暦に基づく時間指示表現に対する意味表現形式の提案

溝渕 昭二†　安藤 一秋‡

† 近畿大学理工学部情報学科
mizo@info.kindai.ac.jp
‡ 香川大学工学部信頼性情報システム工学科
ando@eng.kagawa-u.ac.jp

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．
本稿では，暦に基づく時間指示表現（暦表現）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．時間指示式は，暦表現の持つ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．

A Semantic Representation of Time Expressions Based on Calendars

Shoji Mizobuchi†　Kazuaki Ando‡

† Department of Informatics, School of Science and Engineering, Kinki University
mizo@info.kindai.ac.jp
‡ Department of Reliability-based Information Systems Engineering,
Faculty of Engineering, Kagawa University
ando@eng.kagawa-u.ac.jp

In natural language processing systems, it is very important to understand the meaning of time expressions. Introducing such a function to the systems leads to improve the ability of information processing systems which communicate with human beings. However, the quantitative aspect of time expressions required for the systems has hardly been taken into consideration in the conventional research.
In this paper, we propose a semantic representation of time reference expressions based on calendars, called a time reference equation, which indicates a specified time using a calendar. The feature of a time reference equation is to represent the quantitative meaning of such an expression by a collection of intervals of time calculated from the equation. The representation is the extension of a series of semantics representations originated by Leban.

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非常に重要である．

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「暦法的な時間指示表現」，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」の意味の質的側面とは，それが時間表現あるいは西暦年であるということであり，量的側面とは，それが紀元から 2001 番目に位置する年であるということである．

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，質的側面だけに着目した当該表現の認識方法 [1] （「2001 年」が時間表現であると認識すること）や変換方法 [2] （「今日」を「today」と翻訳すること）の提案であり，量的側面までも含めた理解方法（M 手法）[3] (「2001 年」を「紀元から 2001 番目の年」と解釈すること) についてはあまり議論されていないのが現状である．その理由は，情報抽出や機械翻訳における暦表現の認識や変換が質的側面のみを必要とするからである．

しかしながら，冒頭で述べたような高度な情報変換を行うシステムにおいて，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，意味の量的側面にも着目する必要がある．

暦表現の量的な意味に関係する研究としては，本手法のベースとなった Leban らの手法（L 手法）[4], これより記述性の高い Bettini らの手法（B 手法）[5] や Ning らの手法（N 手法）[6] などがある．しかしながら，そこで提案されたいずれの手法においても，限定された範囲の時間しか記述できない，あるいは，暦の区画にあった時間しか記述できないなどの問題点を持つ．

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，従来，時点や期間，あるいは，それらの集合という形で別々に記述されてきた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．

本論文の構成を次に記す．2 章では，暦表現の意味的な性質について述べる．3 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．4 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，5 章では，まとめと今後の課題について述べる．

2　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．暦表現の具体例を次に記す．

絶対暦表現	2001 年 3 月，8 月末日，日曜日，2003 年の第 7 週，1 月と 2 月
相対暦表現	今日，10 年後，2001 年から 2 年後，2 年ごと

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．本論文で対象とするのは，絶対暦表現とする．

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．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**
   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．このようなことから，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」のように単一の時間を指示するものや,「2003 年と 2005 年」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，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，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

によって指示される．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．

3 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，時間軸上において定義されるコレクション，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．

以降では，時間軸，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．

3.1 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，これを時間軸と呼ぶ．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．

例えば，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．

3.2 コレクション

コレクションは，暦表現が示す時間の範囲を表すものである．コレクションの定義の前にまずその要素となるインターバルを定義する．

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．

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

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，$\langle -\infty\rangle = \langle -\infty, -\infty\rangle$ と $\langle +\infty\rangle = \langle +\infty, +\infty\rangle$ を定義する

二つのインターバルの間には，表 1 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[7] 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，特に二つのものについては，表 2 に示すような関係として定義する．

表 1 インターバルに成り立つ基本関係

I_1 equals(=) I_2	
I_1 meets I_2	
I_1 met-by I_2	
I_1 before I_2	
I_1 after I_2	
I_1 overlaps I_2	
I_1 overlapped-by I_2	
I_1 starts I_2	
I_1 started-by I_2	
I_1 during I_2	
I_1 contains I_2	
I_1 finishes I_2	
I_1 finished-by I_2	

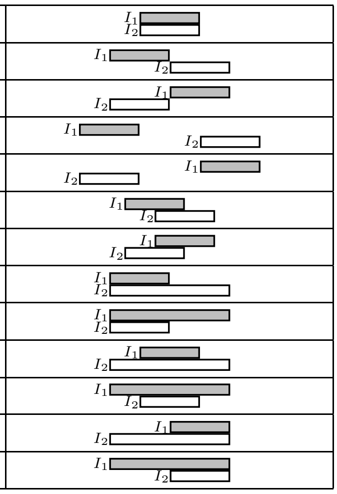

表 2 インターバルに成り立つ複合関係

preceeds	$\prec$	before $\vee$ meets
succeeds	$\succ$	after $\vee$ met-by

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，$\prec$，$=$，$\succ$ 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 $\{\langle 0, 365\rangle\}$ と $\{\langle -730485, +\infty\rangle\}$ で表される．

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．

/[] 演算子

$A/[n]B$ 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,

$$\{\langle 0, 365\rangle\}/[1]\{\langle 0, 31\rangle, \cdots, \langle 334, 365\rangle\}$$

という式で表される．/$[n]$ は，コレクション A の各インターバルに対して，それに覆われるコレクション B のインターバルのうち，先頭 ($n < 0$ のときは末尾) から n 番目のインターバルからなるコレクショ

ンを返す．したがって，先の式に演算を施すと，

$$\{\langle 0,31\rangle\}$$

というコレクションが求められる．

& 演算子

$A\&B$ 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,

$$\{\langle 0,365\rangle\}\&\{\langle -366,-335\rangle,\langle 0,31\rangle,\langle 365,396\rangle\}$$

という式で表される．& は，コレクション A の各インターバルに対して，それに覆われるコレクション B のインターバルからなるコレクションを返す．したがって，先の式に演算を施すと,

$$\{\langle 0,31\rangle\}$$

というコレクションが求められる．

∼ 演算子

$A\sim B$ 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,

$$\{\langle 0,31\rangle,\langle 365,396\rangle\}\sim\{\langle 59,90\rangle,\langle 424,486\rangle\}$$

という式で表される．∼ は，コレクション A の各インターバルから，それと対応関係 (基本的には順序が同じ位置) にあるコレクション B のインターバルまでを覆う最小のインターバルからなるコレクションを返す．したがって，先の式に演算を施すと,

$$\{\langle 0,90\rangle\ \langle 365,486\rangle\}$$

というコレクションが求められる．

| 演算子

$A\mid B$ 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,

$$\{\langle 0,365\rangle\}\mid\{\langle 730,1095\rangle\}$$

という式で表される．| は，コレクション A のインターバルとコレクション B のインターバルからなるコレクションを返す．したがって，先の式に演算を施すと,

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

というコレクションが求められる．

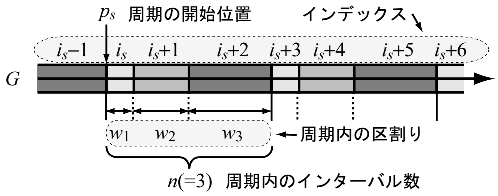

図 1 グラニュラリティー

3.3 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．

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，周期内の区切りを示すインターバルの幅の列を $w_1,\cdots,w_n$ とすると（図 1 参照)，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

$$\langle\langle p_s;i_s;w_1,w_2,\ldots,w_n\rangle\rangle$$

と記す．また，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び，グラニュラリティー G のインデックス i に対応付けられたインターバルを $G(i)$ で表すことにする．

例として，週を表すグラニュラリティー W を次に示す．

$$W=\langle\langle 6;0;7\rangle\rangle$$

$$\begin{array}{ccccccc}
\cdots & W(-2) & W(-1) & W(0) & W(1) & W(2) & \cdots \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\
\cdots & \langle -8,-1\rangle & \langle -1,6\rangle & \langle 6,13\rangle & \langle 13,20\rangle & \langle 20,27\rangle & \cdots
\end{array}$$

グラニュラリティーについて，次の演算子を定義する．

{} 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 $i_1,\cdots,i_n$ とするとき,

$$G\{i_1,\ldots,i_n\}=\{G(i_1),\cdots,G(i_n)\}$$

と定義する．なお，{} 演算子の特別な記法として以下のものを定義する．

$$G\{*\}=G\{\cdots,-1,0,1,\cdots\}$$
$$G\{-\infty\}=\{\langle -\infty\rangle\},\quad G\{+\infty\}=\{\langle +\infty\rangle\}$$

{} 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,

$$\begin{aligned} W\{*\} &= W\{\cdots,-1,0,1,\cdots\} \\ &= \{\cdots,\langle -1,6\rangle,\langle 6,13\rangle,\langle 13,20\rangle,\cdots\} \end{aligned}$$

で表すことができる．

3.4 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，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．

例として，グラニュラリティー Y が

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

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.

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

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）.

$$\begin{gathered} (\underline{Y\{-2001\}} \sim \underline{Y\{+\infty\}})/[2001]Y\{*\} \\ \downarrow \\ \underline{(\{\langle -730485,-730120\rangle\} \sim \{\langle +\infty\rangle\})}/[2001]Y\{*\} \\ \downarrow \\ \{\langle -730485,+\infty\rangle\}/[2001]\underline{Y\{*\}} \\ \downarrow \\ \underline{\{\langle -730485,+\infty\rangle\}/[2001]\{\cdots,\langle 0,365\rangle,\cdots\}} \\ \downarrow \\ \{\langle 0,365\rangle\} \end{gathered}$$

「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 3 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．

表 3 暦表現に対する時間指示式

西暦	$Y\{-2001\} \sim Y\{+\infty\}(= AD)$
2001 年 3 月	$AD/[2001]Y\{*\}\&Y\{*\}/[3]M\{*\}$
8 月末日	$Y\{*\}/[8]M\{*\}\&M\{*\}/[-1]D\{*\}$
日曜日	$W\{*\}/[1]D\{*\}$
2003 年の第 7 週	$AD/[2003]Y\{*\}/[7]W\{*\}$
1 月と 2 月	$Y\{*\}/[1]M\{*\} \mid Y\{*\}/[2]M\{*\}$

4 評価

1 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 4 に示す．また，項目別の評価内容は次のとおりである．

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

表 4 評価結果一覧

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

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，本手法を除く 4 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），あるいは，その定義だけを行うもの（B 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．

5 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．この式は，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，量的な側面からも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は，人間とコミュニケーションするための技術が必要とされる．人間を取り巻く環境には，常に時間情報が含まれているので，本手法はそのような場面において活用できる．

今後の課題としては，本手法の適用範囲の拡張が挙げられる．本手法は，絶対暦表現に限られているので,「今日」や「2 日後」などの相対暦表現の意味についても記述できるように拡張する必要がある．

参考文献

[1] 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).

[2] 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).

[3] 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).

[4] Leban, B., McDonald, D. D. and Forster, D. R.: A Representation for Collections of Temporal Intervals (1986).

[5] Bettini, C., Wang, X. S. and Jajodia, S.: A General Framework for Time Granularity and Its Application to Temporal Reasoning, *Annals of Mathematics and Artificial Intelligence*, Vol. 22, No. 1-2, pp. 29–58 (1998).

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

[7] Allen, J. F.: Maintaining Knowlegdge About Temporal Intervals, *Communications of the ACM*, Vol. 26, No. 11, pp. 832–843 (1983).