

テキストにおける イベント間の時間的関係の抽出

東 正造, 山田篤, 松本 裕治
奈良先端科学技術大学院大学 情報科学研究科

自然言語で記述された文章では、テンス、アスペクトなどの様々な言語的掛かりによって、対象世界での出来事間の時間的な関係が表現されている。本研究では、世界内での出来事をイベントとして捉え、これらの手掛かりを用いて、文章からイベント間の時間的関係を抽出する方法を提案する。本論文で提案するイベント間の時間的関係抽出の方法は、(1)文章からイベント計算可能な記述に変換し、(2)そのイベント記述に対して、時間に関する言語情報、常識的な知識である一般知識、記述分野の領域知識とその知識を統合する Event Calculus の枠組を使って、イベント間の相対的な時間関係を抽出し、(3)最後に各イベントが発生する絶対的な時間を決定する、という3つの処理から構成される。本論文の手法は、一般知識と領域知識を分離しているため、領域知識を変えるだけで、様々な分野の文章においても扱えるという特徴がある。このような手法を用いて、原子力ユニットの故障記述に対して、実際にイベント間の時間的関係を抽出し、その結果、本方法の有効性を確認した。

Analysis of Temporal Relation in Texts

Shozo Azuma, Atsushi Yamada and Yuji Matsumoto
Graduate School of Information Science, Nara Institute of Science and Technology

In natural language texts, temporal relations among events are expressed by various linguistic clues. In this paper, we propose a method to extract temporal relations among events, which is composed of three processes: (1) translating events from natural language texts to formal representation, (2) determining temporal relation among events using a framework of Event Calculus, which integrates linguistic information and the common sense of general knowledge and domain knowledge, and (3) deciding the absolute time of each event. Since domain knowledge is separated from general one, we can determine temporal relation among events concerning phrase of a various field just by changing the domain knowledge. Using this method, we extracted temporal relations among events from texts about an atomic power plant. The result shows the usefulness of the method.

1 はじめに

近年、過去における大量の情報を効果的に活用するため、計算機を利用することが多くなっている。しかし、人間向きに蓄積された情報をプレーンテキストの形でそのまま計算機に蓄積しても運用を行うのは難しい。そこで計算機上である対象に関する事件記述や事故・故障記述のような事象を記述したテキストの理解が行えれば、事象間の関係が明らかになるため、プレーンテキストのままより、情報を系統的に蓄積すること

が可能となり、効果的な運用に役立てることもできる。本稿では、このような背景のもと、計算機上での事象の内容を認識させる機構の構築を目的とする。テキストの内容を理解するには、文章が記述する対象物の状態がどのように変化していくかを理解する必要がある。このような文章の内容を理解することは、状態の変化の流れを理解することであり、そのために、状態の流れを変化させるイベントに注目する。自然言語表現によって表現される対象世界の状態には、ある時点における静的状態とイベントの発生によって変化する動的

状態がある [山本 92]。ゆえに、時間的な概念を含む自然言語表現の意味する対象世界の状態変化を捉えるためには対象世界で発生したイベント間の時間的な関係を推論する必要がある。しかし自然言語の文章は基本的には、イベントの発生順序どおりに記述されておらず、また全ての時間的な関係が明示的に書かれているわけではない。したがって、本稿では、言語の表面に現れる時間的な関係を示す表現である語彙アスペクト、テンス、アスペクト、時間指示語とそれ以外に言語の意味的な一般知識、その記述対象の領域知識まで考慮し [Ter93]、それらを Event Calculus [KS86] を用いて統合して、イベント間の時間的な関係の抽出を行う。さらに原子力ユニットの故障記述に対して、実際にイベント間の時間的な関係の抽出を行い、その結果に基づいて、本研究の手法の有効性について評価する。

2 言語知識ベース

2.1 テンス情報

発話時点基準にしてイベントを位置付ける文法形式がテンスである。用言は、それぞれに固有の語彙アスペクトをもつが、語彙アスペクト自身は発話時点と直接に関係できないので、特定の時制標識と結合してはじめて発話時点と結び付く。現代の日本語においては二つの時制標識ル形とタ形を持つ。それぞれの時制の指定は発話時点 t_0 と会話が目している状況 (以下関心状況と呼ぶ) を R とすると、ル形なら $t_0 \leq R$ (R より t_0 は時間的に同時か前) であり、タ形なら $R < t_0$ (t_0 より R は時間的に前) である。

2.2 語彙アスペクト情報

用言は、ある状況を言語的に多彩に表現するもので、用言自体は、瞬間的な動作、継続的な動作、そして状態など様々な状況に関する情報を持っている。語彙アスペクトとは、そのような用言の意味に現れる状況を規定したものである。これは用言が表示する状況である表示状況、その状況を作る前提となる状況である前状況、その表示状況の後に続く含意、結果等を表す後状況と表示状況の開始と終了を示す時点から規定される。

用言の表示状況は持続的なもの (dur) と瞬間的なもの (min) の二種類に分類できる。これにより表示状況が持続的なものは、表示状況がある時間幅をもつので、その前後に開始点と終了点が前状況、後状況を分けるものとして必要であるが、表示状況が瞬間的なものは表示状況がある時点として捉えられるので、それ自身

が前状況、後状況を分けるものとなる。前状況は、表示状況と同じかまたは異なる状況という二種類に分類できる。後状況は、表示状況と同じかまたは異なる状況に加えて、その結果や効果を表すものという四種類に分類される。このことから [金子 95] に基づき、表示状況・前状況・後状況とその境界によって、語彙アスペクトを分類すると次の表 2.1、および 2.2 の 10 種類に分類される。V は用言の持つ状況を表し、nil は開始、または終了点が明確でないことを示す。res は表示状況 V の結果を表し、eff は表示状況 V の効果を表す。

表 2.1: 持続的な表示状況を持つデータ

識別子 id	前状況 before	開始点 start	表示状況 indicate	終了点 end	後状況 after
d1	V	nil	V	nil	V
d2	V	nil	V	end	not V
d3	not V	start	V	end	not V
d4	not V	start	V	end	not V(eff)
d5	not V	start	V	nil	V(res)
d6	not V	start	V	nil	V(eff)

表 2.2: 瞬間的な表示状況を持つデータ

識別子 id	前状況 before	表示状況 indicate	後状況 after
m1	not V	V	not V
m2	not V	V	not V(eff)
m3	not V	V	V(res)
m4	not V	V	V(eff)

具体例で考えると、持続的な動作を持つ (d3) サ変動詞「運転する」は、

aktionsalt(key(運転), before(not 運転),
start(start), indicate(運転),
end(end), after(not 運転))

という語彙アスペクトで表され、瞬間的な動作を持つサ変動詞 (m4) 「停止する」は、

aktionsalt(key(停止), before(not 停止),
indicate(停止), after(停止))

という語彙アスペクトで表される。以上のデータ構造によって、実際には図 2.1 のような時間関係が得られる。

2.3 アスペクト情報

語彙アスペクトだけの場合は、関心状況と表示状況は一致する。アスペクトは、その関心状況を変更する働きがある。アスペクトは語彙アスペクトのタイプに

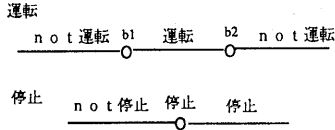


図 2.1: 語彙アスペクトから得られる時間的關係

よって、接続するものと接続しないものがあり、また語彙アスペクトのタイプによって、同じアスペクトでも表示する状況が変わってくる。次に表 2.1, 及び 2.2 の語彙アスペクトと主なアスペクトによって、関心状況がどう変わるか示す。

- 開始点 (start) に変更
d1 から d6 までの継続動詞+(はじめる、だす、かける、かかる)
- 終了点 (end) に変更
d1 から d6 までの継続動詞+(おわる、おえる、きる)
- 動作中 (indicate) の内部に変更
d1 から d6 までの継続動詞+(つづける、つつある、ている、)
- 後状況 (after) に変更
m1 から m4 までの瞬間動詞+(ている)、m3,m4 の瞬間動詞+(つづける)
- 繰り返し動作に変更
m1,m2 の瞬間動詞+(つづける)

2.4 時間指示語情報

時間指示語とは、表示状況の成り立つ時間を指示するものでテンスより具体的かつ詳細に表示状況の成立時間を指示する。ここでは、次の四種類の時間指示語を考える [金子 95]。

- 基本的直示例:「今」
関心状況が発話時点 t_0 を含む。
- 派生的直示例:「今日、去年」
この指示語による発話時点を起点とした指示される時間を関心状況の絶対時間とする。
- 歴法的指示例:「9月20日9時」
この指示語による時間を関心状況の絶対時間とする。

- 述語的指示例:「学校に来る前」
述語的指示によって、主にイベント間の相対的時間関係を得る。

時間指示語の前のイベントの関心状況を r_1 、後のイベントの関心状況を r_2 として、述語的指示による相対的時間関係の抽出に関してまとめたものが、表 2.3 である。表での時間関係は、関心状況のものであるから、二つの関心状況間の関係は、時点ではなく、時間幅をもったもの間の関係であるので、その関係は、Allen の時間論理 [All84] に基づいて得られる。

表 2.3: 述語的指示におけるイベント間の時間的關係

種類	時間関係	類語
原因・理由	$r_1 < r_2$	から、ので、より、ため
順序関係 (1)	$r_1 > r_2$	前、直前、以前
順序関係 (2)	$r_1 < r_2$	後、次、以後、以来、以降
順序関係 (3)	$r_1 \supset r_2$	中、うち、途中、最中、間
時点表示	$r_1 < r_2$	時、際、瞬間

2.5 常識的な言語情報

2.5.1 一般知識ベース

一般知識ベースには、二つの用言間の因果関係、排他関係、同義関係を蓄積する。いずれの関係も二つの言葉の関係なので、表 2.4 のように蓄積する。このとき、各イベントの対象物が同じと仮定する。そうすると例えば、表 2.4 から入学というイベントが卒業というイベントの前に成立することが得られる。その他の関係も同様に表 2.4 から得られる。

表 2.4: 一般的な関係

関係	表現形式	時間関係	例
因果	$\text{cause}(x, y)$	$x < y$	$\text{cause}(\text{入学}, \text{卒業})$
排他	$\text{exclusive}(x, y)$	$x < y$ or $x > y$	$\text{exclusive}(\text{就寝}, \text{起床})$
同義	$\text{same}(x, y)$	$x \equiv y$	$\text{same}(\text{中断}, \text{停止})$

但し、時間関係における x, y はそれぞれの関心状況を示す。また、同義関係については、 x, y が同一イベントであることを示す。

次にイベント計算ではイベントの前後の状態を用いる必要があるので、イベントの動作と対応する状態を取る必要がある。そのため、イベント・状態対応関係を蓄積する。

$\text{correspond}(\text{イベント } x, \text{状態 } y)$

これからイベント x の時の前後の状態は、 y という状態を持つか、not y という状態を持つかである。

最後にイベントが取る格に別のイベントが割り当てられる場合の規則について述べる。この時、動詞がどのような格を取り得るかが重要である [小泉 89]。例えば、「停止する」のような動詞では、必ず動作主格、対象格、源泉格、目標格のいずれかに別のイベントを持つ。「停止する」で考えた場合、この言葉は、動物や機械、乗物が前提として動いていた状態があるか、または水道・ガス・電気・電話などが供給されていた状態でないが使われない。つまり前提として、運動や供給という状態が存在し、それを停止するわけである。したがって、「停止する」が持つ動作主格・対象格のイベント状態と停止するというイベントでは、停止するというイベントが必ず後に起こり、しかもその前のイベントの動作主格・対象格などの深層格に割り付けられたものに対して「停止する」というイベントが起こるのである。「停止する」というイベントを e_1 、そのイベント記述の深層格に割り付けられたイベントを e_2 とすると

$$e_2 < e_1$$

という関係が得られる。他のイベントを格に取るものも同様に行う。

2.5.2 領域知識ベース

領域知識ベースには、領域固有の知識を蓄積するので、蓄積方法を一般的に規定することはできない。ただ、この領域知識を変更するだけで、いろいろな分野に対して、時間的関係の抽出を行える。そこで本研究で扱うテキストの一つである原子力に関する事象記述のための原子炉に関する知識について述べる。原子力関連では、どの機器がどの機器と接続関係にあるかが重要である。例えば、母線への電力供給が止まれば、その後、それと接続関係にある保障母線への電力供給が停止することが分る。接続関係が分ってれば、どちらが先に電力供給が停止したかが求められる。そのため、本研究では、接続関係をリンクで表現して、領域知識として、用いる。

[所外電力, 母線, 保障母線, 非常用 DG]

このように、例えば、所外電力からの電力供給が停止すると、母線が停止し、その後、保障母線が停止、非常用 DG が起動すると、保障母線が電力供給が再開され、その後母線への電力供給も回復するというような関係が得られる。

3 時間的関係の抽出

イベント間の時間的関係抽出の全体の流れを図 3.1 に示す。

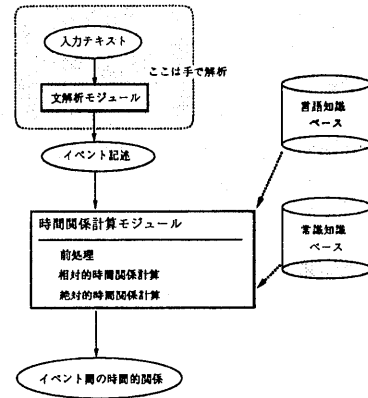


図 3.1: イベント間の時間的関係の抽出の流れ

3.1 文解析モジュール

テキストを形態素解析し、イベント記述を生成する。イベント記述は計算機上でイベント計算可能なデータ構造である。本稿では、文章構成を上から順に、文章、文、イベント、深層格、時間格という 5 階層構造として捉える。4 層、5 層について、ここで考えているイベント記述は、行為、識別子、動作主格、経験者格、道具格、対象格、源泉格、目標格、場所格、時間格で、また時間格の詳細は、テンス、アスペクト、時間指示語 (4 種) である。時間格以外の各スロットは、イベント同定とどの対象物に注目しているかを理解するために使われる。例えば、「原子炉運転 (1) 中、所外電力 751 からの 34.5kV 電力供給 (2)...(略)」という文の (1) のイベント記述は、次のように書ける。

((行為 運転)
(識別子 exp1)
(対象格 原子炉)
(時間格 (時間指示語 (述語の指示 中))))

3.2 時間関係計算モジュール

時間関係計算は、前処理、相対的時間関係計算、絶対的時間関係計算という三つのフェーズからなる。

3.2.1 前処理

先程のイベント記述は、入力から直接得られるものである。複合イベントを一度に記述したものを単一イベントに変換したり、同じイベントを異なる表現をしたものを同一イベントとして同定したりする必要がある。前の例では、例えば「停電」という複合イベントがあれば、電力供給 (a) と供給停止 (b) という二つのイベント (a,b) に分割する。同一イベントの同定は、イベント記述のロットが同一である時、同一イベントとする。

3.2.2 相対的時間関係計算

● Event Calculus の運用

実際、二つのイベント間の時間関係を統合計算するために、Event Calculus [KS86] の枠組を使うが、そのままは扱えない。Event Calculus では、状態を変化させるイベントの時点だけを考慮したイベント計算を行っている。しかし一般にイベントを示す用言、特に動詞は、語彙アスペクトの説明でも明らかなように状態を示すもの、瞬間的な動作を示すもの、継続的な動作を示すものとそれ自身が状態を表したり、ある一定の時間幅を持った動作を示したりと様々である。したがって、文が注目している点は、状態を変化させる時点だけでなく、その状態に注目することもある。このためイベントという概念の他に、文章が注目している状況である関心状況という概念を加えた。例えば、「私は働いたので、病気になった。」という文は、イベントの時点だけに注目して考えると図 3.2の左のように解釈され、関心状況を加えたほうは、図 3.2の右のような解釈になる。

また Event Calculus では、“Mary is hired as lec-

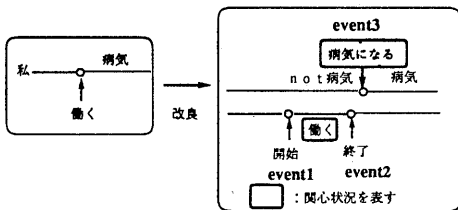


図 3.2: 改良例

turer.”という文があると Mary が雇われるというイベントの後に lecturer という状態を保持するというような規則を作って、イベントの前や後の状態を規定しているが、本研究では、語彙アスペクトにイベントの他に状態を含めて表現しているため、語彙アスペクトと

イベント記述からイベントと状態を規定する。イベント記述と語彙アスペクトから状態記述を作り上げる方法を図 3.3に示す。このような運用の工夫をすること

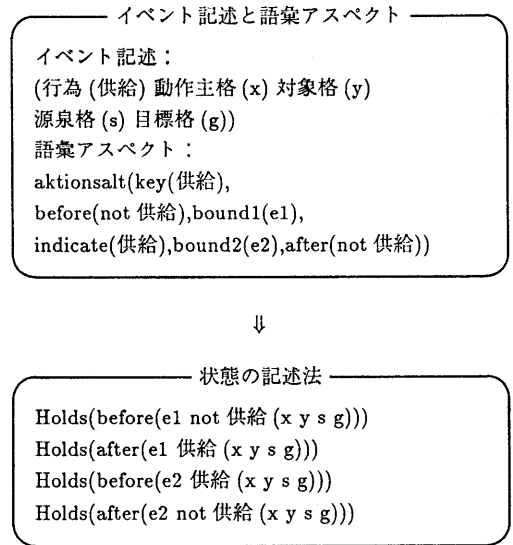


図 3.3: 状態の規定

で前章の知識から得られる二つのイベント間の時間関係が統合可能となる。

● 相対的時間関係計算の処理

上の状態の記述法と前章の知識を用いた処理順序は次のようになる。

- step1: イベント記述を入力
- step2: 語彙アスペクトを利用して、イベント・状態・関心状況を抽出
- step3: アスペクトによる関心状況の変更
- step4: テンスによる発話時点との時間関係の抽出
- step5: 時間指示語による時間関係の抽出
- step6: 接続語による時間関係の抽出
接続語の規則は、述語の指示と同様
- step7: 一般知識による時間関係の抽出
- step8: 領域知識による時間関係の抽出
- step9: Event Calculus によるイベント間の時間計算

3.2.3 絶対的時間関係計算

このモジュールでは、時間指示語のうちの基本的直示、及び歴法的指示によりイベントの絶対時間を獲得し、またイベント間の時間、すなわち状態が継続する

時間を領域知識により獲得する。そして一つの絶対時間軸上でのイベントの位置、すなわちイベント発生時間を計算する。

step1: 時間指示語による歴法的時間決定

基本的直示と歴法的指示から決定する。

step2: 領域知識を利用した状態継続時間の決定

領域知識から状態継続時間を決定する。

以上によりイベント間の時間的關係を抽出する。

3.3 解析例

原子力ユニットの故障記述を例として以下に解析の結果を示す。

入力文

1994年9月21日、10時32分、原子炉運転(1)中、所外電力751からの34.5kV電力供給(2)が停止した(3)ため、母線12Bの電力供給(4)が停止し(5)、B系列保障母線16と17への電力供給(6)も中断した(7)。そのため、設計通りB非常用DGが自動起動し(8)、母線16と17への電力供給(9)を再開した(10)。電力751の停電(11)は、樹木が送電線を直撃した(12)ことが原因であった。

以上の入力文に対して、前節の手順に沿って、解析すると図3.4の結果が得られる。

ここでe3とe12、e4とe13が同一イベントとして、認識されるのは、イベント(11)の「停電」を単一イベントにして分離したイベント記述とイベント(2)、(3)のイベント記述において行為と源泉格スロットが同一になり、イベント同定が行える。

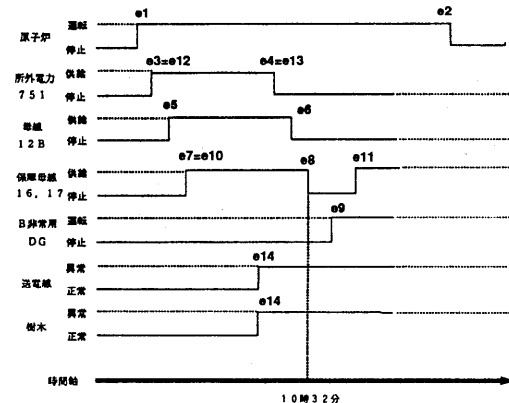


図 3.4: イベント計算結果

3.4 評価

人が事象記述テキストを読む場合と比較して、イベント間の時間的な順序關係を抽出するという観点からは、かなり一致していると思われる。しかし、イベントの絶対時間における時間的關係という観点からは、あまり良い結果とは言えない。これを改善するには、各イベントに関連する状態の継続時間の正規分布を取り、正規が一番高いところの値をデフォルト値として用いるなどの工夫が必要となると考えられる。

4 結論

本稿での時間的關係抽出の方法では、言語的な手がかりを基に状態とその変化に注目して、言語表現から得られる部分的な時間に対する情報を統合する枠組を示した。また、時間に関する言語情報を言語知識ベース、及び常識的な時間情報を一般知識ベース、そして記述分野に依存した時間情報を領域知識ベースに分けて蓄積したので、様々な分野の記述に対しての処理を可能とした。

参考文献

- [山本 92] 山本唯史. 日本語の動作表現中の空間的概念の解析—日本語物語文のプロット理解—. 情報処理学会自然言語処理研究報告, Vol. 87, No. 7, pp. 51-58, Jan 1992.
- [金子 95] 金子亨. 言語の時間表現. ひつじ研究叢書(言語編) 第7巻. ひつじ書房, 1995.
- [小泉 89] 小泉保, 船城道雄, 本田. 日本語基本動詞用法辞典. 大修館書店, 1989.
- [All84] J.F. Allen. Towards a general theory of action and time. *Artificial Intelligence*, No. 23, pp. 123-154, 1984.
- [KS86] Robert Kowalski and Marek Sergot. A logic-based calculus of events. *New Generation Computing*, Vol. 4, pp. 67-95, 1986.
- [Ter93] Paolo Terenziani. Integrating linguistic and pragmatic temporal information in natural language understanding: The case of "When Sentence". *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, Vol. 2, pp. 1304-1309, 1993.