

暦に基づく時間指示表現の意味表現形式

近畿大学 溝渕 昭二

香川大学 安藤 一秋

■ 概要 ■

自然言語処理システムにおいて、時間表現の意味を理解することは非常に重要である。そして、このような機能を実現することにより、人間とコミュニケーションを図る情報処理システム、特に、言語によって示される事態の時間的関係を視覚化したり、形式化された時間情報を言語化したりするなどの高度な情報変換を必要とするシステムの能力を向上させることができる。しかしながら、従来の研究では、そのようなシステムの実現に必要な時間情報の量的側面があまり考慮されていないという問題があった。本論文では、暦に基づく時間指示表現を対象として、その意味を記述するための表現形式である時間指示式を提案する。暦に基づく時間指示表現は、時間表現のうち暦が規定する時間の区割りに基づいて特定の時間を指し示す表現であり、質的な意味と量的な意味を併せもつ。提案する時間指示式は、その量的な意味を最終的にはコレクションと呼ばれる区間の集合で表すという特徴を持つ。これは、参照可能な時間の範囲と精度が限定されているという Leban らの手法が持つ欠点を改善したものである。

■ 1. はじめに ■

言語を媒体として人間とのコミュニケーションを図る情報処理システム、特に、言語によって示される事態の時間的関係を視覚化したり、形式化された時間情報を言語化したりするなどの高度な情報変換を必要とするシステムにおいて、時間表現の意味を正確に理解する機能を実現することは重要である。なぜなら、時間表現は言語が描写する事態の理解に必要な不可欠な時間情報を含み、そして、それは多くの言語表現において明示的にあるいは暗示的に出現するからである。

時間表現には、「2001 年」や「今日」のように暦が規定する時間の区割りに基づいて特定の時間を指し示す表現が存在する。この表現は、数を使って表される内容であり、時間表現の中で最も具体的な内容を提供する。本論文では、この表現を「暦に基づく時間指示表現」、以降では、単に「暦表現」と呼ぶことにする。

暦表現の意味を考えた場合、それは質と量の二つの側面を併せもつ。例えば、「2001 年」の質的な意味とは、それが時間表現あるいは西暦年を表すということであり、量的な意味とは、それが紀元から 2001 番目に位置する年を表すということである。

さて、冒頭で述べたようなシステムにおいて暦表現の表す時間情報を変換する場合、そこには数値化可能な情報が含まれて

いるので、暦表現の表す量的な意味の把握は特に重要になる。しかしながら、従来の研究では、質的な意味ばかりが注目され、量的な意味についてはあまり考慮されていなかった。

そこで、本論文では、暦表現を対象として、その意味の量的側面を記述するための手法を提案する。本手法では、暦表現の指し示す時間の範囲を時間指示式と呼ばれる式で表す。これは、暦表現の示す時間の範囲を最終的にはコレクションという一つの数値情報に変換するものである。したがって、本手法は先に述べたような高度な情報変換に活用することができる。

本論文の構成を次に記す。2 章では、先行研究とその問題点について述べる。3 章では、時間表現における暦表現の位置付けとその意味的な性質について述べる。4 章では、暦表現の意味表現形式である時間指示式を提案する。5 章では、時間指示式の処理系の実装とそれの上での工夫について述べる。6 章では、先行研究において提案された手法と本手法とを比較により評価する。最後に、7 章では、まとめと今後の課題について述べる。

■ 2. 先行研究とその問題点 ■

暦表現を取り扱った研究は、言語学やコンピュータ科学の分野などで行われている。以降では、各分野での研究を概観し、それらが抱える問題点について述べる。

言語学の分野において時間を対象とする研究のほとんどは、時間表現をテンスやアスペクトなどの質的なカテゴリに分類するものである。このような状況において、金子は時間表現の形式的記述方法を提案している [1]。これは、コンピュータで時間表現を取り扱うための基盤を提供するものであるが、暦表現の内容は、例えば、 $t = 「1 月」$ (t は指し示している時間) のように、式の中に言語表記が埋め込まれた形でしか参照されない。つまり、暦表現に限って言えば、その意味を記述するような形式を提案するまでには至っていない。

自然言語処理の分野では、言語表現の解析や生成を行う枠組みの中で時間表現が取り扱われている。暦表現を取り扱った従来の研究を見た場合、そのほとんどは、機械翻訳や情報抽出における質的な意味だけに着目した当該表現の認識方法 [2] (「2001 年」が時間表現であると認識すること) や変換方法 [3] (「今日」を「today」と翻訳すること) の提案である。このような状況において、量的な意味の解釈 (「2001 年」を「紀元から 2001 番目の年」と解釈すること) まで至った研究としては、溝渕らの手法 (M 手法) [4] が挙げられる。この手法では、あらかじめ定めたパターンに値を埋め込むような表現形式が提案されているが、そこに埋め込まれる数値には、特定の暦のものが想定されているので、すべての暦表現の量的な意味を記述

する表現形式としては不十分である。

人工知能やデータベースの分野では、暦に基づいた時間の表現形式を取り扱った研究がなされている。ここで提案された代表的な手法としては、本手法のベースとなった Leban らの手法（L 手法）[5] [6] [7]、これより記述性の高い Bettini らの手法（B 手法）[8] や Ning らの手法（N 手法）[9] などがある。いずれの表現形式においても、暦に基づいた時間を数値情報として捉えることはできるが、参照可能な時間の範囲が限定されていたり（紀元前の時間が参照できなかったりすること）、参照可能な時間の種類が限定されていたり（「…から…まで」にあたる時間帯が参照できなかったりすること）、あるいは、自然言語に対する考慮が欠如しているなどの問題点を持つ。

以上のことより、暦表現を取り扱う、あるいは、それに関連する先行研究が抱える問題点は、これまで提案されてきた表現形式では暦表現の量的な意味を十分に記述できないということである。この原因としては、次のようなことが考えられる。

- (1) 特定の暦に依存
- (2) 参照可能な時間の範囲が限定
- (3) 参照可能な時間の種類が限定
- (4) 自然言語に対する考慮の欠如

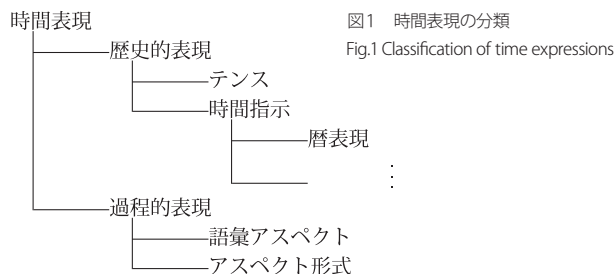
本論文で提案する時間指示式は、暦表現の意味表現形式として、そのような先行研究が抱える問題点を改善するものである。

■ 3. 暦表現 ■

暦表現とは、暦に基づいて事態の歴史的な出現範囲を指し示す表現であり、時間表現の中では、最も具体的な内容を示す。以降では、時間表現における暦表現の位置付けと、その意味的な性質について述べる。

■ 3.1 時間表現における位置付け ■

人間は、自他が体験したり、予期したりする事態の様々な側面を言語によって描写することができる。時間表現とは、そのうちの時間的な側面を描写するために言語に備わった表現形式である。金子 [1] によると、時間表現は、図 1 に示すとおり、



絶対暦表現	2001 年 3 月 26 日, 5 月 26 日, 8 月末日, 日曜日, 毎月の第 1 日曜日, 2003 年の第 7 週, 平成, 春, 1 月と 2 月
相対暦表現	今日, 先月, 来年, 10 年後, 今から 3 年前, 2001 年から 2 年後, 平成 10 年から 4 年間, 2 年ごと, 3 年おき

表1 暦表現 Table 1 Calendrical expressions

歴史的表現と過程的表現に区分され、さらに、歴史的表現がテンス¹と時間指示²に、過程的表現が語彙アスペクト³とアスペクト形式⁴に区分される。

本論文で対象とする暦表現とは、それらのうちの時間指示に下位区分される時間表現である。暦表現の具体例を表 1 に示す。暦表現には、絶対的なものと相対的なものの区分がある。前者は、時間を指示するときに基準とする時間が固定されているものであり、後者は、そうでないものである。

両者を調査すると、絶対暦表現が相対暦表現の一部として出現する場合がある。例えば、「2001 年から 2 年後」という相対暦表現の中に「2001 年」という絶対暦表現が出現しているような場合である。したがって、それらの意味表現形式を考える場合は、先に絶対暦表現から始める必要がある。このような理由から、本論文で対象とするのは、絶対暦表現とする。

[1] 「～た」を代表とする形式で、発話時間との相対的關係によって事態の歴史的な位置を示す表現

[2] 「いつも」や「2003 年」に代表される形式で、事態の歴史的な位置を示す表現

[3] 用言にもともと備わっている事態内部の推移状況を示す表現

[4] 「～ている」を代表する形式で、事態内部の推移状況を示す表現

■ 3.2 意味的な性質 ■

暦表現の主要な役割は、事態の歴史的な出現範囲を指し示すことである。暦表現によって示されるその範囲が備える性質を考えた場合、少なくとも次の五つの性質が挙げられる。

(1) 直線上の時間を指示すること

現代では、時間は限りなく伸びる直線として認識されている [10]。したがって、暦表現の指示する時間は、その全体あるいは一部となる。

(2) 幅を持つ時間を指示すること

Poppel [11] によれば、人間が認知できる時間には個人差はあるものの精度において限界がある。また、日常的に指示される時間や、機器により測定される時間にも精度において限界がある。確かに人間は時間を点として想起することができるが、上記のことを考慮すると、時間を区間とみなすことのほうが妥当である。

(3) 一つ以上の時間を指示すること

暦表現には、「2003 年」や「2003 年 3 月」のように単一の時間を指示するものや、「2003 年と 2005 年」や「毎日」のように複数の時間を指示するものがある。つまり、暦表現は一つ以上の時間を指示する。

(4) 暦をベースにして時間を指示すること

暦とは、ある規則に基づいて分割した時間の区画に名前や番号を与えるものであり、それらを機能語等の他の範疇の語とともに言語表現の中に配置したものが暦表現である。したがって、暦表現は暦をベースにして時間を指示する。

(5) 統合的に時間を指示すること

暦表現を構成する最小成分は、「2001 年」や「3 月」などのように暦の区画に合致する時間を表すものである。その成分によって表せない時間は、「2001 年 3 月 26 日」や「2001 年から 2005 年まで」のように成分同士を記述形式 (YYYY/MM/

DD 等) や機能語と組み合わせることにより指示される。つまり、暦表現についても他の言語表現と同様に、成分を統合して指示する時間を表すという性質を持つ。

■ 4. 暦表現の意味表現形式 ■

暦表現の意味表現形式として時間指示式を提案する。時間指示式とは、次に挙げるコレクション、グラニュラリティー、演算子という三つの成分から構成される式である。式中の演算子が示す各種演算を適用することにより、時間指示式は最終的にコレクションに変換される。

コレクション 暦表現が指し示す歴史的な時間の範囲を示すもの。時間軸上のインターバル(区間)の集合として定義される。**グラニュラリティー**⁵ 時間軸を分割することによってできた時間の区画に番号を与えるもの。これにより、年、月、日などの期間の系列(…、2000 年、2001 年、2002 年、…)の時間的な位置が決まる。

演算子 コレクションとグラニュラリティーに対して各種演算を施すもの。前者に対しては、「/[]」、「&」、「~」、「|」の 4 種類、後者に対しては、「{ }」の 1 種類を定義する。いずれも演算結果としてコレクションを返す。

以降では、時間軸、インターバル、コレクション、グラニュラリティー、時間指示式について述べる。

[5] この概念を表す用語には、グラニュラリティー (granularity) とカレンダー (calendar) が使われるが、後者は暦全体を指す意味が強く、また、最近の論文では、前者の方がよく使われていることから、本論文ではグラニュラリティーと呼ぶことにする。

■ 4.1 時間軸 ■

本手法では、 $-\infty$ から $+\infty$ までの整数を目盛りとして持つ数直線により時間を表すこととし、この数直線を時間軸と呼ぶことにする。時間軸は、前章で述べた意味的な性質のうちの (1) を満足するものである。時間軸に関しては、さらに、基本単位とゼロ時間を導入する。基本単位とは、時間軸に割り当てた時間の単位であり、ゼロ時間とは、時間軸の原点(数直線の 0 の目盛り)に対応付ける時刻のことである。

基本単位は、単に時間軸の連続する目盛り間の幅を決定するだけでなく、最小の単位としての役割も持つ。また、基本単位とゼロ時間によって、時間軸の各目盛りに対応する時刻が変化する。したがって、本手法によって測られる時間の精度や内容は、基本単位とゼロ時間によって定まる。

例えば、基本単位を「秒」とし、ゼロ時間を「2001 年 1 月 1 日 0 時 0 分 0 秒」とすると、時間軸上の 1 に対応する時刻は、「2001 年 1 月 1 日 0 時 0 分 1 秒」となる。また、基本単位を「日」とし、ゼロ時間を「2001 年 1 月 1 日」とすると、時間軸上の 1 に対応する時刻は、「2001 年 1 月 2 日」となる。以降では、特に断らない限り、基本単位とゼロ時間をそれぞれ、「日」と「2001 年 1 月 1 日」として話を進めていく。

■ 4.2 インターバル ■

時間軸上の相異なる二つの点を l, u ($l < u$) とするとき、次のように定義される範囲 $\langle l, u \rangle$ をインターバルと呼ぶ。

$$\langle l, u \rangle = \{p \mid l \leq p < u\}^6$$

インターバルは、意味的な性質のうちの (2) を満足するものである。なお、特別なインターバルとして、 $\langle -\infty, -\infty \rangle$ と $\langle +\infty, +\infty \rangle$ を定義する。例えば、「2001 年」と「西暦」(2001 年 1 月 1 日の 730,485 日前から無期限) が指示する時間には、それぞれ $\langle 0, 365 \rangle$ と $\langle -730485, +\infty \rangle$ が対応する。

インターバルの端点のうち、小さい方を下限、大きい方を上限と呼ぶ。インターバル $\langle l, u \rangle$ を単に I とするとき、その下限と上限は次のように表す。

$$I^- = l, I^+ = u$$

インターバルの下限から上限までの時間を幅と呼ぶ。インターバル I の幅は、 $|I|$ と記し、次のように定義する。

$$|I| = I^+ - I^-$$

二つのインターバルの間には、図 2 に示すような 13 の基本的な関係が成り立つ。これらの関係は Allen [12] によって提案されたものであるが、それらの組み合わせのうち、特に五つのものについては、表 3 に示すような関係として新たに定義する。

二つのインターバルを I_1, I_2 とするとき、それらに対する演算として $\sim$ を次のように定義する。

$$I_1 \sim I_2 = (\min(I_1, I_2), \max(I_1, I_2))$$

ただし、

$$\min(a, b) = a \text{ と } b \text{ のうちの小さいほう}$$

$$\max(a, b) = a \text{ と } b \text{ のうちの大きいほう}$$

とする。

[6] 区間を表す一般的な定義は、 $[l, u] = \{p \mid l \leq p \leq u\}$ であるが、本論文では、「時間の」という意味を添えたいのと、参考文献 [5] でも同様の形式が使われているために、そのような形式を採用した。

$I_1 \text{ equals } I_2$	$(I_1^- = I_2^-) \wedge (I_1^+ = I_2^+)$	
$I_1 \text{ meets } I_2$	$I_1^+ = I_2^-$	
$I_1 \text{ met-by } I_2$	$I_1^- = I_2^+$	
$I_1 \text{ before } I_2$	$I_1^+ < I_2^-$	
$I_1 \text{ after } I_2$	$I_1^- > I_2^+$	
$I_1 \text{ overlaps } I_2$	$(I_1^- < I_2^-) \wedge (I_1^+ > I_2^-) \wedge (I_1^- < I_2^+)$	
$I_1 \text{ overlapped-by } I_2$	$(I_1^- > I_2^-) \wedge (I_1^+ < I_2^+) \wedge (I_1^+ > I_2^-)$	
$I_1 \text{ starts } I_2$	$(I_1^- = I_2^-) \wedge (I_1^+ < I_2^+)$	
$I_1 \text{ started-by } I_2$	$(I_1^- = I_2^-) \wedge (I_1^+ > I_2^+)$	
$I_1 \text{ during } I_2$	$(I_1^- > I_2^-) \wedge (I_1^+ < I_2^+)$	
$I_1 \text{ contains } I_2$	$(I_1^- < I_2^-) \wedge (I_1^+ > I_2^+)$	
$I_1 \text{ finishes } I_2$	$(I_1^- > I_2^-) \wedge (I_1^+ = I_2^+)$	
$I_1 \text{ finished-by } I_2$	$(I_1^- < I_2^-) \wedge (I_1^+ = I_2^+)$	

表2 インターバルに成り立つ基本関係
Table 2 Primitive relations between intervals

precedes	<	before	∨ meets
succeeds	>	after	∨ met-by
in	⊂	starts	∨ during
over	⊃	started-by	∨ contains
intersects	⊄	equals	∨ ⊂ ∨ ⊃ ∨ overlaps ∨ overlapped-by

表3 インターバルに成り立つ複合関係
Table 3 Compound relations between intervals

■ 4.3 コレクション ■

インターバルを要素とする集合において、そこに含まれるインターバルのすべての対に成り立つ関係が、 $<$, $=$, $>$ のいずれかであるとき、この集合のことをコレクションと呼ぶ。コレクションは、意味的な性質のうちの(3)を満足するものである。例えば、「2001年」と「西暦」が指示する時間は、それぞれ $\{<0, 365>\}$ と $\{<-730485, +\infty>\}$ で表される。

コレクションには、要素となるインターバルを有限個含む場合と無限個含む場合がある。前者のコレクションを有限コレクション、後者のコレクションを無限コレクションと呼ぶ。

コレクションについて、 $/[]$, $\&$, $\sim$, $|$ という4種類の演算子を以下に定義する。これらのうち、 $|$ 演算子だけが対称性を持ち、残りの演算子は非対称性を持つ。

■ $/[]$ 演算子 ■

二つのコレクションと正数をそれぞれ、 $A = \{a_1, a_2, \dots, a_l\}$, $B = \{b_1, b_2, \dots, b_m\}$, n とするとき、 $A/[n]B$ は、次に示す手続きによって求められるコレクション C に等しい。

```
C ← ∅;
for(i ← 1; i ≤ l; i ← i + 1) {
  k ← 0;
  for(j ← 1; j ≤ m; j ← j + 1) {
    if(a_i ≥ b_j) {
      k ← k + 1;
      if(k = n) {
        C ← C ∪ {b_j};
      }
    }
  }
}
```

$A/[n]B$ は、「 A に含まれる B の前から n 番目」という表現に相当するものである。例えば、「1月」(2001年だけに限定、つまり、2001年に含まれる月の1番目)の意味は、前述した基本単位とゼロ時間に従うと、

$\{<0, 365>/[1]\{<0, 31>, \dots, <334, 365>\}$

で表すことができる。さらに、これに上記の手続きを施すことによって、

$\{<0, 31>\}$

というコレクションが求められる。

なお、本来の $/[n]$ 演算子は、 n に負数も取ることができる。この場合、 $A/[n]B$ は「 A に含まれる B の後ろから n 番目」という表現に相当する。ただ、そのために必要な手続きが、基本的に n が正数のものと同様なので、ここでは紙面の関係から割愛する。

■ $\&$ 演算子 ■

二つのコレクションをそれぞれ、 $A = \{a_1, a_2, \dots, a_l\}$ と $B = \{b_1, b_2, \dots, b_m\}$ とするとき、 $A \& B$ は、次に示す手続きによって求

められるコレクション C に等しい。

```
C ← ∅;
for(i ← 1; i ≤ l; i ← i + 1) {
  for(j ← 1; j ≤ m; j ← j + 1) {
    if(a_i ≥ b_j) {
      C ← C ∪ {b_j};
    }
  }
}
```

$A \& B$ は、「 A の B 」という表現に相当するものである。例えば、「2001年の1月」(1月は、2000年から2002年までに限定)の意味は、前述した基本単位とゼロ時間に従うと、

$\{<0, 365>\} \& \{<-366, -335>, <0, 31>, <365, 396>\}$

で表すことができる。さらに、これに上記の手続きを施すことによって、

$\{<0, 31>\}$

というコレクションが求められる。

■ $\sim$ 演算子 ■

二つのコレクションをそれぞれ、 $A = \{a_1, a_2, \dots, a_l\}$ と $B = \{b_1, b_2, \dots, b_m\}$ とするとき、 $A \sim B$ は、次に示す手続きによって求められるコレクション C に等しい。

```
C ← ∅;
i ← 1; j ← 1;
while(i ≤ l and j ≤ m) {
  if(a_i < b_j) {
    C ← C ∪ {a_i ~ b_j};
    i ← i + 1; j ← j + 1;
  }
  else {
    j ← j + 1;
  }
}
```

$A \sim B$ は、「 A から B まで」という表現に相当するものである。例えば、「(2001年から2002年までの)1月から3月まで」の意味は、前述した基本単位とゼロ時間に従うと、

$\{<0, 31>, <365, 396>\} \sim \{<59, 90>, <424, 486>\}$

と表すことができる。さらに、これに上記の手続きを施すことによって、

$\{<0, 90>, <365, 486>\}$

というコレクションが求められる。

[7] ここでの $\sim$ はインターバルに対するものである。

■ $|$ 演算子 ■

二つのコレクションをそれぞれ、 $A = \{a_1, a_2, \dots, a_l\}$ と $B = \{b_1, b_2, \dots, b_m\}$ とするとき、 $A|B$ は、次に示す手続きによって求められるコレクション C に等しい。

```

C ← ∅;
i ← 1; j ← 1;
while(i ≤ l and j ≤ m) {
  if(a_i > b_j) {
    C ← ∅; break;
  }
  else if(a_i < b_j) {
    C ← C ∪ {a_i}; i ← i + 1;
  }
  else {
    C ← C ∪ {b_j}; j ← j + 1;
  }
}
if(C ≠ ∅) {
  C ← C ∪ {a_i, ..., a_l, b_j, ..., b_m};
}

```

$A | B$ は、「 A と B 」という表現に相当するものである。例えば、「2001 年 と 2003 年」の意味は、前述した基本単位とゼロ時間に従うと、

$$\{0, 365\} | \{730, 1095\}$$

と表すことができる。さらに、これに上記の手続きを施すことによって、

$$\{0, 365\}, \{730, 1095\}$$

というコレクションが求められる。

■ 4.4 グラニュラリティー ■

グラニュラリティーとは、時間を、周期的に連鎖し、かつ、整数によって識別可能なインターバルの系列に分割するものである。これは、意味的な性質のうちの (4) を満足するものである。グラニュラリティーは、次の性質を満たす整数からインターバルへの写像 G として定義できる。

(1) すべての整数 i について、

$$G(i) \text{ meets } G(i+1)$$

(2) 整数 i, j, n が $i \equiv j \pmod{n}$ のとき、

$$|G(i)| = |G(j)|$$

性質 (1) は、連続する整数に対応するインターバルは隣接することを示しており、性質 (2) は、周期 n で同じ幅のインターバルが出現することを示している。これらの性質を図示すると、図 2 のようになる。

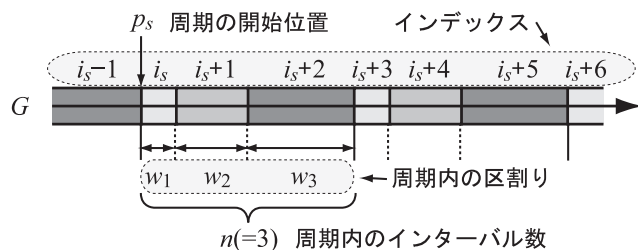


図2 グラニュラリティーの性質

上記の定義は、抽象的なグラニュラリティーの定義であるが、

整数からインターバルへの対応付けまでも含めた具体的なグラニュラリティーの定義は、次のとおりである。

任意に選んだ周期の開始位置を p_s 、 p_s を含むインターバルに対応付ける整数を i_s 、周期内のインターバルの数を n 、周期内の区割りを示すインターバルの幅の列を $w_1, \dots, w_n$ とすると、これらのパラメータから定義されるグラニュラリティー G を

$$\ll p_s; i_s; w_1, \dots, w_n \gg$$

と記し、次のように定義する。

$$G(i) = \langle l_i, u_i \rangle$$

$$\gamma = \begin{cases} i - i_s & (i - i_s \geq 0) \\ i - i_s + 1 & (i - i_s < 0) \end{cases}$$

$$q = \frac{\gamma}{n}, \quad r = \gamma \bmod n$$

[8] $\lfloor X \rfloor = X$ 以下で最大の整数

[9] $\lceil X \rceil = X$ 以上で最小の整数

$$w_0 = 0, \quad \sigma = \sum_{k=1}^n w_k$$

$$\lambda = \begin{cases} \lfloor q \rfloor & (i - i_s \geq 0) \\ \lceil q \rceil - 1 & (i - i_s < 0) \end{cases}$$

$$l_i = \begin{cases} p_s + \lambda \times \sigma + \sum_{k=0}^r w_k & (i - i_s \geq 0) \\ p_s + \lambda \times \sigma + \sum_{k=0}^{r+n-1} w_k & (i - i_s < 0) \end{cases}$$

$$u_i = l_i + w_{r+1}$$

なお、グラニュラリティーにおいて、インターバルに対応付ける整数 i のことをインデックスと呼ぶことにする。

例として、グレゴリアン暦¹⁰の週の系列を W とするとき、これをグラニュラリティーで表すと次のようになる。

$$W = \ll 6; 0; 7 \gg$$

この場合、 $p_s = 6$ 、 $i_s = 0$ 、 $n = 1$ 、 $w_1 = 7$ となるので、グラニュラリティー W において、インデックスが -1 と 1 に対応付けられるインターバルは、上記の定義から次のように計算される。

$$W(-1) = \langle l_{-1}, u_{-1} \rangle = \langle -1, 6 \rangle$$

$$\gamma = 0, \quad q = 0, \quad r = 0$$

$$l_{-1} = 6 + (\lceil 0 \rceil - 1) \times \sum_{k=0}^1 w_k + \sum_{k=1}^{0+1-1} w_k$$

$$= 6 + (-1) \times 7 + 0 = -1$$

$$u_{-1} = -1 + w_{0+1}$$

$$= -1 + w_1 = -1 + 7 = 6$$

$$W(1) = \langle l_1, u_1 \rangle = \langle 13, 20 \rangle$$

$$\gamma = 1, \quad q = 1, \quad r = 0$$

$$l_1 = 6 + \lfloor 1 \rfloor \times \sum_{k=0}^1 w_k + \sum_{k=1}^0 w_k$$

$$= 6 + 1 \times 7 + 0 = 13$$

$$u_1 = 13 + w_{0+1}$$

$$= 13 + w_1 = 13 + 7 = 20$$

グラニュラリティーについて、次の演算子を定義する。

[10] 現在世界標準として用いられている暦。1 年を 365 日とするが、400 の倍数と、100 の倍数を除く 4 の倍数の年を 366 日とする。

■ {} 演算子 ■

グラニュラリティーとインデックスの列をそれぞれ、 G と $i_i, \dots, i_n$ とするとき、 $G\{i_i, \dots, i_n\}$ は次のように定義される。

$$G\{i_i, \dots, i_n\} = \{G(i_i), \dots, G(i_n)\}$$

なお、{} 演算子の特別な記法として以下のものを定義する。

$$G\{*\} = G\{\dots, -1, 0, 1, \dots\}$$

$$G\{-\infty\} = \{\langle -\infty \rangle\}$$

$$G\{+\infty\} = \{\langle +\infty \rangle\}$$

{ } 演算子は、インデックスの列からコレクションを生成するものであり、インデックスの列に整数全体が指定された場合は、「毎…」という表現に相当する意味を表す。例えば、「毎週」の意味は、前述した基本単位とゼロ時間に従うと、

$$W\{*\} = W\{\dots, -1, 0, 1, \dots\}$$

で表すことができる。さらに、これを上記の定義に従って計算することにより、

$$\{\dots, \langle -1, 6 \rangle, \langle 6, 13 \rangle, \langle 13, 20 \rangle, \dots\}$$

というコレクションが求められる。

■ 4.5 時間指示式 ■

時間指示式とは、グラニュラリティーとコレクション、そして、それらに対して定義された演算子から構成される記号列のことである。これは、意味的な性質のうちの (5) を満足するものである。

時間指示式は、優先順位の高いものから順に演算を施すことにより、最終的にコレクションへと変換される。時間指示式に登場する演算子の優先順位を表 4 に示す。いずれの演算子も左結合性を持つので、式中に同じ演算子が複数含まれる場合は、左側から順に演算が行われる。式の演算の順位を変更したい場合は、括弧「()」を使用する。括弧は、それに挟まれた式の優先度を高くする働きを持つ。

優先順位	演算子
高	{ }
↓	/, [], &
↓	~
低	

表 4 演算子の優先順位
Table 4 Precedence of operators

例として、グラニュラリティー Y が

$$Y = \langle 0; 0; 365, 365, 365, 366 \rangle$$

のように定義されているときの「2001 年」に対する時間指示式を次に示す。

$$(Y\{-2001\} \sim Y\{+\infty\})/[2001]Y\{*\}$$

前述した演算の順位に従うと、この式の演算は、次に示すような過程を経て、最終的に解となるコレクションに至る（下線部に対して演算を適用）。なお、式中の $Y\{*\}$ の取扱いについては、次章で述べる。

$$\begin{aligned} & \underline{(Y\{-2001\} \sim Y\{+\infty\})/[2001]Y\{*\}} \\ & \quad \downarrow \\ & \underline{\{ \langle -730485, -730120 \rangle \} \sim \{ \langle +\infty \rangle \} }/[2001]Y\{*\} \\ & \quad \downarrow \\ & \{ \langle -730485, +\infty \rangle \} / [2001]Y\{*\} \end{aligned}$$

$$\begin{aligned} & \downarrow \\ & \{ \langle -730485, +\infty \rangle \} / [2001] \{ \dots, \langle 0, 365 \rangle, \dots \} \\ & \quad \downarrow \\ & \{ \langle 0, 365 \rangle \} \end{aligned}$$

「2001 年」以外のいくつかの時間表現に対する時間指示式を表 5 に示す。この表に示したとおり、時間指示式を使えば、多様な暦表現の意味を記述することができる。

特に、「日曜日」のように無限に存在する時間帯を示すような表現や、「8 月末日」のようにある期間の末尾から時間帯を示すような表現の意味を記述できるという点は、時間指示式が持つ有用な特徴である。

また、時間指示式は、あらかじめ定めた枠組みの中に値を埋め込むような限定的な表現形式ではなく、コレクションとグラニュラリティーをそれらに対する演算子を使って次々と結びつけ、最終的に示したい時間の範囲を記述するような拡張的な表現形式である。拡張的という性質は、多様な意味を記述するために自然言語に備わっている性質であり、ゆえに、時間指示式を使えば、多様な暦表現の意味を記述することができる。

西暦	$Y\{-2001\} * Y\{+\infty\} (= AD)$
平成	$Y\{-12\} * Y\{+\infty\} (= H)$
2001 年 3 月 26 日	$AD/[2001]Y\{*\} \& Y\{*\}/[3]M\{*\} \& M\{*\}/[26]D\{*\}$
平成 15 年	$H/[15]Y\{*\}$
5 月 26 日	$Y\{*\}/[5]M\{*\} \& M\{*\}/[26]D\{*\}$
8 月末日	$Y\{*\}/[8]M\{*\} \& M\{*\}/[-1]D\{*\}$
日曜日	$W\{*\}/[1]D\{*\}$
毎月の第 1 日曜日	$M\{*\}/[1](W\{*\}/[1]D\{*\})$
2003 年の第 7 週	$AD/[2003]Y\{*\}/[7]W\{*\}$
春 (3 月から 5 月まで)	$Y\{*\}/[3]M\{*\} * Y\{*\}/[5]M\{*\}$
1 月と 2 月	$Y\{*\}/[1]M\{*\} Y\{*\}/[2]M\{*\}$

表 5 暦表現に対する時間指示式

Table 5 Time reference equations for calendrical expressions

■ 5. 処理系の実装 ■

時間指示式を解釈し、その解であるコレクションを求める処理系を Java 言語で実装した。

実装に際して問題となるのは、 $Y\{*\}$ のような無限コレクションを表す式の取り扱いである。記憶領域に制限のあるコンピュータ上では、無限コレクションを生成することはできないので、そのような式の解釈には、何らかの対策が必要である。本手法では、現在以下のような対策を施している。

■ (1) /[] 演算子の右側に位置する場合 ■

/[] 演算において、右側に位置するコレクションのインターバルのうち、その計算に必要なのは、左側に位置するコレクションの各インターバルの傘下にあり、かつ、その上限あるいは下限を含む位置から指定された順位の位置にあるものである。例えば、 $AD/[2001]Y\{*\}$ ($AD = Y\{-2001\} \sim Y\{+\infty\}$) の場合、計算に必要な $Y\{*\}$ のインターバルは、 AD のインターバルの下限を含むインターバル $\langle -730485, -730120 \rangle$ から 2001 番目の位置にあるインターバル $\langle 0, 365 \rangle$ までである。このように、演算子の右側に位置するコレクションが無限であっても /[] 演算の計算に必要なのは、そこに含まれる一部

のインターバルであるので、そのようなコレクションを表す式については、優先順位に従った式の展開を保留して、/[] 演算の中で部分的に展開する。

■ (2) 上記以外の場合 ■

フレームと呼ぶインターバルを用意しておき、この場合の無限コレクションに対しては、フレームの範囲に収まるインターバルだけを生成する。フレームの範囲は、自由に変更できる。

■ 6. 評価 ■

2章で述べた四つの手法（B手法、L手法、M手法、N手法）と本手法を、(1) 暦法に対する非依存性、(2) 参照可能な時間の範囲、(3) 参照可能な時間の種類、(4) 言語に対する考慮という四つの項目において比較することにより、本手法の有効性を評価した。この結果、本手法だけが四つの項目をすべて満足することから、本手法が他の手法より有効な表現形式であるといえる。評価結果の一覧を表6に示す。また、項目別の評価内容は次のとおりである。

評価項目	B手法	L手法	M手法	N手法	本手法
暦法に対する非依存性	○	○	×	○	○
参照可能な時間の範囲	○	△	△	○	○
参照可能な時間の種類	○	△	△	△	○
自然言語に対する考慮	×	○	○	×	○

○… 適当, △… 一部不適当, ×… 不適当

表6 評価結果一覧 Table 6 The result of evaluation for 5 methods

■ (1) 暦法に対する非依存性 ■

暦表現が基盤とする暦には、地域や年代によりさまざまなものが存在する。このため、暦表現の意味表現形式は、暦法に対して非依存である必要がある。

この観点から五つの手法を評価した場合、M手法のみが不適当と判断される。これは、M手法のみがグレゴリアン暦を前提としているからである。残りの手法はいずれも、暦を定義する仕組みを持つため、さまざまな暦法の暦表現の意味を記述することができる。

■ (2) 参照可能な時間の範囲 ■

我々が現在使う暦表現では、過去も未来も制限なく目的の時間を参照できる。このため、暦表現の意味表現形式は、時間軸上のすべての時間帯を参照できる必要がある。

この観点から五つの手法を評価した場合、L手法とM手法が不適当と判断される。これは、両者がともにある時点から未来へと向かった範囲内にある時間しか参照できないからである。

■ (3) 参照可能な時間の種類 ■

暦表現は、暦の区画に合致しない時間も参照できる。このため、暦表現の意味表現形式は、そのような時間を参照するための形式を備えておく必要がある。

この観点から五つの手法を評価した場合、本手法とB手法を除く3つの手法が不適当と判断される。これは、それらの手法が、暦の区画に合致する時間しか参照できないもの（L手法、M手法、N手法）、だからである。

■ (4) 自然言語に対する考慮 ■

暦表現は、自然言語一般に見られるように統合的な性質を持つ。このため、暦表現の意味表現形式は、意味を成す成分が結合して全体の意味をなすような形式を持つほうが好ましい。

この観点から五つの手法を評価した場合、B手法とN手法が不適当と判断される。これは、両者がともに一般的な時間情報のモデルとして考案されたものであり、自然言語との関連を考慮していないからである。

■ 6. おわりに ■

本論文では、暦表現の意味表現形式として時間指示式を提案した。暦表現の意味は質と量の二つの側面を併せもつが、従来、自然言語処理の分野ではそれを質的な側面から捉える傾向にあった。本論文で提案した時間指示式は、最終的にコレクションという数値情報に変換することができるので、量的な側面から暦表現の意味を捉えることが可能になった。

今後期待されている知的な情報処理システムにおいては、人間と円滑にコミュニケーションするための技術が必要とされている。人間を取り巻く情報には、常に時間情報が含まれているので、本手法はそのような技術の一要素としての活用が期待できる。

今後の課題としては、まず、処理系の実装の箇所ですべて無限コレクションへの対策である。本手法では、あらかじめフレームを設定することにより対処したが、計算に必要なインターバルがフレームに含まれない場合もあるので、そのような場合の対策が必要である。また、本手法の適用範囲は、絶対暦表現に限られているので、「今日」や「2日後」などの相対暦表現についても拡張する必要がある。さらに、本手法により暦表現の意味を網羅的に記述できるかどうかの調査もコーパス等を使って行う必要がある。

【参考文献】

- 金子亨：言語の時間表現，ひつじ書房 (1995).
- 関根聡：テキストからの情報抽出，情報処理学会会誌，Vol. 40, No. 4, pp.370 - 373 (1999).
- 池原悟，村上仁一，的場和幸：日英機械翻訳のための時間表現の意味と対応関係の解析，情報処理学会論文誌，Vol.44, No. 2, pp.451 - 465 (2003).
- 溝渕昭二，住友徹，泓田正雄，青江順一：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).
- Leban, B., McDonald, D.D. and Forster, D.R.: A Representation for Collections of Temporal Intervals (1986).

6.Chandra, R., Segev, A.and Stonebraker, M.: Implementing Calendars and Temporal Rules in Next Generation Databases (1994).

7.Terenziani, P.: Integrating Calendar Dates and Qualitative Temporal Constraints in the Treatment of Periodic Events, IEEE Transactions on Knowledge and Data Engineering, Vol.9, No.5, pp.763 – 783 (1997).

8.Bettini, C., Mascetti, S., Wang, X.S.: Supporting Temporal Reasoning by Mapping Calendar Expressions to Minimal Periodic Sets, Journal of Artificial Intelligence Research, Vol.

28, pp.299 – 348 (2007).

9.Ning, P., Wang, X.S.and Jajodia, S.: An Algebraic Representation of Calendars, Annuals of Mathematics and Artificial Intelligence, Vol.36, No.1 – 2, pp.5-38 (2002).

10. 荒井喬: 時間の矢, 地人書館 (1987).

11. 田山忠行, 尾形敬次: 意識のなかの時間, 岩波書店 (1995).

12.Allen, J.F.: Maintaining Knowledge about Temporal Intervals, Communications of the ACM, Vol.26, No.11, pp.832 – 843 (1983).

溝渕 昭二
Mizobuchi Shoji

近畿大学講師, 徳島大学大学院博士後期課程修了 (システム工学専攻), 博士 (工学)。元理化学研究所協力研究員。これまでに, 自然言語処理, 情報検索の研究に従事。現在, Web マイニングに関心を持っている。情報処理学会, 電子情報通信学会, 人工知能学会, 言語処理学会各会員。

安藤 一秋
Ando Kazuaki

香川大学准教授, 徳島大学大学院博士後期課程修了 (システム工学専攻), 博士 (工学)。自然言語処理, 情報検索の研究に従事。情報処理学会, 電子情報通信学会, 人工知能学会, 言語処理学会, 教育システム情報学会, 日本データベース学会各会員。