

## 動詞意味素性の付加による 日本語文アスペクトの解析および翻訳について

の が い と い ず り ず り ま ず り み じ かの ひ し  
野垣内 出 鈴木 雅実 標 博史

(KDD 研究所)

### [0] [はじめに]

筆者らは、科学技術文献を対象とした、日英機械翻訳のためのアルゴリズムを研究している。

日本語においてアスペクトは補助用言で表われているが、

- i) 日本語のアスペクト解析は解釈上の多義性を必然的に生ずる。しかし、副詞などの情報を効果的に利用すれば、ある程度 多義は制限できる。
- ii) 補助用言の連鎖の解析は辞書データの増大、またはアルゴリズムの複雑化をまねきやすい。アルゴリズムの複雑化は、文法等の記述力の低下につながるため、記述力のあるアルゴリズムを作成する必要がある。

以上の2点は、日本語を対象とした自然言語処理システムにおいて、アスペクトを扱う場合に問題となる。現在までに 日本語においてアスペクトを解析するアルゴリズムは、いくつか提案されているが、この両者について解決されているアルゴリズムは少ない。

そこで 筆者らは、この2点の解決を目的として一種の動詞意味素性である、アスペクト素性を動詞に付与した アスペクト解析・変換・生成アルゴリズムを作成した。

以下にこの実験アルゴリズムについて報告する。

### [1] [アスペクト解析]

機械翻訳における現象面でのアスペクト解析の問題点は多義性と補助用言の連続である。ここでは、より具体的に考察する。

#### ・ [多義性]

日本語では、アスペクトは「動詞」+・・・+「補助用言」の形で表されるが、補助用言自身が多義性を持つ。

「彼はその本を読んでいる。」[S-1]

において(読み続けている 読んでしまっている)の2つの多義性が生ずる。これは(・・・ている、・・・、でいる)の補助用言自身に多義性があるからである。しかし、ある種の動詞においては、この現象は生じない。

「彼はその本を知っている。」[S-2]においては、(知り続けている)の意味はない。このことからアスペクト解析を行なうために動詞分類を行なうことの必要性は広く知られている。

#### ・ [副詞などの共起]

一般的に日本語文のアスペクトの多義性は副詞などの共起により制限される場合がある。

彼はサリンジャーの小説を読んでいる。[S-3]

「今、読んでいるところである」の意味 進行相

「すでに読んでしまっている」の意味 完了相

彼はすでにサリンジャーの小説を読んでいる。

[S-4]

「もうすでに読んでしまっている」の意味 完了相

これらの例からも明らかなように、副詞などは、共起によってアスペクトの多義性を制限する場合がある。

また

雨が降っている。[S-5]

進行相

雨が毎日、降っている。[S-6]

反復相

のようにアスペクトが変化するものがある。

#### ・ [補助用言の連続]

補助用言に限らず、日本語文では用言が連続し、解析において難点となっている。

ここでは、「動詞」+「補助用言」+・・・+「補助用言」の形のみを考えてみる。

一般的に動詞を類別したものに補助用言を組みあわせてアスペクトを解析する方法では、補助用言もまた分類されていなければならず、かつ用言連続に関しその組み合わせたものの動詞分類がわかっていなければ、ならぬ

い。このために組み合わせた動詞自身を辞書データとして必要とする、またはアルゴリズムが複雑となる。辞書データとして記述した場合には量はかなりのものとなると予想される。

## [2] [アスペクト解析に必要なデータ類について]

### ・ [動詞分類]

動詞、形容詞、形容動詞について、アスペクト解析のために、一種の動詞意味素性であるアスペクト素性を付加し、アスペクト素性の組み合わせにより動詞分類を行った。

#### a、アスペクト素性

fig-1に示す8種のアスペクト素性を用いた。実際には、後述の動詞分類などで示されるように+/-の記号を付加して使用されている。

#### b、動詞分類

fig-2に示すように動詞は8分類されている。データ構造上はアスペクト素性の付加の様子によって分類されている形式をとっている。

#### c、アスペクト形式素の分類

アスペクトを表す補助用言をまとめてアスペクト形式素と呼んでいる。これらのアスペクト形式素もアスペクト素性の付加のされ方によって、9分類されている。

なお、実際の辞書データは各グループを表わす記号で記述しており、辞書データの量は、さほど増加していない。(fig-3)

#### d、副詞、名詞などの分類

副詞はアスペクト解析のために一般的な分類を細分化した形でまとめられている。副詞全体で大きく10グループに分けられ、それぞれが細目に分かれ細目では25項に分類されている。なお、現在のアルゴリズムで使用しているのは、この分類の内ごく少数である。

名詞類はアスペクトの解析の際「反復相」であるかの判定などに使用される。ここでは複数名詞句として「すべて、全員、...」などが分類されている。

また副詞にもアスペクト素性が付加されている。

<u>state</u>	状態
<u>continuity</u>	継続
<u>act</u>	動作
<u>volition</u>	意志
<u>result</u>	結果
<u>repeat</u>	反復
<u>start</u>	起動
<u>completion</u>	完了

fig-1 アスペクト素性

例

A [+state,+cont]	ある、できる
B [-state,-cont]	似る
C [-state,+act,+cont,+compl]	書く、言う
D [-state,+act,+cont,-compl]	与える、降る
E [-state,+act,-cont]	生じる、失う
F [-state,-act,+cont,+vol]	認める
G [-state,-act,+cont,-vol]	苦しむ
H [-state,-act,-cont]	知る、わかる

fig-2 動詞の分類

例

(1) [+state,+cont,+result]	である
(2) [+state,+cont]	ている
(3) [-state,+act,-cont,+result]	てしまう
(4) [-state,+act,-cont,+start]	始める
(5) [-state,+act,+cont,+compl]	通す
(6) [-state,+act,+compl]	ておく
(7) [-state,+act,-cont]	てくる
(8) [-state,+act,+cont,+repeat]	暮らす
(9) [-state,+act,+cont,-compl]	続く

fig-3 アスペクト形式素の分類

### [3] [アスペクト解析の実際]

統語解析結果を再解析する形でアスペクト解析は行なわれる。

アスペクト解析アルゴリズムは、述語とアスペクト形式素のそれぞれに付属しているアスペクト素性の束をとりだす。また同時に文中に筆者らが付加情報と呼ぶアスペクトに關係する副詞、名詞の一部があるかを判定し、あれば処理を行なう。

これらの処理はアスペクト素性の束と副詞などもつアスペクト素性間の演算で行なわれている。

ここでのアルゴリズムは大きく結合ルールと呼ばれるアスペクト素性間の演算ルールと、演算結果をアスペクトに判定するための解釈ルールの2つに分けられる。

ここで行なわれる演算を示す。

- i) 副詞があればそれに応じてアスペクト素性の付加、消去を行なう。
- ii) 2つのアスペクト素性の束を演算し、付加、消去を行なう。
- iii) ii)を左から右へくりかえし行なう。

この演算は原則として、アスペクト素性を消去する方向の演算であるため、ここでの結果は入力された素性数よりも少なくなる。この演算は「動詞+補助用言+・・・+補助用言」が補文構造であることを利用して行なっている。

結合ルールより処理されたアスペクト素性すなわち消去されていないものは1つの束としてまとめられて解釈ルールによってアスペクトを決定する。(fig-4に解釈ルールを示す。)このルールは、解釈できるすべてのアスペクトを出力する手続きで使用されるが、一部に制限的なルールもある。

一例として、あるアスペクト素性があれば一定のアスペクトの解釈を制限しているルールがある。

すなわちあるアスペクト素性の存在によって結果的に

あるアスペクトへの解釈を排するようなルールの記述も可能である。

たとえば[+act,+state,+cont]の束は([+act,+state,+cont]→状態相、進行相)となり、[+act,+state,+cont]の束は、2つの相の可能性はあるが、実際は状態相の解釈ルール中にnot(+act)すなわち+actが存在してはいけないことが記述できる。

- (1) 状態相 [+state,+cont]
- (2) 実現相 [-state,+act,-cont,+result]
- (3) 結果状態相 [+state,+result,+cont]
- (4) 完了相 [+act,+result]
- (5) 進行相 [+act,+cont]
- (6) 完結相 [-state,+act,+compl]
- (7) 起動相 [-state,+act,-cont,+start]
- (8) 反復相 [-state,+act,+cont,+repeat]
- (9) 反復完結相 [-state,+act,-cont,+repeat,+result]
- (10) 反復状態相 [+state,+cont,+repeat]

fig-4 解釈ルール (一部省略)

### [4] [変換・生成の実際]

ターゲット言語である英語では、アスペクトは現在形、完了形などの構造で表現されている。

ここでは、英語側にアスペクトを設定せず、アスペクトは日英共通であると仮定し、直接英語統語構造を変換出力する方法をとっている。ここでもルールと手続きは分離されており、変換テーブルと呼ぶルールによって変換を行なっている。

この変換テーブルは、入力に解析部分で出力されたアスペクト、動詞が「ル形」・「タ形」であるかの別、動詞グループ名であり、出力としては英語表層の形で記述されている。fig-6に変換テーブルの一部を示す。

今日も雪が降り続けている。

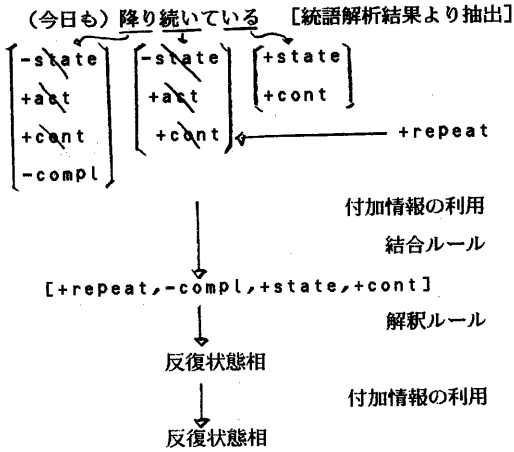


fig-5 アスペクト解析の流れ

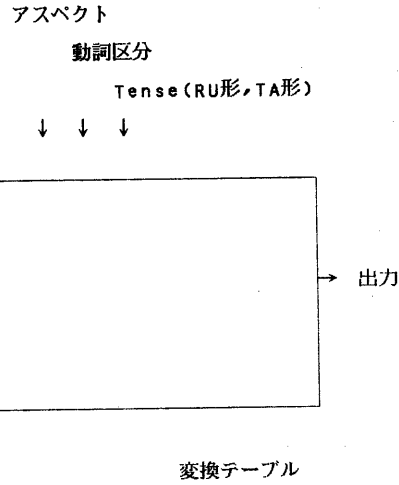


fig-7 変換の流れ

[6] [実験結果]

実験は次のデータで行なわれた。

副詞の共起がない場合

動詞グループ	日本語	英語
1	無相	
	RU形	無相現在
	TA形	無相過去
2	無相	
	RU形	
	TA形	
	状態相	
	RU形	無相現在
	TA形	無相過去
実現相	RU形	無相現在
	TA形	無相過去

fig-6 アスペクト変換テーブル

- ・動詞 基本的な動詞を中心に50動詞, 8分類
- ・アスペクト形式素 20数語, 9分類
- ・副詞 別項のとおり
- ・アスペクト分類 11種
- ・アスペクト素性 9種
- ・結合ルール 約10
- ・解釈ルール 10数ルール
- ・アスペクト変換テーブル 約90行

・ [データ類の作成および調査]

本実験に使用された言語データの分類および内容について説明する。

・動詞分類 分類数およびアスペクト素性の内容などについては、現在のところ文献 i), ii) を参考にしている。基本的な50単語について独自に調査を行なったものを使用している。

・アスペクト素性 これも文献 i) を参考にしているが、解析・変換アルゴリズムの中で使用されていること

を考慮して、result、repeat、startを追加した。

・アスペクト形式素の分類と内容 アスペクト形式素についても独自に調査して、アスペクト素性を与えている。

・副詞 ここで扱ったのは単独で副詞となるものでいわゆる副詞句については扱っていない。

・結合ルール（演算ルール） 文献 1)の意味解釈ルールなどを原形としているが、現在は内容的にかなり変化してしまっている。

・解釈ルール 独自に作成

・変換テーブル 独自に作成

#### ・【記述性】

前述したように、今回の実験では主としてアルゴリズムの可能性、またルールやテーブルの記述性がよいこととアルゴリズムの動作の確認のために行なわれた。

テスト文と呼ぶ例文を計算機上でアスペクト解析、変換、生成の各処理を行ない、「ルール」の手直しを行なった。テスト文はアスペクト解析、変換、生成の実験のために収集または作成された、かなり短い文章である。すくなくとも動詞分類ごとにテスト文は用意された。

実験においては「ルール」「テーブル」類の手直しやチェックの作業が大部分を占め、全体のアルゴリズムの修正は、ごく一部ですんでいる。修正の方法は、まず正しいアスペクト名が処理結果として出てきているか、またあるアスペクトからの変換、生成結果の確認、また原文と翻訳文との対応も確認している。

テスト文においては、アスペクト解析・変換・生成のそれぞれ基本アルゴリズムにはあまり修正を加えずに正しい翻訳結果が出ている。

この実験の結果、以下のことが判明した。

#### (1) 細かい演算の記述に富む

このアルゴリズムでは、一種の意味解釈にあたる結合ルールの細部、解釈ルールなどを中心に演算は素性について行なわれる。このことは動詞を分類または類別した結果について演算を行なう方法より細かい規則が記述できた。

すなわち 動詞分類より細かいアスペクト素性についての規則や演算が記述可能であった。

#### (2) 分類について

動詞分類などでは、素性を与えているため、分類作業が、はん雑かつ一貫性を失なうとの予想もあったが、実際の作業にあたっては、一種のチェック・テストを行ない、分類したため、作業量も比較的少なく、一貫性を失なうことなく行なえた。

#### ・【補助用言の連続】

補助用言の連続のアスペクト解析については既に報告したので、ここでは変換・生成について報告する。このアルゴリズムでは変換プロセスでターゲット言語でそのアスペクトを表わす統語の形を与えているので、生成についてはアスペクト生成について特別のアルゴリズムがあるわけではない。すなわち、連続に関しては変換プロセスの中で処理を行なっている。現状では意味上では一応正しい文を生成しているとはいうものの、いわゆる硬い感じのする文となっている例もある。変換ルールにおいてより自然な文とアルゴリズムの合理性・辞書データ類の数量的な妥当性（むやみに大きくならないこと）の三者を両立することは今後の課題である。

#### ・【多義性について】

実験は、独自に作成したテスト文と呼ぶ文を中心として行なったが、この作業と並行して、実際の文を中心とした小調査も行なった。

実際の文として調査対象に、「CCITT オレンジブック（日本語版）」「人工知能関係についての記事（数編）」を選んだ。文中で、アスペクトの多義性が、どのような語によって制限されているか、また どの語によって制限しているかを推定している。以下に例を示す。

現用の装置の多くは、12Vかそれ以下の電源で動作するように設定されている。[S-10]

スクランブルとディスクリンブル手順の詳細は付録に述べられている。[s-11]

エキスパートシステムの実用性が増すことは、これまでの研究者も強調している。[s-12]

黒板モデルは音声理解のために考えられたが、その後もなく画像理解にも応用されている。[s-13]

GMではその後、コンペア上の機械部品を認識して仕分けを行なうシステムを試作している。[s-14]

実際に文章を調査した結果、さまざまな語が、アスペクトの多義性を制限していることが、わかった。

[s-10]の例では、「現用の」が「ている」の多義、つまり進行相か完了相の2つの可能性を完了相1つに制限している。これは、「現用装置ならば設定はすんでいるはずだ。」という専門知識・常識による。また「語」について何かルールを記述するにしても[s-10]の「現用の」は文の中では次の名詞にかかっており、直接アスペクト形式素「ている」または「設計する」には、かかっていない。このために実験的なルールを記述しない限り、今回作成の実験プログラムでは、解析・変換・生成すべて可能であるが、多義性を解消することは不可能である。これはプログラムでは、名詞句の内部の意味解析は行なっていないためである。

[s-11]の例では、「述べる」の「に格」（「に」をマーカとする表層格）が埋れば、完了相であるというルールが書けそうである。またこの例でも常識を利用して多義をなくしているとも考えられる。

また、実際の文章を調べても、副詞類の共起による多義性の解消は多く、50弱の文で20数例あった。

一般にアスペクトの多義性は副詞などが支配していな

い場合には「彼はその本を読んでいる」[s-1]のように残る。文脈や常識や専門知識などが複合して日本語の場合はアスペクト形式素が本来もつ多義を解析しているものと考えてよい。つまり、アスペクト解析だけでなく機械翻訳全体を通じての問題につきあたる。

#### [7] [意味処理としてのアスペクト処理]

今回の実験では、アスペクト解析、変換はこのソフトウェア構造上、統語処理レベルの上位に位置しており、意味処理レベルにある。

一般の機械翻訳のための意味処理に比べると、データ量、ルール数も少なく、極めてシンプルであるが、一応意味処理を行なっていると考えられる。現在の実験プログラムに限ってしまえば、「ルール」「データ」は解析・変換において独立したテーブルとして扱っているためにプロセスは、複雑な動きをしていない。

また、意味上、最上位にくる解析結果は、アスペクト名（状態相など）であり、ソース言語とターゲット言語間でアスペクトは共通であると仮定しているため、意味処理にあたる変換部ではテーブルによる処理のみでアスペクトをターゲット言語の統語構造上にマップできている。

多義性の解消を別にすれば、アスペクト解析・変換は意味処理を行なうにしてもシンプルな形で実現できることがわかった。

ここでの意味処理がシンプルである理由としては、動詞をアスペクト解析のためだけに分類していることと、分類がさらに素性に分離され、ここでの演算が可能であり、その演算が単純であることと考えられる。また意味処理の出力は英語での、そのアスペクトを表わす統語構造であり、英語、日本語ともに言語学からの研究が進んでおり、比較的安定した言語調査が可能だったことも一因にあげられる。

[8] [さいごに]

「補助用言の分析と解析」

情報処理学会 第29回全国大会 4N-3

アスペクト素性の演算によるアスペクト解析・変換・生成の実験アルゴリズムについて報告した。

今後は動詞分類なども再考、また変換ルールの調整などを行なう、予定である。

最後に日頃御指導いただくKDD研究所鍛冶所長、野坂副所長、樽松次長、武田第一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、井倉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ます。

[9] [参考文献]

- i) 井上和子 1976 「変形文法と日本語 下」  
第5章 <大修館書店>
- ii) O.Akira 1971 " comparison of  
English and Japanese, with special  
reference to tense and aspect"  
working papers in Linguistics,  
University of Hawaii 3.4
- iii) 金田一春彦 編 1976  
「日本語動詞のアスペクト」<むぎ書房>
- iv) 中石 実 1980  
「日英語比較講座 第2巻 文法」  
第3章 テンス、アスペクトの比較  
<大修館書店>
- v) 第一特別研究室 1984~1985  
「KDD 研究所 第一特別研究室・内部資料」
- vi) 日本ITU協会 1978  
「CCITT オレンジブック 第Ⅷ巻」(vシリーズ)  
<日本ITU協会>
- vii) 日経エレクトニクス 1983 「人工知能」  
日経エレクトニクス・ブックス  
<日経マグロウヒル社>
- viii) 野垣内 出 他 1984

彼 Ⅱは Ⅱその Ⅱ道 Ⅱを Ⅱ走っ t4\_13 て t101\_20 いる tb407\_7 Ⅱ.

2/2

He has run the road.

彼 Ⅱは Ⅱその Ⅱ道 Ⅱを Ⅱ走っ t4\_13 て t101\_20 いる tb407\_7 Ⅱ.

He is running the road.

彼 Ⅱは Ⅱ彼女 Ⅱを Ⅱ待っ t3\_13 て t101\_20 いる tb407\_7 Ⅱ.

2/3

He is waiting her.

彼 Ⅱは Ⅱ彼女 Ⅱを Ⅱ待っ t3\_13 て t101\_20 いる tb407\_7 Ⅱ.

He has been waiting her.

雪 Ⅱが Ⅱ降り t1\_4 続い tg436\_14 て t101\_20 いる tb407\_7 Ⅱ.

1/3

It is snowing.

今 Ⅱも Ⅱ雪 Ⅱが Ⅱ降り t1\_4 続い tg436\_14 て t101\_20 いる tb407\_7 Ⅱ.

1/1

Now it has been snowing.

彼 Ⅱは Ⅱそれ Ⅱを Ⅱ認め t8\_1 始め td401\_1 た t27\_7 Ⅱ.

1/1

He began to admit it.

彼 Ⅱは Ⅱこの Ⅱ言葉 Ⅱの Ⅱ意味 Ⅱが Ⅱわかっ t4\_13 て t101\_20 き tg455\_4 た t27\_7 Ⅱ.

1/1

He came to understand the meaning of this word.

この Ⅱジュウタン Ⅱは Ⅱ壁紙 Ⅱと Ⅱ合っ t4\_13 て t101\_20 い tb407\_1 ない t25\_7 Ⅱ.

This rug does not match the wall paper.

1/1