

日本語文における述語任意成分の
意味解析実験について
野垣内 出、鈴木 雅美、榎本 博史
(KDD研究所)

[0. はじめに]

筆者らは、結合価（バレンシー）文法を中間構造とする日英機械翻訳のアルゴリズムの研究を進めている。この結合価文法の枠組の中で筆者らは日本語文中の「名詞（句）+助詞」のストリングを述語との関係に基づいて必須成分と任意成分の2種の成分に区分して考察している。

必須成分はその「名詞（句）+助詞」がなければ文が成立しないものとして定義している。また任意成分は必須成分以外のものと定義している。

例えば「彼は書店で雑誌を買った。」[S1]において必須成分は「彼は」「雑誌を」となり、任意成分は「書店で」にあたる。

本報告では、任意成分を解析する際に生じる問題点とその解析手法について論じる。

・任意成分解析の問題点

日本語文における任意成分は「名詞（句）+助詞」で表層上表現される。したがって名詞と任意成分を示すマーカである助詞類の一方または組合せにより一意に深層格を決定するのであれば問題はない。しかし、実際に名詞もマーカの助詞も一意に深層格を決定できない。助詞を中心に考察をしても、ひとつのマーカ（助詞）が表層上から数種の深層格へのマーカとして機能しており、多義性が生じている。

たとえば助詞「て」は、その示す深層格は原因格、理由格、道具格、手段格、付帯格、時間格、条件格、場所格などのように多様である。この任意成分の多義の問題は機械翻訳システムの問題として考察すれば訳文が数多く出力される、または誤訳が出来るという問題となって表れてくる。日英機械翻訳の場合では、日本語の任意成分は、英語側では「前置詞+名詞」となることが多いため、対応する前置詞の誤出結果が決まらないということにつながる。

このため本報告では動詞分類と詳細な名詞意味素性の利用による任意成分解析のためアルゴリズムを作成したので報告する。

[1. 任意成分に関する基礎調査]

任意成分解析のための一次データ作成のために、1674文について基礎調査を行なった。

この基礎調査の概要については報告をすでに行なっているため、ここでは、調査のあらましを説明する。

・データおよび調査

国際通信に関する勧告集であるCCITT orange book中の動詞および基礎動詞計250語を選んだ。また調査文献よりこの250語を使用した例文1674文を抽出して例文データとした。

日本語における任意成分は表層上に前述した様に原則的に「名詞（句）+助詞」の形で表わされる。この内、助詞は深層格を示すマーカとして機能している。この例文データより「名詞（句）+助詞」に関して必須成分ではないものをぬき出し、深層格名を人手により付加し基本データとした。この基本データは動詞対応であり、機械翻訳における対象言語である英語との対応を知るために対訳が付加されている。また英語における対応任意成分のマーカをぬき出した。

・逆引辞書

基本データは日本語マーカ、格名、動詞（日本語・英語）、対応英語マーカと英日各例文のデータの集合である。このデータ間の互いの相関関係、例えば日本語マーカと深層格名の関係はマーカの多義性を知る上で重要な資料となるため、データ間の対応関係を調査するための逆引辞書を作成した。この逆引辞書より作成した任意格マーカの深層格への多義性の一例を【Fig-1】に示す。なお、逆引辞書における対応関係は以下の通りである。

1. 日本語マーカ
2. 日本語マーカ対応英語
3. 日本語マーカの格名1,2
4. 日本語マーカと日本語動詞
5. 日本語マーカ対応英語と英語動詞
6. マーカ対応英語と格1,2

・マーカと任意成分深層格との対応

調査にもとづき、日本語任意成分の多義性を次の3種に分類した。

a. マーカに対応して深層格が推定可能

(マーカに多義性が生じないもの)

例「・・・の後に」「・・・のごとく」など

b. マーカと名詞との対応により深層格が推定可能

(マーカには多義があるが名詞との共起関係で多義を制限できるもの)

例「・・・ので」など

c. マーカと名詞対応でも深層格が決定できない。

(マーカと名詞との共起関係においても多義が制限できないもの)

例「・・・で」「・・・に」など

任意成分調査用例文(1674文)において

1. 「NP+で」の意味内訳(200文)

格名	%	対訳英語前置詞
a. 原因格	7.5%	because of
b. 理由格	3.5%	in
c. 道具格	20.0%	with, for
d. 手段格	19.0%	by
e. 付帯格	17.0%	in
f. 時間格	4.5%	in, after
g. 条件格	1.5%	for, because of
h. 場所格	27.0%	at, in, behind

2. 「NP+ごとく」(3文)

a. 付帯格	100.0%	like
--------	--------	------

3. 「NP+ので」(52文)

a. 理由格	73.0%	because of
b. 原因格	27.0%	because of, for

[Fig-1 マーカによる多義性の例]

[2. 解析方法]

任意成分に関する基本調査より調査例文の一部を示す。

例1：定められた変調モードで直列伝送を行なう。

(..Using a synchronous
transmission procedure)

2：これらの測定は受信機の電源部で行なう。

(These measurements apply with
the power supply of the receiver.)

3：任意の時点でデータ伝送を行なっている国際回線の割合

(The proportion of international
circuits which will at any time
be carrying data transmissions)

例1.2.3の「で」が表す深層格はそれぞれ道具格、位置格、時間格である。これらの文は「で」の多義性を示すものであるが、任意成分となっている名詞句に注目して、統語解析に用いている5種の意味素性（後述の[Fig-3]の1~5に相当する）を与えてみるとすべて+absという抽象的な名詞を示す意味素性となる。これらの文は助詞と動詞が同じものであるから、名詞自身から得られた情報がもし意味素性+absのみであるとすると、常に3種の多義性が生じることとなる。したがって、より細かい名詞意味素性分類が必要であろう。

一方、動詞によって任意成分深層格との共起の仕方に差があると考えられる。任意成分は本来はすべての動詞に共起しうる成分であり、その成分がなくとも文は非文とはならない性格を持つものであるが、調査文例中には動詞「合う」と一般的な任意成分である時間格の共起の例はなく、理由や条件を意味する深層格との共起が多い。このような共起関係の差が適当に意味分類された動詞ごとに利用できれば、多義性の解消に有効である。

すなわち任意成分の多義性の解決のためには、次に示す名詞中心と動詞中心の2つの解析方向が考えられるであろう。

1. Cause (略: Cause) =原因

動作・作用の原因となりえるもの。

例：風でロウソクが消えた。

2. Reason (略: Reason) =理由

動作・作用の理由となるもの。

例：花子が風邪で学校を休んだ。

3. Directional (略: dir) =志向

動作の志向的内容を表わすもの。

例：入学祝いに何をあげようか。

4. Conditional (略: Con) =条件

動作を決める判断基準を示すもの。

例：花子の返事で行くか行かないかきめる。

5. Comparative (略: Com) =比較

物事・人についての比較

例：太郎は花子より大きい。

6. Reference (略: Ref) =関連

動作の対象を表す。

例：経済的側面に関して討議する。

7. Path (略: Path) =経路

移動を表わす動詞で動作の経路となるもの。

例：シベリヤ経由でパリに行く。

8. Locative (略: Loc) =位置

動作・作用の行なわれる場所・抽象的場

例：公園で遊んだ。

9. Situation (略: sit) =状況

動作・作用の行なわれる背景を表わすもの。

例：花子は雨の中をぬれて帰った。

10. Theme (略: Theme) =提題

「は」でマークされ、文全体を修飾するもの。

例：春は桜がいい。

11. Time (略: Time) =時間

動作・作用の行なわれる時間を表すもの。

例：花子が8時に学校に着いた。

12. Quantity (略: Quant) =量・水準

物事の量的属性・水準を表わす。

例：私は水を一杯のみたい。

13. Source (略: Source) =起点

動作の時間的・空間的起点となるもの。

例：花子は学校から6時に帰る。

14. Goal (略: Goal) =目標

動作の帰着点を表わすもの。

例：花子は6時に駅に着いた。

15. Symmetry (略: sym) =一致

動作・作用の随伴者となるもの。

例：花子は太郎と映画に行った。

16. Instrument (略: Inst) =道具

動作の道具となるもの。

例：花子は鉛筆で答案を書いた。

17. Means (略: Means) =手段

動作の抽象的道具となるもの。

例：商人は大売出しで人を集めめた。

18. Manner (略: Manner) =付帯

動作の様態を表わすもの。

例：生徒達は一列に並んだ。

19. Role (略: Role) =役割

動作の付帯状況のうち役割を示すもの。

例：花子は秘書として働いている。

[Fig-2 任意成分深層格一覧]

・名詞中心の解析

数量的な制限が可能であることとアルゴリズムの一般化のために、実際に名詞一語一語を辞書データとして記述する方法ではなく名詞の意味素性の使用を前提とする。したがって名詞意味素性とマーカとの2項関係により深層格を推定する。名詞に関する情報以外は記述しにくくなるが深層格の推定ルールの記述は比較的容易である。

・動詞中心の解析

一般的に動詞（述語）から任意成分を深層格推定するためのデータとして名詞（意味素性）、任意成分マーカの対応などが利用できるほか、動詞自身の情報つまり意味と任意成分深層格との共起性なども利用できる。意味的に近い動詞は同じ深層格が共起しやすいことから、動詞も適当な意味分類を行なえば、辞書データの量を小さくすることが可能であり、深層格の推定ルールの記述も簡潔となると予想できる。

以上のように多義性の解消のためには2方向の解析が考えられる。先に述べたように調査の結果から任意成分の多義は名詞中心の解析のみでは解消できない。したがって動詞・名詞両方向からの解析方向が必要となる。また2方向からの解析を適当な方法で調節する必要がある。

[3. 意味解析の実際]

本項では実験用に作成されたアルゴリズムの概要について説明する。このアルゴリズムでは動詞・名詞の両者の情報を利用している。

・入力

任意格の意味解析は統語解析結果を入力としている。この統語解析結果は文中よりあるストリングが任意成分であると判定されれば、任意格を表す表層マーカと共に出力されている。

・出力

意味解析の結果として[Fig-2]に示す深層格名⁴⁾である意味ラベルを出力する。機械翻訳システム全体としては、この出力は変換部（対象言語におけるマーカと、変換されたストリングを出力する）に渡される。

・名詞

各名詞には28種類の意味素性[Fig-3]の束が付加されている。素性の一部([Fig-3]の1~5)はすでに統語解析に使用していたものを流用している。

・動詞意味分類

Nilsenの動詞分類、任意成分に関する基本調査に基づき動詞意味分類を行なった。[Fig-4]に分類を示す。各動詞にはこの動詞分類No.⁵⁾を付加した。

・マーカ+名詞素性テーブル

任意格マーカと名詞素性の組合せで推論可能な任意格（深層格）を与えるテーブルである。このテーブルは原則的に可能性を示しているだけで一意に結果を与えるものではない。

・動詞分類+マーカテーブル

任意マーカと動詞分類の組合せにより、深層格に格優先順位を与えるテーブルである。ここには先の調査などをもとに動詞分類ごとの深層格の共起性が与えられており、より共起性が高いものが最終選択される。

・解析過程

1) 統語解析結果より任意成分が表層の任意格マーカと共に抽出される。表層のマーカとそれに伴なうストリング（主に名詞句）はすでに分離されており、ストリングは対応する名詞素性を値として持つ。

2) 表層マーカと名詞素性の2つの項をキーとして深層格を推定する。ここで推定ルールは、「マーカ+名詞素性テーブル」に記述されており、可能性のある深層格を全て出力する。（名詞中心の解析）

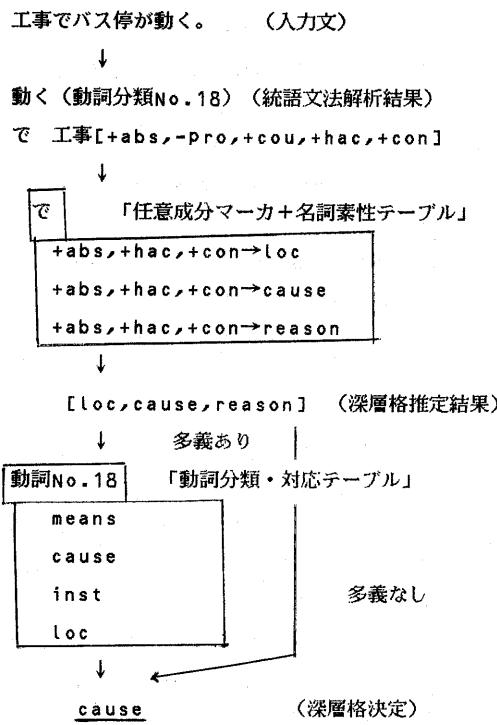
3) 2)の出力である深層格がもし一意に決定されれば解析は終了する。しかし表層のマーカは前述の様に多義性があるために一意に決定できない場合が少くない。もし、多義性があれば、複数の深層格が出力して次の過程に送られる。

4) 動詞分類とマーカによる決定

統語文法解析結果より、ある述語の中の任意成分であることが記述されている。各任意成分は、その任意成分自身が属する述語の動詞意味分類にしたがってここで決定は2)の結果の中より共起関係により深層格を再選択する形で決定する。（動詞中心の解析）[Fig-5]

1. **±Human** (略: ±Hum)
人または身体の一部を表す名詞であるかどうかを示す
2. **±Abstract** (略: ±Abs)
抽象名詞であるか具体名詞であるかを示す
3. **±Countable** (略: ±cou)
英語に直して数を数えることができるかどうかを示す
4. **±Personel** (略: ±Pro)
固有名詞であるかどうかを示す
5. **±Animate** (略: ±Ani)
動物であるかどうかを示す
6. **±Spacial** (略: ±Spa)
空間的広がりを持つかどうかを示す
7. **±Locational** (略: ±Loc)
空間に位置を占るかどうかを示す
8. **±Physical Object** (略: ±Phy)
物体であるかどうかを示す
9. **±Capacity** (略: ±Cap)
容量を持つかどうかを示す
10. **±Transparent** (略: ±Tra)
透過性があるかどうかを示す
11. **±Measure** (略: Msu)
計量する道具・手段であるかどうかを示す
12. **±Force** (略: For)
自律的な力または脅威があるかどうかを示す
13. **±Volitional** (略: Vol)
意志的な主体であるかどうかを示す
14. **±Content** (略: Cot)
思想・情報などの内容を持つかどうかを示す
15. **±Directable** (略: Dir)
一方から他方へ方向づけられたものであるかどうかを示す
16. **±Standard** (略: Stn)
基準・レベルを表すかどうかを示す。
17. **±Requital** (略: ±Req)
報酬・補償・仕返しを表すかどうかを示す
18. **±Possesible** (略: ±Pos)
所有されているものまたは属性であるかどうかを示す
19. **±Manner** (略: ±Man)
動作・行為・変化の状態を表すかどうかを示す
20. **±Set** (略: ±Set)
定義によって指定される集合になっているかどうかを示す
21. **±Event** (略: ±Eve)
出来事・催物であるかどうかを示す
22. **±Non-intensional event** (略: Nin)
不慮の事故性を持つかどうかを示す
23. **±Stative** (略: ±Sta)
状態を表すかどうかを示す
24. **±Controllable** (略: ±Con)
制御可能かどうかを示す
25. **±Living thing** (略: Liv)
生物であるかどうかを示す
26. **±Human activity** (略: ±Hac)
人間の活動に関するものであるかどうかを示す
27. **±Temporal** (略: ±Tem)
時間性を持つかどうかを示す
28. **±Attempt** (略: Att)
試みによって結果が変化するものであるかどうかを示す

[Fig-3 名詞意味素性一覧]



【Fig-5 任意成分解析過程の概略】

1.一項動詞A	15.知覚
2.一項動詞B	16.心理現象
3.一項動詞C	17.所有
4.移動	18.変更
5.移送	19.創造
6.変換	20.破壊
7.伝達	21.許容
8.収納	22.援助
9.付着	23.激励
10.包囲	24.期待
11.被覆	25.強制
12.接触	26.対称述語A
13.接近	27.対称述語B
14.感覚	28.特別述語

【Fig-4 動詞分類一覧】

【4. 実験結果および今後の方針】

・データおよびルール・テーブル類

「マーカ+名詞素性テーブル」すなわち任意成分マーカと名詞意味素性のマッチングにより、深層格を推定するテーブルに記述された推論ルールは「で」17行、「に」10行、「に関して」1行となっており、推論ルールは比較的少なくてすんでいる。

・名詞意味素性

機械翻訳システムとして任意成分解析以降のステップで深層格の意味レベルと意味素性が対象言語側に渡され、英語の前置詞を最終決定する。すなわち、任意成分解析で深層格の場所格が与えられたとすれば、英語側では(in, on...)などの前置詞の内から1つ選ばなければならない。このための情報も意味素性として与えているため名詞意味素性は本来の任意成分の解析で必要な以上に細分類されている。

・解析実験

実験は任意成分に関する基礎調査の例文などを中心とした比較的短い文約80文を対象として代表的な任意マーカ「で」「に」「に関して」の3種について行なった。「で」「に」は前述した様にマーカ自身が多義性を持ち名詞との共起関係からも多義を制限できないものである。「に関して」は逆に多義を生じないマーカである。

実験結果

マーカ	例文数	正解率
「で」	50文	45文 (90%)
「に」	20文	17文 (85%)
「に関して」	2文	2文 (100%)

なお正解率は正しい深層格が優先順位の第一番目として出力されたものの数である。現在まで80文程度の解析実験しか行なっていないが、各ルールやテーブルに検討を加えることにより、より大量の文の解析時にも動詞との共起関係を利用していることから80%前後の解析率を維持できると予想している。また解析に失敗した主な原因は動詞分類別の優先度テーブルの順位によるものと名詞意味素性付加の誤りの2つが大きな原因である。

つた。

・今後の方針

今後、解析率の向上のために各推論ルール、動詞分類別の格優先順位などの再検討を行なう。また名詞の意味素性も、より多くの実験に基づいて再検討を行ないたい。

【5. さいごに】

日本語文における任意成分の解析手法について述べた。日本語文の任意成分は助詞によってマークされているが多義性が生じ、たとえば機械翻訳などのように深い意味解析を行なう必要がある場合には、問題が多い。

この多義性の解消のため、名詞・動詞の分類を行ない推定ルールと動詞との共起関係を利用した新たな任意成分解析方法について述べた。

最後に日頃御指導いただき KDD 研究所鍛冶所長、野坂副所長、博松次長、武田第1特別研究室長に感謝する。

言語調査上の御協力をいただいた㈱日本IR 友清氏をはじめとする調査グループの方々、プログラム作成の御協力をいただいた㈱SCC 佐野氏に御礼申し上げる。

【6. 参考文献】

1) 野垣内ほか「日本語文における述語の任意成分について」第31情報処理学会全国大会(2H-10)

2) 任意格調査の例文の出典

学習研究社「国語大辞典」

国立国語研究所「動詞の意味用法の記述的研究」

文化庁「外国人のための基本用例辞典」

研究社「新英和辞典」

岩波書店「国語辞典第2版」

小学館「国語大辞典」

ITU協会 CCITT勧告集 オレンジブック

日経マグロウヒル「人工知能」

3) 水谷、石綿ほか「文法と意味1」2章

”結合価からみた日本語文法” <朝倉>

または

鈴木ほか「結合価に基づく英日機械翻訳の試み」

自然言語処理研究会(49-2)

4) 柴谷他「言語の構造 意味・統語編」

<くろしお出版> 1982

5) Don L.F. Nilsen

“Toward a semantic specification of
deep case” <MAUTON> 1972

6) Eugene A. Nida 成瀬訳

「翻訳学序説」<開文社出版> 1972

工事　||で　||バス停　||が　||動く t1_7　||.

The bus station is shifted because of the construction.

風　||で　||ろうそく　||が　||消え t8_1 た t27_7　||.

The wind blew the candle out.

同様の t10_9　||処理　||が　||CRT画面　||上　||で　||行なえる t8_7　||.

The same process is possible on the CRT.

L I S P　||処理系　||が　||L I S P　||言語　||で　||記述さ t7_3 れる t11_7
||.

L I S P system is described by L I S P language.

ループ4　||は　||主管庁　||によって　||線　||の　||保守　||に　||考え t8_1 ら
れ t12_1 た t27_7　||.

The Loop4 was designed by the administration for the maintenance of the line.

子供達　||は　||一　||列　||に　||並ん t5_14 だ t28_7　||.

The children stood in a line.

インパルスカウント　||に関して　||保守限界　||は　||勧告さ t7_3 れ t11_1 ない
t25_7　||.

The maintenance limits is not recommended for the impulse counts.

回路　||は　||差動雑音　||に　||敏感だ t10_7　||.

The circuit is sensitive to the differential noise.

西洋　||の　||文化　||は　||日本　||で　||完全な t10_8　||形　||で　||受け入れ t
8_1 られ t12_1 た t27_7　||.

Western culture was accepted in the perfect form in Japan.

[付録 任意成分を含む例文の機械翻訳例]