がケーキを食べた)
バスで隣の人に窓を開けられ、髪がばさばさになる。(隣の人が窓を開ける)

1の例は自動詞の受身で、英語では存在しない日本語特有の間接受身に含まれる．

2の例は他動詞の受身で、迷惑の受身となるものである．間接受身は他動詞でも成り立つ．「ケーキを食べられた」の例では、「ケーキ」は「食べる」の直接の対象となるが、「ケーキを食べる」という行為は、「他者に対する行為とは言えず、間接受身に分類されるものとしている．間接受身の最大の特徴として、能動文の「*N*が」は「*N*に」に必ず変化し、それ以外の形をとらない．

ここで、寺村の考察による態の対応と、これを受けた庭の能動文と受動文の規則的な対応を表 2.1 にまとめる．表中の「*A*, *B*, *C*」は格要素の名詞変数、「*V*」は述語となる動詞変数、「*.reru*」は受動表現の接辞を意味している．

本研究では、能動文に相当する記述を条件パターンとし、受動文に相当する記述を帰結パターンとする．変換規則は1つの条件パターンと1つの帰結パターンの組である．能動文に条件パターンが適合するとき、帰結パターンを用いて受動文を作ることができる．さらに結合価パターンに条件パターンが適合するとき、文の場合と同様に受動態のパターンが作成できると考えられる．

表 2.1: 変換規則集

#	条件パターン	帰結パターン	典拠
1	<i>A</i> が <i>V</i>	<i>A</i> に <i>V.reru</i>	寺村, 庭
2	<i>A</i> が <i>B</i> と <i>V</i>	<i>A</i> に <i>B</i> が <i>V.reru</i>	寺村, 庭
3	<i>A</i> が <i>B</i> に <i>C</i> を <i>V</i>	<i>A</i> に/から <i>B</i> が/に <i>C</i> を/が <i>V.reru</i>	寺村, 庭
4	<i>A</i> が <i>B</i> を <i>C</i> に <i>V</i>	<i>A</i> に/から <i>B</i> を/が <i>C</i> が/に <i>V.reru</i>	寺村, 庭
5	<i>A</i> が <i>B</i> に <i>V</i>	<i>A</i> に <i>B</i> が <i>V.reru</i>	寺村, 庭
6	<i>A</i> が <i>B</i> の <i>C</i> を <i>V</i>	<i>A</i> に <i>B</i> が <i>C</i> を <i>V.reru</i>	寺村, 庭
7	<i>A</i> が <i>B</i> を <i>C</i> と <i>V</i>	<i>A</i> に <i>B</i> が <i>C</i> と <i>V.reru</i>	寺村, 庭
8	<i>A</i> が <i>B</i> を <i>V</i>	<i>A</i> に <i>B</i> が/を <i>V.reru</i>	寺村, 庭
9	<i>A</i> が <i>B</i> を <i>C</i> から <i>V</i>	<i>A</i> に <i>B</i> が/を <i>C</i> から <i>V.reru</i>	寺村

第3章 受動パターンの作成

本章では，寺村や庭の考察より得られた変換規則（表 2.1）を日本語語彙大系の結合価パターンに適用し，受動文に対応する結合価パターンの作成を行う．

3.1 能動パターンと受動パターン

能動態で記述してある文を能動文と呼ぶことから，本研究では，能動態で記述してある結合価パターンを能動パターン，同様に，受動態で記述してある結合価パターンを受動パターンと呼ぶことにする．

3.2 受動パターンの作成手順

以下の手順により，能動パターンから受動パターンを作成する．

- 手順 1

能動パターンと表 2.1 の条件パターンを照合する．ここでは，格助詞の並びが最も多く適合する条件パターンを選択する．

- 手順 2

選択された条件パターンと対になっている帰結パターンを参照する．これに従い，能動パターンの格助詞のうち，条件パターンと適合した部分を変換する．

- 手順 3

変換したパターンの述部に，受動態の接辞を意味する「*.reru*」を付与し受動パターンとして出力する．

なお，上記の手順で作成される受動パターンは能動パターンと意味属性制約が対応している．

3.3 受動パターンの作成例

受動パターンの作成例を示す。下記の能動パターンは、一見すると「AがV」「AがBにV」「AがBにCをV」の3つの条件パターンに適合しうる。ここでは、3つの条件パターンの中で要素の並びが最も多く適合する「AがBにCをV」という条件パターンにより、受動パターンが作成されている。

能動パターン：N1がN2にN3を伝える

受動パターン：N1から/にN2に/がN3を/が'伝える'.*reru*

意味属性制約：N1(3主体) N2(3主体) N3(“悲報/訃報”)

また、名詞変数の直後に記述されている助詞のみを、条件パターンと照合させる。そのため、以下の能動パターンは、名詞変数N3の格助詞は「に」となる。そのため条件パターン「AがBをCにV」が適合し、帰結パターンにより受動パターンが作成される。

能動パターン：N1がN2をN3に/から/より教わる

受動パターン：N1から/にN2が/をN3に/が/から/より'教わる'.*reru*

意味属性制約：N1(4人) N2(1001抽象物 1236人間活動) N3(4人)

一方で、能動パターンの格要素が条件パターンの格要素よりも多い場合、適合部分のみから帰結パターンに従い格助詞を変換し、受動パターンを作成する。例えば、以下に示す能動パターンでは、格助詞が「が」「を」「から/より」「に/へ/まで」の並びとなっている。この場合は、「AがBをCからV」に対する条件パターンが該当し、該当部のみ格助詞の変換が行われる。

能動パターン：N1がN2をN3から/よりN4に/へ/まで移転する

受動パターン：N1にNが/をN3から/よりN4に/へ/まで'移転する'.*reru*

意味属性制約：N1(3主体) N2(1165権利・義務 1168制度(経済)) N3(3主体)

3.4 受動パターンの作成結果

日本語語彙大系における動詞の能動パターン 12,297 件に対して、表 2.1 を使用した結果、11,310 件の受動パターンが作成できた。

第4章 受動文の意味解析

本章では、受動パターンを使用した受動文の意味解析を説明する。

4.1 パターンを用いた意味解析

言語表現の処理における意味解析とは、言語表現に社会的約束として対応する概念を求めることである [5]。パターンを用いた意味解析では、パターンと言語表現が適合により、言語表現に含まれる概念を捉えるということを表す。

また、言語表現の処理の応用として翻訳を行う場合、パターンに翻訳用の情報を付与しておくことで工学的な処理を行うことができる。

そこで、本研究では、受動文を機械的に処理するために受動パターンを用いた意味解析を行う。受動パターンでの意味解析により、受動文の言語処理が可能となり、受動文から能動文を生成するといった処理ができる。

受動パターンは、能動パターンより作成されているため、能動パターンに付与されている情報を用いた言語処理が可能である。

4.2 意味解析の手順

受動文の意味解析は以下の手順により行われる。

- 手順 1
入力文（受動文）を形態素解析し、単語に一般名詞意味属性を付与
- 手順 2
文法レベルで入力文と受動パターンを照合
- 手順 3
意味属性制約を充足する受動パターンを選択

上記の手順3により意味解析は完了する。

また、意味解析により、受動パターンと対応する能動パターンを使用した能動文の生成が可能となる。入力文と生成された能動文を観測する際に、入力文とパターンの適合具合の観測が容易に行える。

4.3 解析の様子

受動文「太郎に花子が愛される。」の意味解析の様子を示す。

まず、形態素解析によって入力文の単語の品詞を解析し、名詞に一般名詞意味属性番号を付与する。以下に示す図4.1では、形態素解析の結果、名詞「太郎」と「花子」は、「人」という名詞意味属性における下位の「男」と「女」を表す意味属性番号がそれぞれ付与されている。

-
1. /太郎 (1900,NI:48,NI:464,NK:69,NK:27)
 2. +に (7430)
 3. /花子 (1900,NI:49,NI:5,NK:70,NK:67)
 4. +が (7410)
 5. /愛さ (2432,愛する,愛さ,NY:31,KR:0600h00,IY:1311)
 6. +れる (7116)
 7. +。(0110)
 8. /nil
-

図 4.1: 形態素解析結果

次に、受動パターンと、入力文の名詞「太郎」、「花子」の格助詞、そして、述部を照合する。今回、入力文に対して2件のパターン(図4.2)が照合された。

-
- パターン番号：9
受動パターン：N1にN2が/を'愛する'.*reru*
意味属性制約：N1(4人)N2(*)
用言意味属性：31感情動作
-
- パターン番号：10
受動パターン：N1にN2が/を'愛する'.*reru*
意味属性制約：N1(4人)N2(72人間 親族関係)
用言意味属性：31感情動作
-

図 4.2: パターンの検索結果

そして入力文の名詞がパターンと適合するかを意味属性制約により判断する。名詞「太郎」は、パターン番号 9, 10 の名詞変数 $N1$ と適合する。一方、名詞「花子」はパターン番号 10 における意味属性制約 (72 人間 親族関係) を充足していない。しかし、パターン番号 9 の意味属性制約を充足している。よって入力文は 2 つの名詞変数の制約を充足するパターン番号 9 の受動パターンと適合し、感情動作を表す文であると解析される。

変換規則により作成された受動パターンは、能動パターンと対になっている。

受動パターン： $N1$ に/によって/から $N2$ が/を '愛する'.*reru*
能動パターン： $N1$ が $N2$ を 愛する

受動文の意味解析の結果、適合した受動パターンと対になる能動パターンにより、能動文の生成ができる。

入力文：花子が太郎に愛される。
出力文：太郎が花子を愛する。

第5章 予備実験

能動文と受動文が、格の移動で対応することから、受動文と能動文の対応を評価することで、変換規則により作成された受動パターンの正しさを判断することができると思われる。そこで、寺村の受動態についての考察 [3] に使用された受動文の例文を抽出し、受動パターンを使用した解析を行う。そして、変換規則による受動パターンが正しく作成されているか検証する。ただし、変換規則の正しさを調べるのが目的であるので、第4章における受動文の解析手順3での複数候補の選択の性能は問わないことにする。

5.1 評価方法

受動パターンと受動文が意味的に正しく適合しているかの判定を行う。本実験では、1文につき複数のパターンが適合するので、受動文を能動文に復元することでパターンの適合を観測する。これは、受動文と能動文を復元することでパターン適合の観測が容易となるためである。そのため、第4章で述べた処理方法を用いて、入力文を能動態に復元した結果を以下の通り評価する。

- 受動パターンが入力文の格要素を満たしているか評価
- 能動文に変換し、意味的に正確であるかを評価

また、1つの入力文に対して、複数の候補が得られるが、そのうち1つでも、上記の2つの評価を満たした受動パターンがあれば、変換規則は正しく適用されたと判断する。2つの評価を満たす例を以下に示す。

入力文：その国は国王によって治められた。

出力文：国王がその国を治める。

受動パターン：N1に/によってN2が/を'治める'.*reru*

能動パターン：N1がN2を治める

意味属性制約：N1(3主体)N2(385国家 154国民 388場所)

5.2 入力文

本実験では，寺村の受動態についての考察 [3] に使用されている受動態の例文 57 文を入力文として使用する．[3] では例文がカタカナで表記されている場合があるが，形態素解析を正しく行うため，ひらがな表記に直して入力文として使用する．

彼八刑事二怪シマレタ。 彼は刑事に怪しまれた。

また，基本的な受動文を得るため，受動文の文末を「れる」、「られる」のみに加工して実験に使用する．

直孝は祖母に育てられた。 直孝は祖母に育てられる。
多くの人に愛されている。 多くの人に愛される。

5.3 実験結果

入力文 57 文に対し，37 文が受動文から能動文に復元することができ，カバー率は 65%(37 文/57 文) となった．本実験では，変換規則で作成した受動パターンは，格助詞「が・に・を」を使用した受動文を正しく解析できた．また，受動パターンの不備を検証するため，誤りがあった事例を表 5.1 に示す．

表 5.1: 予備実験における変換失敗の内訳

分類	原因	頻度
1	形態素解析の誤り	1
2	受動パターンの助詞が不足	15
3	格要素の意味属性制約	3
4	格要素の個数に変化	1

5.4 誤り分析

受動パターンの作成における不備を検出するため，事例ごとに誤り分析を行う．

5.4.1 分類1「形態素解析の誤り」

以下に誤り事例を示す。

入力文：日本の領空は外国機に飛ばれる。

出力文：外国機に [が日本の領空を から/より へ/に/まで 飛ぶ]。

正解文：外国機が日本の領空を飛ぶ。

出力文の括弧内は、受動パターンが適合した範囲を表す。「 ϕ 」は結合価パターンの格要素に名詞が照合されていないことを表す。

入力文の「外国機」は、能動文でのガ格の名詞として解析されるべきである。しかし、実際の出力ではそのようになっていない。そこで使用したパターンを参照すると、次の通りであった。

能動パターン：N1 が N2 を N3 から/より N4 へ/に/まで 飛ぶ

受動パターン：N1 に N2 が N3 から/より N4 へ/に/まで '飛ぶ'.*reru*

意味属性制約：N1(3 主体 536 動物 (個体)990 乗り物 (本体 (移動 (空圏))))

N2(388 場所 2610 場)N3(388 場所 2610 場)N4(388 場所 2610 場)

上記に示される助詞の対応では、入力文の「外国機」は受動パターンの変数 N1 または N4 が選択候補となり、意味属性制約により変数 N1 の名詞として適合されるべきである。しかし、名詞の形態素解析の結果を参照したところ、次の通りとなっていた。

/外国 (1100,NI:385,NI:463) +機 (6280,NI:2595,NI:2588)

「機」は接尾辞として解析されており、一般意味属性番号の 2595, 2588 が付与された。これらは「単位」「回数・回数等」を意味する。したがって、「飛行機」という意味で解析されておらず、形態素解析における名詞意味属性の付与誤りであることがわかった。これは変換規則では対応できない。

5.4.2 分類2「受動パターンの助詞が不足」

以下に誤り事例を示す。

(2a) 入力文：彼は刑事から怪しまれる。

出力文：刑事から [が彼を怪しむ]。

正解文：刑事が彼を怪しむ。

以下に使用されたパターンを示す。

能動パターン：N1がN2を怪しむ
受動パターン：N1にN2が/を'怪しむ'.*reru*
意味属性制約：N1(4人)N2(*)

使用された受動パターンの変数 N1 と N2 の助詞には、助詞「から」がどちらにも存在しない。そのため、名詞「刑事」が受動パターンの名詞変数と照合されなかった。

入力文が受動パターンと正しく対応するためには、「N1に」ではなく「N1から」が受動パターンに必要である。同様に誤り事例を示す。

- (2b) 入力文：金閣は義満によって建てられる。
出力文：義満によって [が 金閣を に/へ 建てる]。
正解文：義満が金閣を建てる。

以下に使用したパターンを示す。

能動パターン：N1がN2をN3に/へ建てる
受動パターン：N1から/にN2が/をN3が/に/へ'建てる'.*reru*
意味属性制約：N1(3主体)N2(389施設938像・書画863建造物928碑)
N3(388場所2610場)

(2a)と同様に、受動パターンの各変数の助詞に、複合助詞「によって」が存在しない。そのため、名詞「義満」が受動パターンの名詞変数と照合されなかった。入力文が受動パターンと正しく対応するためには、「N1に」ではなく「N1によって」が受動パターンに必要である。

以上(2a)と(2b)により表1の変換規則に改良するべき点が示された。

5.4.3 分類3「格要素の意味属性制約が満たされていない」

以下に誤り事例を示す。

- 入力文：家の流儀が当主から次に当主となる子に受け継がれる。
出力文：家の流儀が [次に当主となる子が を 当主から/より 受け継ぐ]。
正解文：次に当主となる子が家の流儀を当主から受け継ぐ。

以下に使用したパターンを示す。

能動パターン：N1がN2をN3から/より受け継ぐ
受動パターン：N1にN2が/をN3から/より'受け継ぐ'.*reru*
意味属性制約：N1(3主体)N2(362組織1936仕事)N3(3主体)

入力文を意味的に正しく能動文へ変換するためには、名詞句「家の流儀」が受動パターンの $N2$ に適合しなければならない。ここで、「家の流儀」の形態素解析結果を参照すると次の通りであった。

/家 (1100,NI:447,NI:865,NI:73,NI:387) +の (7410) /流儀 (1100,NI:1222,NI:1035)

名詞「流儀」の意味属性番号 1222, 1035 は、「流儀」、「方法」を意味する名詞として正しく解析されている。しかし、受動パターンの変数 $N2$ は意味属性制約により「流儀」を選択することができなかった。これは表 2.1 の規則の問題ではなく、日本語語彙大系における問題である。なお、日本語語彙大系の一般名詞意味属性に対する制約条件の変更は、本研究では対象としておらず、対策を設けない。

5.4.4 分類 4 「格要素の個数に変化を要する」

以下に誤り事例を示す。

入力文：アーサー王子が両親をラビック王に殺される。
出力文：アーサー王子が [ラビック王が 両親を 殺す]。
正解文：ラビック王がアーサー王子の両親を殺す。

ここで使用したパターンは次の通りである。

能動パターン： $N1$ が $N2$ を 殺す
受動パターン： $N1$ に $N2$ が / を '殺す'. <i>reru</i>
意味属性制約： $N1$ (4人 535 動物) $N2$ (4人 535 動物)

入力文の格要素として考えられる名詞は「アーサー王子」、「(アーサー王子の) 両親」、「ラビック王」の 3 つである。しかし、動詞「殺す」の全ての能動パターンは、2 つの格要素しか選択できない。そのため、入力文の格要素を、能動文として意味的に正しく変換できなかった。

この格要素の個数に変化を要する問題は、庭によって述べられた「 N の受身」と対応していると考えられる。能動文で動作の対象となる補語を「 N の」の形で修飾している要素が、受身文の「 N が」になる型のものである。庭が述べた規則では、格要素が 3 つの文に対して適用されているが、パターンに定義されている格要素が 3 つ未満の場合と対応できていなかった。

入力文を正しく能動文に変換するならば，日本語語彙大系に以下の結合価パターンが必要になる．なお，名詞変数 $N3$ を囲む「 $\#n$ 」は，任意のパターン記述要素を表す記号である．

能動パターン： $N1$ が $\#1[N3$ の] $N2$ を 殺す
受動パターン 1： $N1$ に $N3$ が $N2$ を 殺す'.*reru*
受動パターン 2： $N1$ に $N3$ の $N2$ が 殺す'.*reru*
意味属性制約： $N1(4$ 人 535 動物) $N2(4$ 人 535 動物) $N3(*)$

しかし，能動パターンの変形までは本研究の対象としないため，本研究ではこの誤り事例に対策を設けない¹．

¹この事例より類推すると次のような受動文が考えられる：「この象は鼻が寝められた。」，「この象は鼻を寝められた。」

第6章 変換規則の改良

本章では、予備実験において受動パターンがカバーできていなかった「から」、または、「によって」に対応させるため、変換規則の改良を行う。

6.1 改良の方針

- 助詞「から」

表 2.1 の変換規則で作成した受動パターンは、「から」が含まれる受動文と適合させる場合、以下のパターンでしか対応できていなかった。

能動パターン：A が B に C を V
受動パターン：A に / から B が / に C を / が V.reru
能動パターン：A が B を C から V
受動パターン：A に B が / を C から V.reru

そこで、他の受動パターンに対して「から」を付与するため、次の基本的な構文を考える。

$$X \text{ が } Y \text{ を } V$$

この構文にて、[3] では、『V が感情・感覚・思考の動きを表す動詞をとる場合、Y を主役に立てた受動表現では、感情や感覚が X から発して Y に到ると、通常見立てることができる。思考の動きを表す動詞も同様に考えることができる』と述べられている。[3] を踏まえ、感情・感覚・思考の動きを表す動詞が含まれる受動パターンで、動作主となる格要素に格助詞「から」を追加する。

- 複合助詞「によって」

受動パターンは入力文に複合助詞「によって」が使用されている場合、全く対応できていない。入力文に「によって」が付属する名詞は能動文に

において主格となっている．そのため，受動パターンにおいて動作主となる格要素に複合助詞「によって」を追加する．

6.2 改良を行った変換規則

改良を行った変換規則を表 6.1 示す．

表 6.1: 改良後の変換規則集

#	条件パターン	帰結パターン
1	A が V	A に/によって/から * $V.reru$
2	A が B と V	A に/によって/から * B が $V.reru$
3	A が B に C を V	A に/から/によって B が/に C を/が $V.reru$
4	A が B を C に V	A に/から/によって B を/が C が/に $V.reru$
5	A が B に V	A に/によって/から * B が $V.reru$
6	A が B の C を V	A に/によって/から * B が C を $V.reru$
7	A が B を C と V	A に/によって/から * B が C と $V.reru$
8	A が B を V	A に/によって/から * B が/を $V.reru$
9	A が B を C から V	A に/によって B が C から $V.reru$

助詞「から*」は V が「感情動作・知覚動作・思考動作」となるパターンのみ有効である．この判別には，日本語語彙大系の結合価パターンにおける用言意味属性を参照する．

6.3 改良の効果

表 6.1 の変換規則を使用し実験を行った結果，入力文 57 文に対し，49 文を正しく能動文に復元することができ，カバー率は 86%(49 文/57 文)に向上した．この結果，受動パターンの助詞が不足しているという問題が 15 件から 3 件へと減少し，受動パターンの助詞不足が改善された．

表 6.2: 改良後の変換失敗の内訳

分類	原因	頻度
1	形態素解析の誤り	1
2	受動パターンの助詞が不足	3
3	格要素の意味属性制約	3
4	格要素の個数に変化	1

第7章 評価実験

受動パターンの網羅性と正確性を検証するためにオープンテストとして評価実験を行う。

7.1 評価方法

第5章の予備実験と同様に評価を行う。

- 受動パターンが入力文の格要素を満たしているか評価
- 能動文が意味的に正確であるか評価

以上の評価を満たした出力があれば、受動パターンが受動文をカバーしているものとし、正解の出力とする。また、変換規則により作成した受動パターンのカバー率と精度を以下の式で評価する。

7.1.1 カバー率

受動パターンが入力文に対応しているかを判断する際に、入力文に対して複数の能動文を出力する場合がある。出力のうち1つでも正解となる出力が得られた場合、受動パターンが入力文を網羅していると考え、以下の式でカバー率を求める。

$$\text{カバー率} = \frac{\text{1つでも正解となる出力が得られた入力文数}}{\text{全入力文数}} \times 100[\%]$$

7.1.2 精度

変換規則により作成した受動パターンの精度を以下の式で求める。

$$\text{精度} = \frac{\text{正解と判断した出力件数}}{\text{全出力件数}} \times 100[\%]$$

7.2 入力文

本実験では、一般的な学習用辞書から収集した単文 [6] の受動文 100 文を使用する。以下に例を示す。

- 4月に新しいお札が発行される。
- 泥棒は近所の人に捕まえられた。

7.3 実験結果

入力文 100 文のうち、56 文が意味的に正しく受動パターンで判断できた（カバー率 56%）。また、入力文 100 文に対し、157 件の出力候補が得られ、62 件が意味的に正しかった（精度 40%）。この結果、格助詞「が・を・に」や助詞「から・によって」を使用した受動文に対してはカバーできているといえる。以下に意味的に正しいと判断した例を示す。

入力文：3人の反体制作家はその国から追放された。

出力文：[が 3人の反体制作家を その国から/より に/へ/まで 追放する]。

入力文：あのジェット機事故で多くの貴い生命が失われた。

出力文：多くの [が 貴い生命を あのジェット機事故で 失う]。

入力文：ことしも新しい雑誌が次々と創刊された。

出力文：ことしも次々と [が 新しい雑誌を 創刊する]。

入力文：そのショーは失敗とされた。

出力文：[が そのショーを 失敗に/と する]。

入力文：そのレスラーは猪木の勢いに圧倒された。

出力文：[猪木の勢いが そのレスラーを 圧倒する]。

入力文：その外交官はアメリカへの入国を許可された。

出力文：[が アメリカへの入国を その外交官に 許可する]。

入力文：その候補者は選挙違反で検挙された。

出力文：[その候補者を 選挙違反で 検挙する]。

入力文：その公開討論会はテレビで放送された。

出力文：[が その公開討論会を テレビで 放送する]。

入力文：その国有地に新しい病院が建てられた。

出力文：[が 新しい病院を その国有地に/へ 建てる]。

入力文：その小説は評論家に徹底的に叩かれた。

出力文：徹底的に [評論家が その小説を 叩く]。

7.4 誤り分析

本実験において能動文への出力が失敗したもの、および、意味的に正しいと判断できなかった出力の内訳を表 7.1 に示す。

表 7.1: 評価実験における変換失敗の内訳

分類	原因	頻度
1	入力文の形態素解析の誤り	6
2	適切な用言受動パターンの作成不足	1
3	受動パターンの助詞が不足	11
4	格要素の意味属性制約	9
5	複数の候補からの選択	10
6	格要素の個数に変化	5
7	文の曖昧性	1
8	日本語語彙大系の網羅性	1

予備実験では見られなかった新たなタイプの誤り事例について分析を行う。

7.4.1 分類1 「入力文の形態素解析の誤り」

以下に誤り事例を示す。

入力文：その外交官は本国に強制送還された。

出力文：出力なし

正解文：その外交官を本国に強制送還する。

次に入力文に対して、使用するべきパターンを示す。

能動パターン：N1がN2をN3に/へ送還する

受動パターン：N1から/に/によってN2が/をN3が/に/へ'送還する'.*reru*

意味属性制約：N1(3主体)N2(3主体)N3(388場所 2610場 362組織)

受動パターンを参照すると、「外交官」は名詞変数 $N2$ と、「本国」は名詞変数 $N3$ と適合して能動文が出力されると考えられる。しかし、実験では出力が得られなかった。

そこで、入力文の形態素解析結果を参照したところ次のとおりであった。

その外交官は本国に強制送還された。

1. /その (4200, 其の, 其の)
2. /外交官 (1100, NI:266)
3. +は (7530)
4. /本国 (1100, NI:385, NI:463)
5. +に (7430)
6. /強制 (1220, NI:1826)
7. +送還 (1220, NI:2175)
8. +さ (2432, する, さ)
9. +れ (7113, れる, れ)
10. +た (7216)
11. +。 (0110)
12. /nil

図 7.1: 入力文の形態素解析結果

入力文に使用されている動詞は「送還する」ではなく、「強制送還する」となっている。そのため動詞「送還する」のパターンと照合されておらず、パターンによる解析が行われなかった。これは、「強制」が副詞として解析されていないことに起因する¹。この問題に対しては、形態素解析をより正確に行うことで改善されが、本研究の対象としない。

7.4.2 分類 2 「適切な用言受動パターンの作成不足」

以下に誤り事例を示す。

入力文：その囚人は終身刑を言い渡された。

出力文：出力なし

正解文：その囚人に終身刑を言い渡す。

入力文に対して出力文を得ることができなかった。「言い渡す」の受動パターンを参照したところ次の通りであった。

能動パターン：N1はN2を言い渡す

意味属性制約：N1(4人)N2(“判決”)

名詞変数 N1 の助詞が「は」となっており、表 6.1 の条件パターンに該当しないため、受動パターンが作成されていなかった。本研究で使用した変換規則では対応できず、パターンごろに受動パターンを作成しなければならない。

¹しかし、入力文の「強制」を複合動詞とみなす場合もあり、その場合、日本語語彙大系のパターンが不足していたともいえる。

また、「言い渡す」という能動パターンには、入力文の「囚人」に対応する格要素が存在しない。正解文の出力を得るためには、「囚人」に対応する格要素を追加した以下に示すパターン対が必要となるが、次は新たに「分類6，格要素の個数に変化を要する」問題へと転化する。

能動パターン： $N1$ は $N2$ を $\#1[N3$ に/へ] 言い渡す
 受動パターン： $N1$ から/によって $N2$ が/を $\#1[N3$ が/に/へ]'言い渡す'.*reru*
 意味属性制約： $N1(4人)N2(“判決”)N3(*)$

分類3「受動パターンの助詞が不足」

第5章の予備実験の結果より、変換規則に「から」、「によって」を追加したが、それでも助詞が不足した。以下に誤り事例を示す。

(3a)

入力文：あの大リーガーの場外ホームーには度胆を抜かれた。

出力文：あの大リーガーの場外ホームーには [が 度胆を 抜く]。

正解文：あの大リーガーの場外ホームーが度胆を抜く。

以下に使用したパターンを示す。

能動パターン： $N1$ が $N2$ の度胆を 抜く
 受動パターン： $N1$ に/によって/から $N2$ が度胆を '抜く'.*reru*
 意味属性制約： $N1(*)N2(4人)$

使用した受動パターンの変数 $N1$ には、助詞「には」が存在しない。そのため、名詞句「あの大リーガーの場外ホームー」が受動パターンの名詞変数 $N1$ に適合しなかった。入力文を正しく解析するには、受動パターンを「 $N1$ には」とする必要がある。

次に(3a)と性質の異なる助詞の問題を次の事例で考える。

(3b) 入力文：あの政治家はテロにより命を奪われた。

この入力文は次に示す2つの能動文に復元されることが考えられる。

復元例1：(組織が)テロで政治家の命を奪う。
 復元例2：テロが政治家の命を奪う。

復元例1では，入力文に明記されていない「組織」という動作主が「テロ」という手段を用いて政治家の命を奪うと解釈できる．一方，復元例2では「テロ」という擬人化された主格が政治家の命を奪うと解釈できる．したがって，能動パターンと受動パターンの対として，少なくとも以下のパターンが必要であり，さらにいくつかの受動パターンを追加しなければならない．

既存パターン
能動パターン：N1がN2を奪う
受動パターン：N1に/によってN2が/を'奪う'.reru
意味属性制約：N1(4人*)N2("タイトル"2385生命現象)

追加パターン
能動パターン：N1がN3でN4のN2を奪う
受動パターン1：N4は/がN3によりN2を'奪う'.reru
受動パターン2：N4は/がN1に/から/によりN3でN2を'奪う'.reru
受動パターン3：N4はN1のN3に/から/によりN2が/を'奪う'.reru
受動パターン4：N1に/から/によりN3でN4のN2が/を'奪う'.reru
意味属性制約：N1(4人*)N2("タイトル"2385生命現象)N3(*)N4(*)

表7.1の分類3に該当する入力文を，助詞の追加で意味を正しく解析できるものと，できないものに分類する．助詞の追加だけでは意味を正しく解析できないものとは，助詞の追加により，分類6「格要素の個数の変化」，および，分類7「文の曖昧性」の問題へと転化するものである．

助詞の追加で解析できるもの
<ul style="list-style-type: none"> ・この文書には最終条件が含まれる。 ・その件については銘々の判断に任せられた。 ・その雑誌のクイズにはたくさんの回答が寄せられた。 ・その事件については裁判所の判断に委ねられた。

助詞の追加だけでは解析できないもの
<ul style="list-style-type: none"> ・あの政治家はテロにより生命を奪われた。 ・昨日その工場で役人による抜き打ち監査が行われた。 ・今日の交通ストで約9万人の通勤客が足を奪われた。 ・その言葉は別の意味にも用いられる。 ・これは速度の意味で用いられる。 ・このシステムにより交換機からデータが収集される。

太字で強調した助詞は，助詞が受動パターンに存在しなかったものを表す．本実験では「により・による・で・には・については・にも」の助詞が不足していることが分かった．

7.4.3 分類5「複数の候補からの選択」

以下に誤り事例を示す。

入力文：このホテルはよく商談の場に使われる。
出力文：よく [このホテルが を 商談の場に 使う]。
正解文：このホテルをよく商談の場に使う。

以下に候補として使用した2つのパターンを示す。

能動パターン1：N1がN2をN3に使う
受動パターン1：N1から/に/によってN2が/をN3が/に'使う'. <i>reru</i>
意味属性制約：N1(3主体) N2(-2702順序 1190金銭 2670時間) N3(*)
用言意味属性：利用

能動パターン2：N1がN2をN3に使う
受動パターン2：N1から/に/によってN2が/をN3が/に'使う'. <i>reru</i>
意味属性制約：N1(3主体 760人工物*) N2(*) N3(*)
用言意味属性：利用

上記に示した2つのパターンは、意味属性制約のみが異なる。受動パターン2は、受動パターン1と比べ、自由度の高い意味属性制約が掛けられていることがわかる。正解文を出力するには、受動パターン2が使用されるべきである。しかし、今回候補選択に問題があり、意味属性を一部でも厳密に満たす受動パターンが優先されたため、正解となるパターンが出力されなかった。

7.4.4 分類7「文の曖昧性」

以下に誤り事例を示す。

入力文：あの会社は誇大広告で非難された。
出力文：誇大広告で [が あの会社を 非難する]。
正解文1：(ある団体が) あの会社の誇大広告を非難する。
正解文2：(ある団体が) 誇大広告のあの会社を非難する。
正解文3：(ある団体が) あの会社を誇大広告で非難する。

入力文に対し、正解文が二通り考えられる。正解文を出力するには、入力文に対し、以下の解釈がそれぞれ必要である。

解釈1：あの会社は誇大広告が原因で(ある団体に)非難された。
解釈2：(ある団体が)あの会社の誇大広告を(対象として)非難した。
解釈3：(ある団体が)誇大広告を利用してあの会社を非難した。

常識的には、入力文の文意は解釈1である。常識というのは、入力文の外の知識を含んでいる。解釈1に限定する処理は常識や文脈をとらえた上での処理であり、意味理解という範囲である。しかし、パターンに基づく意味解析は、文のもたらず解釈を出力することを処理範囲としているため、解釈1から解釈3まですべて出力するべきである。これらの解釈を満たすためには、受動パターンの助詞の不足、格要素の追加など、対策を処理の段階ごとに設け、1つの能動パターンから複数の受動パターンを作成する必要がある。本研究で作成した変換規則では、1つの能動パターンに対して、1つの受動パターンを作成したが、複数の受動パターンが必要となる場合に対応できていない。これらの問題に対応するために、1つの能動パターンに対して複数の受動パターンを用意する必要があることがわかった。

7.4.5 分類8 日本語語彙大系の網羅性

以下に誤り事例を示す。

入力文：この話は女性の前ではちょっとはばかりる
出力文：出力なし
正解文：女性の前ではこの話をはばかりる。

以下に動詞「はばかりる」のパターンを示す。

能動パターン：N1が人目を憚る
受動パターン：N1に/によって人目が/を'憚る'.reru
意味属性制約：N1(2 具体 1001 抽象物 1235 事)

「はばかりる」の能動パターンは日本語語彙大系において上記の1つのみである。能動パターンを使用した意味解析において、入力文に「人目を憚る」が含まれていないと適合させることができない。これは、パターン中の「人目を」が、「はばかりる」の命題表現における必須格として定義されているからである。よって、入力文において、「人目を」という格要素が記述されていないため、受動パターンが照合されず、解析することができなかった。この「人目を」を省略してある文に対して、日本語語彙大系では網羅できていないため、本研究では問題の対策を行わない。

第8章 おわりに

本研究では、大局的に能動パターンから変換規則による受動パターンの作成を行い、受動パターンの網羅性を検証した。具体的には、日本語語彙大系の結合価パターンを9つの変換規則に従い、格助詞を変形し、受動文の格助詞に対応する受動パターンを11,310件作成した。このパターンを用いて、寺村の受動態の考察に基づいた規則が適用されているかを確認したところ、変換規則において助詞の不足という問題があった。そこで、変換規則に改良を行い、大局的に受動パターンを作成し直した。

評価実験において、受動文パターンの網羅性と正確性を検証するため、学習辞書などで使用された単文の受動文に対して受動パターンでの意味解析を行った。その結果、受動文100文に対する受動パターンのカバー率は56%であった。また、誤り事例の分析により、作成したパターンに動作主、対象物、あるいは手段を表すための多様な格助詞を網羅するための助詞が、改良した変換規則においても不足していることがわかった。同時に、助詞による文の曖昧性が存在する場合、任意格となる格要素の追加や、格助詞の示す意味を正確に捉えたパターンの追加が必要となることがわかった。

今後の課題は、大局的に作成した受動パターンに対し、局所的に動詞ごとの差異を吸収するための修正を行うことである。

謝辞

本研究を進めるにあたり，種々の御助言を頂きました鳥取大学工学部知能情報工学科
計算機工学講座C池原研究室の池原悟教授に心から御礼申し上げます．本論文に関して
御指導を頂きました鳥取大学工学部知能情報工学科計算機工学講座 AB の菅原一孔教授
に心から御礼申し上げます．

本研究を進めるにあたり，御指導を頂きました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，細部にわ
た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言を頂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] Masato Tokuhisa, Kumiko Endo, Yuya Kanazawa, Jin'ichi Murakami, Satoru Ikehara: Evaluation of Pattern Generalization Effect under Development of Pattern Dictionary for Machine Translation, Proceedings of the Conference of Pacific Association for Computational Linguistics(PACLING 2005), pp.311-318, 2005.
- [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] 寺村秀夫: 日本のシンタクスと意味 vol. I, くろしお出版, 1982.
- [4] 庭三郎: 現代日本語文法概説, 2004.
<http://www.geocities.jp/niwasaburoo/index.html>
- [5] 徳久雅人, 池原悟, 村上仁一: 非線形言語モデルに基づく文型パターン型言語知識ベースの開発, 電子情報通信学会ソサイエティ大会講演論文集, 基礎・境界, シンポジウム「言語の意味と思考過程」, pp.49-50, 2007.
- [6] 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回年次大会発表論文集, pp.372-375, 2005.
- [7] 福田泰介, 徳久雅人, 村上仁一, 池原悟: 受動態に変換した結合価パターン辞書, 電子情報通信学会技術研究報告, 思考と言語, TL2009-43, pp.19-24, 2010.