中間表現を用いた日英時間表現の意味的対応関係の解析

的場 和幸　池原 悟　村上 仁一
鳥取大学大学院工学研究科
{matoba,ikehara,murakami}@ike.tottori-u.ac.jp

1 はじめに

日英機械翻訳においては，時間表現の意味のずれが翻訳品質を低下させる原因の一つとなっている．時間表現が表す概念には，事象の時間的位置を表すテンス，動作や状態の時間的な局面を表すアスペクトなどのカテゴリーがある．これを言語表現から見た場合，例えば，英語では，過去と完了では，異なる時間表現表現を用いるが，日本語では，同じ表現で過去と完了を表すなど，両言語間でずれが生じている．

時間表現の解析によって，客観的な意味を捉える研究には，日本語動詞の時間的性質を素性によって図式化しアスペクト解析を行ったもの [1]，また，数量表現を含む名詞句の時間表現を，意味によって分類し，形式表現に変換するアルゴリズムを提案したもの [2] などがある．しかし，これらの研究は日本語の時間表現を解析したものであり，翻訳を対象とはしていない．

従来の機械翻訳における時間表現の翻訳では，表層上の対応関係から作成された翻訳規則が用いられていた．しかし，言語表現上の違いを克服するためには，原言語から客観的な時間関係を抽出し，それを目的言語の枠組みの中で捉え直す仕組みが必要である．

そこで，本稿では，動詞と時間副詞の時間的性質を手がかりに，日本語時間表現から言語に依存しない中間表現を介して英語への翻訳を行う方法を提案する．

2 時間表現の翻訳手順

本稿では，日本語時間表現の英語への翻訳を，以下の手順に沿って行う．

1) 時間関係の抽出

テンスの過去，非過去，現在，未来を表す一般的な語尾形式「～タ」「～ル」「～テイル (テイタ)」「～ル+ダロウ」の場合について，動詞と副詞の時間的性質を手がかりに日本語時間表現から時間関係を抽出する．抽出された時間関係を，発話時・事象時，そして，事象を捉える基準時点である参照時 [3] の，3つの時点の関係として表す．

2) 中間表現を介した時間表現の翻訳

日本語側から抽出した時間関係を，英語の枠組みの中で捉え直すため，英語側からも現在形，過去形，未来形，進行形，完了形があらわす時間関係を，発話時・事象時・参照時からなる時点の関係で表す．翻訳は，両言語から得られた時間関係の対応づけにより行う．

3) 制約条件を用いた時間表現の修正

日本語から見て同じ時間表現であっても，英語側の用法によって時間表現に違いが現れる場合がある．例えば，進行形は，一時的な動作を表す場合に用いられ，長期的な動作や状態を表す場合には用いられない．そこで本研究では，結合価パターン [4] を用いて用法の推定を行い，文法的に正しい時間表現に修正する．

以上の翻訳手順の流れ図を図1に示す．

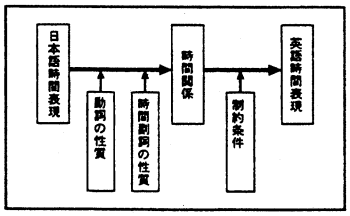

図 1: 時間表現の翻訳手順

以降では，語尾形式が「ル」をとるものをル形,「タ」をとるものをタ形，同様にテイル形，ダロウ形と表記する．

3 時間関係の表現形式

本稿では，文が表す時間関係を S(発話時)，E(事象時)，R(参照時) の3つの時点の関係として表す．S(発話時) とは，発話が与えられた時点であり，E(事象時) は，事象が発生した時点である．R(参照時) とは事象を捉えている基準時点である．以下に例文を挙げる．

(例文 1)　　私は家まで走った．

例文1では,「走った」という過去の事象をその時点までさかのぼって捉えている．そのため，E(事象時) と R(参照時) は，一致しており，S(発話時) 以前に位置する．つまり，時間関係は，次の図のようになる．

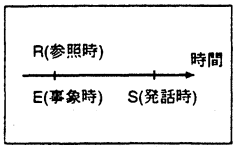

図 2: S(発話時),E(事象時),R(参照時) の時間関係

同時関係を「=」，順序関係を「→」で表すと，図2は，以下のように表せる．

$$\mathbf{E{=}R \rightarrow S}$$

以降では上記の表現形式を用いることにする．

4 時間表現の解析

4.1 動詞分類

日本語において時間的性質は，動詞の語彙的内容とも深く関わっているため，時間関係を抽出するにあたって，動詞の分類が必要となる．本稿では，従来の分類法 [5] を参考に，動詞を時間的性質から状態動詞，動作動詞，変化動詞に分類する．

表 1: 動詞分類表

動詞の種類		例
状態動詞		ある，いる
動作動詞	内的動作動詞	思う，考える
	外的動作動詞	走る，書く
変化動詞		着く，終わる

状態動詞は，存在や属性など，主体の状態を表す動詞である．形容詞も状態動詞に含める．動作動詞は，主体の動作を表す動詞である．本研究では，従来の分類に加えて動作動詞をさらに思考や感覚など人の内的な事象を捉える内的動作動詞と,「走る」や「飛ぶ」のように人や物の動態的な運動を捉える外的運動動詞に分類する．変化動詞は，主体の変化を表し，運動終了後，主体の状態が変わる動詞である．

4.2 中間表現への置き換え

日本語動詞における時間関係の解析においては，時間表現から発話時，参照時の関係，事象時，参照時の関係をそれぞれ抽出し，それを結合することにより行う．

以下の表2で語尾形式と動詞の分類から発話時と参照時の時間関係を決定する規則を示す．なお，表中の「＊」は，すべての動詞に適用可能である．発話時と，事象を捉える基準時である参照時との関係は，時間概念のテンスに相当する．

表 2: 発話時・参照時の時間関係決定規則

語尾形式	動詞の種類	時間関係
ル形	状態動詞，内的動作動詞	S=R
	外的動作動詞，変化動詞	S=R or S → R
テイル形	*	S=R
タ (テイタ) 形	*	R → S

つまり,「走った」では，過去の事象を捉えているため，発話時と参照時の関係は「R → S」となり,「走っている」では，発話時を基準ににおいて捉えているため,「S=R」となる．

また，以下の表3で，語尾形式と動詞の時間的性質から事象時と参照時の関係を抽出する規則を示す．事象時と，参照時の関係は時間概念のアスペクトに相当する．

表 3: 事象時・参照時の時間関係決定規則

語尾形式	動詞の種類	時間関係
ル・タ形 (完了相)	*	E=R
テイル・テイタ形 (継続相)	動作動詞	E(P)=R
	変化動詞	E → R

事象時が E(P) となるものは，動作継続を表すとする。状態動詞はテイル (テイタ) 形を取らないため，空欄になっている．

以下に例文を示す．

(例文 2)　　ここに本がたくさんあった．

例文2は，状態動詞の夕形であるから，発話時と参照時の関係は，表2から「R → S」．事象時と参照時の関係は，表3から「E=R」となる．よって，発話時，事象時，参照時の関係は「E=R → S」となる．

また，時間関係には時間副詞も重要な役割を果たしているため，時間副詞の時間関係の詳細化に用いる．以下に例を示す．

(例文 3)　　彼は明日遊園地に行く．

例文3のように変化動詞「行く」がル形をとると，表2，3の規則から時間関係は次の二通りの候補が得られる．

$$\mathbf{S{=}R{=}E \text{ or } S \rightarrow R{=}E}$$

しかし，時間副詞「明日」に着目すると，発話時の一日後を基準にして事象「行く」を捉えているため「**S → R**」という条件が得られる．よって例文 3 の時間関係は次の一つに絞り込むことができる．

$$\mathbf{S \rightarrow R{=}E}$$

5 時間表現の翻訳

本稿では，日本語側から抽出した時間関係を，英語に変換することで時間表現の翻訳を行う．単文における英語時間表現形式を以下の表 4 に示す [3].

表 4: 英語時間表現形式

	R=S	R → S	S → R
E=R	単純現在	単純過去	単純未来
E → R	現在完了	過去完了	未来完了
R → E	単純未来	過去以後	

なお，事象時が E(P) となっているものに関しては，進行形を取るものとする．空欄は，対応する時間表現がないことを表す．以下に例を示す．

(例文 4)　　彼らは通りを歩いていた．

例文 4 では，動作動詞がテイタ形で使われているため，4.2 節の規則から時間関係は次の通りである．

$$\mathbf{E(P){=}R \rightarrow S}$$

表 4 から対応する英語時間表現は，単純過去形となり，事象時が E(P) であるから，過去進行形に翻訳される．

6 時間表現の制約

時間表現の翻訳にあたっては，原言語である日本語の動詞，副詞から時間関係の抽出を行い，英語側の時間表現を決定してきた．しかし，目的言語側の用法によって時間表現に違いが出る場合がある．

(例文 5)　　娘は高等学校にかよっている．
My daughters goes to the high school.

例文 5 の動詞「かよう」は，本稿の動詞分類では，動作動詞に属するため，5 節から英語時間表現は進行形となる．しかし，実際の文では，単純現在形となっている．これは，一時的な動作も長期的な動作も表す日本語のテイル形と，一時的な動作にしか用いられない英語の進行形の違いにより起こる．

目的言語側の用法に依存した誤りを避けるために，本稿では，目的言語の文法に沿った時間表現に修正する制約条件を導入する．

6.1 結合価パターン

用法に依存した英語時間表現を正しく翻訳するためには，前もって用法を知っておく必要がある．そこで，本稿では，結合価パターン [4] を用いて用法の推定を行い，得られた情報をもとに英語時間表現を修正する．

結合価パターンは，名詞と用言の関係を記述したものであり，文法的，意味的な情報を含んでいる．本稿では，日本語語彙大系 [4] に掲載されている「構文意味辞書」の結合価パターンを使用する．一般の文型と慣用的表現の文型をあわせて約 14,800 件の日本語文型パターンにまとめられている．以下の表 5 に結合価パターンの例を示す．

表 5: 結合価パターンの例

属性変化 (動作 受け身不可)	N1(962 機械) が N2(*) で/から疲れる N1 be fatigued with N2
属性 (状態 受け身不可)	N1(4 人 535 動物) が N2(*) で/に 疲れるている N1 be tired from N2

結合価パターンには，意味属性と，その用言が持つ日本語の文型パターンを示し，対応する英語の文型パターンが記載されている．意味属性の右側には，変形情報が "()" で囲って示してある．変形情報とは，英語の文型パターンが進行形，受け身形に変形しうるのかを示している．(動作) が進行形可，(状態) が進行形不可を表す．また，特定の助動詞をともなって英語の文型パターンが変化する場合には，助動詞を小書きにして後ろに付記してある．また，この結合価パターンでは，格要素の名詞の代わりに一般名詞意味属性で記述されている．一般名詞意味属性とは，名詞の意味的用法を整理，体系化したシソーラスである．

6.2 結合価パターンを用いた制約条件

本研究で作成した制約条件を以下の表 6 に示す．

表 6: 制約条件

	条件	制約
(1) 進行形に関する 制約条件	事象時が E(P)	事象時を E に
	結合価パターンの 変形情報が状態	
(2) 完了形に関する 制約条件	E → R	E=R
	英語文形パターンが 「be 動詞 + adj」	

表 6 の「(1) 進行形に関する制約条件」を例文 5 に適用すると，マッチする結合価パターンの変形情報は (状態) であるから，規則によって事象時は，**E(P)** から **E** に修正される．よって最終的な英語時間表現は単純現在形になる．

また，日本語では,「変化動詞+テイル (テイタ) 形」で変化の結果状態を表すのに対して，英語では，最初から状態として捉え，完了形ではなく,「be 動詞 + adj」型や状態動詞で表すことが多い．例えば「濡れている」は，変化動詞+テイル形であるが,「be wet」のように訳される．そこで「(2) 完了形に関する制約条件」によって，状態そのものを捉えるように修正する．

7 評価実験

本稿で提案した翻訳方法の精度を対訳コーパスを用いて評価する．コーパスには，アンカー和英辞典を用い，ル形・タ形それぞれ 200 文，ル+ダロウ形・テイル (テイタ) 形それぞれ 100 文をランダムに取り出し人手で実験を行った．取り出した文はすべて単文である．以下の表 7 に結果を示す．なおル形については，現在形と未来形の複数の候補が得られる場合がある．

表 7: 翻訳結果 (和英辞典 600 文)

	一意選択正解	複数選択正解	不正解
タ形 (200)	92.5%(185)		7.5%(15)
ル形 (200)	51.5%(103)	41.5%(83)	7.0%(14)
テイル (タ) 形 (100)	65.0%(65)		35.0%(35)
ル+ダロウ形 (100)	86.0%(86)		14.0%(14)
平均 (600)	73.2%(439)	13.8%(83)	13.0%(78)

表 7 の実験結果から正しい英語時間表現が，タ形については，92.5%，ル+ダロウ形では，86.0%求まるという結果を得た．また，候補が一つに決定できない場合があるル形については，一意正解率が 51.5%であったが，候補の中に正解が含まれる割合は，93.0%となった．テイル (テイタ) 形については，一意正解率が 65.0%にとどまった．

8 考察

8.1 正解率について

本稿で提案した翻訳手法の評価実験結果より一意正解率は，73.2%(439/600) であった．しかし，ル形において時間表現を一意に判断できない文を，すべて現在形とすると，86 文中 76 文が正解となり，それを含めた正解率は，85.8%(515/600) であった．また，本稿では英語側の動詞の推定に結合価パターンを用いることにより，正しい時間表現に修正する制約条件を導入した．その結果，テイル (テイタ) 形において正しい英語時間表現に修正できた文が，100 文中 23 文存在した．

8.2 不正解率について

テイル (テイタ) 形については，一意正解率が 65.0%と，もっとも低い結果になった．テイル (テイタ) 形は，英語において進行形・完了形・単純形など様々に訳されるためである．

また，タ形において以下のような不正解文が存在した．

(例文 6)　朝晩めっきり涼しくなった．
It is quite cool these days mornings and evenings.

選択した訳：単純過去形
正解　　：単純現在形

「涼しくなった」という過去の事象を直接的に捉えているのではなく，過去の結果の現在の状態を表しているため，現在形が使われていると思われる．この文の時間表現を正しく訳すためには，発話時・事象時・参照時の他に，話者がどの局面に着目しているかについての情報が必要だと思われる．

9 おわりに

本稿では，日本語時間表現の英語への翻訳を，発話時，事象時，参照時の 3 つの時点の時間関係からなる中間表現を介して行う方法を提案した．その際，日本語側の動詞，時間副詞を時間的性質により分類し，時間関係特定の手がかりとした．また，英語側の用法によって時間表現に違いが現れることがあるため，結合価パターンを用いて用法の推定を行い，正しい時間表現への修正を行った．提案した手法を和英辞典 600 文に適用した結果，複数選択正解を含む正解率 87.0%を得た．

今後は対象とする語尾形式を増やし，時間関係を詳細化することで正解率を向上させる必要がある．

参考文献

[1] 古瀬，中園，野村：アスペクト情報の素性図式化と日本語解析への応用，情報処理学会 87-NL-63-2 (1987).

[2] 溝渕，住友，泓田，青江：日本語時間表現の一解釈法，情報処理学会論文誌 vol.40，No.9，pp.3408-3419 (1999).

[3] Reichenbach,H : *Elements of Symbolic Logic*, Collier-Macmillan，London，England (1947).

[4] 池原，宮崎，白井，横尾，中岩，小倉，大山，林：日本語語彙大系，岩波書店 (1997).

[5] 国立国語研究所: 現代日本語動詞のアスペクトとテンス，秀英出版 (1985).

[6] 的場，池原，村上：日英時間表現の意味と対応関係の解析，情報処理学会 01-NL-146-9，pp.53-60 (2001).