

機械可読辞書の語釈文を用いた動詞間の時間関係の解析

森部 正二郎, 奥村 学

email:{moribe,oku}@jaist.ac.jp

北陸先端科学技術大学院大学情報科学研究科情報処理学専攻

1 はじめに

テキストの抄録作成など、テキストの知的処理ではテキストの内容を理解することが重要な役割を果たす。テキストの理解には、テキスト中の動詞間の時間関係を明らかにしておくことが必要である。しかし、通常のテキストは必ずしもイベントの発生順に記述されているとは限らない。

動詞間の時間関係を解析する手法には、時制情報を利用する方法(手法1)や、イベント間の時間関係に関する一般知識を利用する方法(手法2)がある。手法1は複文中の動詞間の時間関係を解析する場合などに用いられるが、時制情報だけでは曖昧が生じることが多い。手法2は、動詞間の時間関係を、あらかじめ人手で作成した一般知識として持っているの、時制情報に関係なく時間関係を決定できる。しかし一般知識を人手で作成する問題点として、客観性を欠いた知識になる可能性がある、手間がかかるので、大規模な知識を作成できない、などが考えられる。そこで、本研究では日本語基本動詞辞書 IPAL[情処 87]の語釈文から、一般知識として利用可能な知識を、自動的に獲得することを目的とする。

2 時間関係の定義

本研究における時間関係とは、例えば「鶏が卵を暖める」「ひよこが生まれる」における2つの動詞が、「暖める」の後「生まれる」が起こる、というような関係を示す。このような時間関係を13に分類したのが表1である[Allen 83]。

表 1: 時間幅を有する二つのイベントの時間関係

X before Y	XXX	Y after X	YYY	X meets Y	XXX	Y met-by X	YYY
X overlaps Y	XXXX	X during Y	XXX	Y contains X	YYYYY	Y finished-by X	YYYYY
X starts Y	XXX	X equal Y	XXX				
Y started-by X	YYYYY						

本研究では動詞間の時間関係として、表1の分類を用いる。

3 知識の獲得

語釈文からどのように一般知識を獲得するかについて解説する。例として、次の文中の「暖める」と「孵す」の時間関係を調べる。

例：にわとりが卵を暖める。にわとりが卵を孵す。

「暖める」の語釈文は次のような複文である。

暖めるの語釈文：

卵を孵すために親鳥が抱いて、温度を高くする。

この場合、「孵す」は「暖める」の語釈文中に存在するので、「暖める」とその語釈文中の動詞の間の時間関係を解析しておけば、時間関係が決定できる。

語釈文は

- 動詞を別の単語で言い替えたイベント・状態(主節の動詞)
- 動詞の原因、目的、方法などを表すイベント(従属節の動詞)
- 動作主や対象の状態を表す格要素

から構成されている。語釈文からこれらのイベントや状態を抽出し、時間関係でリンクした意味ネットワークを、動詞の語義を表す語義ネットワークとする。この語義ネットワークを全ての動詞に対して作成し、一般知識として蓄積する。

4 語義ネットワークの作成

IPAL 中の全ての語釈文から語義ネットワークを自動的に作成するシステムを図1のように作成した。

入力された語釈文は図中の前処理において形態素、構文解析される。動詞の語義は KNP[Kurohashi 92]で行なわれるが、語義が解析できない動詞は語義ネットワークから削除し、語義が誤っている場合は実験段階に人手で修正する。

残りの小節で、図中の処理について解説する。

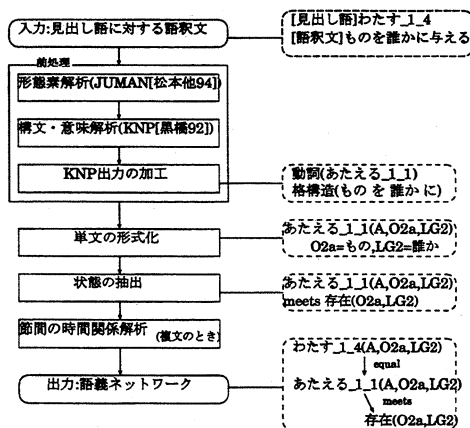


図 1: 語義ネットワーク作成システム

4.1 単文の形式化

図 1 中の語釈文を例として、単文の形式化について解説する。

本研究では動詞をあたえる_{1.1}(A,O2a,LG2)のように表す。‘あたえる_{1.1}’は動詞の読みと語義番号を示す(以降見出し語と呼ぶ)。括弧内の記号は、述語素と呼ばれる IPAL 中の情報で、動詞が持つ格の制約(動詞と名詞句、名詞句と名詞句の関係)を表す。本研究ではこの述語素を、格要素を表す変数として扱う。つまり単文は動詞の述語素 O2a にヲ格の名詞句「もの」、LG2 に二格の名詞句「誰か」を入れ換えた形式になる。

4.2 状態の抽出

文中の格には動作主や対象を表す格(ガ,ヲ)と、動作主や対象のイベント前後などの状態を表す格(カラ,ニ,へ,...)がある。述語素はこれらの格の種類を示す。語義ネットワークにおける状態抽出とはこれら二つの格(述語素)から動作主や対象の状態を獲得することである。

あたえる_{1.1}(A,O2a,LG2)の場合、対象を表す述語素 O2a と対象のイベント語の場所を表す LG2 から存在(O2a, LG2)という状態が抽出される(時間関係は「あたえる meets 存在」。)

4.3 節間の時間関係の解析

語釈文が複文である時、節間の時間関係を、時制情報で解析する。通常、時制情報から時間関係を解析するには、情報として各節の時制情報を用いる。し

かし語釈文の場合、基本的に非過去形でアスペクトがない。そこで、節間の接続形式と、従属節の時制情報だけから時間関係を解析する。

入力文が複文の場合は図 2 のような語義ネットワークとなる。

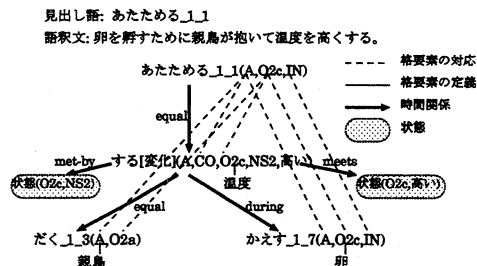


図 2: あたえる_{1.1}の語義ネットワーク

5 語義ネットワークの評価システム

前節のような過程で作成した語義ネットワークが、手法 2 において一般知識として利用可能であるかを評価するために、語義ネットワークを用いた時間関係解析システムを、図 3 のように作成した。

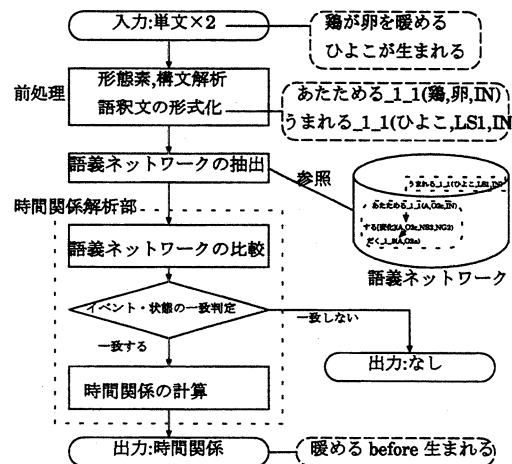


図 3: 時間関係解析システム

二つの単文を入力とし、図 1 と同様の前処理を行なう。ここで、動詞の語義が間違っていた場合、人

手で修正する。決定した語義から語義ネットワークを参照する。以降の処理は次の節で解説する。

5.1 時間関係解析部

解析例として、
「子供が怪しい男につぶてを打つ」
「子供が怪しい男につぶてを当てる」
の二つの単文を入力した場合の処理を解説する。「打つ」と「当てる」の語義ネットワークは図4のようになる。(図中 A=子供, O2a=つぶて, LG2=男)

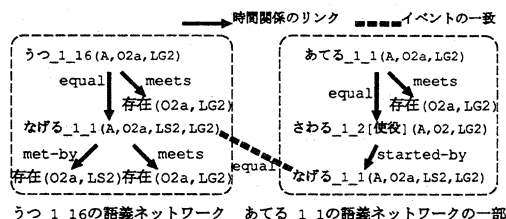


図 4: 時間関係解析例

まず、両方の語義ネットワーク間に一致するイベントか状態(イベント優先)があるか調べる。一致条件は、見出し語が一致していることと、格要素が一致あるいは類似していることである。格要素が類似しているかは分類語彙表[分類 96]の5桁の分類番号が一致しているかで判断する。¹

一致していた場合そのイベント・状態を同時とみなす。最後に見出し語間の時間関係のリンクをたどって時間関係を計算する。「打つ」と「当てる」の場合は「打つ」から「当てる」へたどる。矢印が逆方向な場合(started-by)は、矢印が「当てる」の方向になるように時間関係を変える(→ starts)。時間関係の計算は、二つの連続する時間関係をあらかじめ計算しておく(例: A starts B, B meets C = A before C) [Allen 83] 方法で処理する。

図4の場合の時間関係は「打つ starts 当てる」と出力される。一致するイベントが複数あった場合は複数の時間関係が出力される。

6 評価実験

6.1 評価方法

ここでは、作成した語義ネットワークが一般知識として、どの程度利用可能かを評価する。評価は、ま

¹例えば、ひよこ='15620 6 2 2', ひな='15620 6 2 1'の場合、最初の5桁が一致しているので類似している

ず時間関係解析システムで、依存関係のある動詞間の時間関係を決定できるかを調べる。入力に対する正解は、人間が決めておく。次に、時制情報を用いた手法1で時間関係を解析し、システムの結果と比較する。

入力はIPAL中の動詞だけを用いた、省略のない単文とする。さらに単文化した状態で、人間が時間関係を決定できるものを選択する。本研究では次のような文を入力とした。

- A. 1992 年度毎日新聞記事 (100 の複文を単文化した 100 組の単文)
- B. 物語文 (ごんぎつね)(35 の複文を単文化した 35 組の単文)
- C. 人手で作成した単文 (32 組の単文)

6.2 結果

語義ネットワークを用いた時間関係解析システムで、入力文を解析した結果を表2に示す。(再現率=正解/入力数, 適合率=正解/解答数)

表 2: 時間関係解析結果の精度 (動詞の語義修正後)

入力文	入力数	解答数	正解	再現率	適合率
人手で作成	32	16	12	0.375	0.75
新聞記事	100	15	11	0.11	0.733
物語文	35	21	19	0.6	0.905

どの入力に対しても適合率が高いが、物語文以外は再現率が低い。この結果から、語義ネットワークが表す知識は、物語文のような入力に対して有効だと推測される。

次に新聞記事と物語文を、手法1を用いて解析する。その結果をシステムの出力結果と比較した(表3)。(括弧内は物語文だけを入力とした数)

表 3: 手法1とシステムの出力結果の比較

		システム			合計
		正解	不正解	出力なし	
手法1	正解	20(15)	1(1)	64(13)	85(29)
	不正解	10(4)	5(0)	35(2)	50(6)
合計		30(19)	6(1)	99(15)	135(35)

システムが出力した解答36のうち、30(83%)が正解だった。さらに正解のうち10(33%)は手法1で不正解だった時間関係を決定できた。しかし、手法1でもシステムでも解析できない時間関係が40(全体の30%)も残っている。

これを物語文に限定すると、システムは手法1で不正解だった時間関係の67%を決定し、両方の手法で解析できない時間関係は2(全体の6%)だけだった。

7 考察

ここでは、本システムが物語文で、時間関係の多くを決定できた理由について考察することにより、語義ネットワークがどのようなドメインに対して有効なのかを検討する。まず物語文の特徴について考える。

特徴の一つは、文が平易であることである。物語文の格要素は「人」「物」「場所」を表す簡単な名詞句である。物語文以外の入力では、動詞間の時間関係が、名詞句の表す意味に依存することがあった。

例1. 大統領がタラップ上に現れる。礼砲が響く。

この場合、動詞間に依存関係はない。しかし、名詞句「礼砲」はイベント「現れる」に対して行なわれ、その結果「響く」が生じている。このような例は新聞記事に多く、語義ネットワークでは決定できない。

また、物語文の動詞は移動、発生、消滅、といった状態変化を表す動詞が多く、特殊な語義を持たない。

例2. 傷に薬を塗る。傷に包帯を巻く²。

この場合、「巻く」は「傷を覆う、隠す」などの意味を暗示しているが、語釈文には記述されていないため解析できなかった。「ごんぎつね」では、このような動詞は見当たらなかった。

もう一つの特徴として、物語文は、二つの単文で同じ名詞句の状況変化を述べている。今回対象とした「ごんぎつね」は、主に「ごん」または「兵十」の状況変化を表すイベントから構成される。これにより動詞間の時間関係は、イベント前後の「ごん」「兵十」の状態の一致から決定される。

次に、前述のような物語文の特徴から時間関係を決定できた例を示す。

例3. 兵十が川でびくを持つ。兵十が川から土手に上がる。

例4. ごんが森の中に穴を掘る。ごんが穴に住む。

例1は両方とも「兵十」が動作主の文である。最初の文は「持つ」と同時に兵十が川にいることを示し、後の文は兵十が川から土手に移動することを示す。二つのイベントは「兵十が川にいる」という状態が一致することから時間関係が決定する。

例2は名詞句「穴」が最初の文で発生している。後の文では場所を表す「穴」が存在していることがイベントの発生条件となる。従って「穴」の発生後に起こることが決定する。

²「巻く」の語釈文：身体の部分の周りに布状物をぐるりと付ける。

以上の点から、本システムは、ある主題(物語文では「ごん」「兵十」など)の状況変化を表し、平易な文から構成される文章を対象とした場合に有効である。

8 おわりに

本研究では日本語基本動詞辞書 IPAL 中の語釈文を形式化して、見出し語の意味を表す語義ネットワークを作成した。そして、語義ネットワークが一般知識として利用できるかを調べるために、語義ネットワークを用いた時間関係解析システムを作成し、新聞記事や物語文中の動詞の時間関係を解析した。物語文を入力とした場合、入力数の約6割の時間関係を決定できた。さらに手法1で解析できなかった時間関係の66%を決定できた。本研究で作成した語義ネットワークは、物語文のような、ある主題の状況変化を表し、平易な文から構成される文章に対して有効であることを考察した。

参考文献

- [東 96] 東 正造, 山田 篤, 松本 裕治. 「テキストにおけるイベント間の時間関係の抽出」 情報処理学会 112-19, 1996, pp133-138.
- [富浦 89] 富浦洋一 「国語辞典の語義文からの動詞の上位-下位関係の抽出」 情報処理学会 73-3, 1989, pp133-138.
- [情処 87] 「計算機用日本語基本動詞辞書 IPAL-解説編」 情報処理振興事業協会技術センター, 1987.
- [松本他 94] 松本裕治, 黒橋禎夫, 宇津呂武仁, 妙木裕, 長尾真. 「日本語形態素解析システム JUMAN 使用説明書 version3.0beta」 京都大学工学部長尾研究室, 1994.
- [Kurohashi 92] Dynamic Programming Method for Analyzing Conjunctive Structure in Japanese. COLING 92, 1992, pp23-28.
- [Allen 83] Allen, J.F. "Maintain knowledge about temporal intervals", Commun. ACM, Vol26, No.11, pp.832-843, 1983.
- [分類 96] 中野洋. 「分類語彙表」形式による語彙分類表(増補版) 第1分冊, 第2文冊, 国立国語研究所, 1996.
- [金子 95] 金子亨. 「言語の時間表現」 第7巻, ひつじ書房, 1995.