

Title	機械可読辞書中の語釈文を用いた動詞間の時間関係の解析
Author(s)	森部, 正二郎
Citation	
Issue Date	1997-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1040
Rights	
Description	Supervisor:奥村 学, 情報科学研究科, 修士

修 士 論 文

機械可読辞書中の語釈文を用いた 動詞間の時間関係の解析

指導教官 奥村学 助教授

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

森部正二郎

1997 年 2 月 14 日

要 旨

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。テキスト理解には、テキスト中の動詞間の時間関係を明らかにすることが必要であるが、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。

本研究では動詞間の時間関係の解析を目的とする。動詞間の時間関係を解析する方法の一つとして、人間が持っているイベント間の時間関係に関する一般知識を知識ベースとして用意しておき、それを利用する方法がある。しかし一般知識を人手で作成するには手間がかかるうえ、客観性を欠いた知識になる可能性がある。本研究では、イベント間の因果関係や順序関係に関する知識を計算機用日本語辞書 IPAL の語釈文から半自動的に導く。語釈文は動詞がどのような状態・状態の変化を表しているのかを抽出するのに利用される。抽出した状態を時間関係でリンクした意味ネットワークを用いて動詞が表す語義を表し、一般知識として蓄積する。また語釈文から抽出した一般知識を用いて動詞間の時間関係を導く方法を示す。本研究で提案した手法を計算機上に実装し、例文を処理させた結果、物語のような時間的推移の明確な文に対して約 60 % の時間関係を決定できた。

目次

1	序論	1
2	動詞間の時間関係の解析手法	3
2.1	時制情報を用いた動詞間の時間関係の解析手法	3
2.2	一般知識を用いた時間関係の解析手法	6
3	研究に利用した辞書	8
3.1	計算機用基本動詞辞書 IPAL	8
3.2	分類語彙表 (増補版)	11
4	語釈文の形式化	12
4.1	述語素を用いた単文の形式化	13
4.2	語義ネットワークにおける動詞間の時間関係の解析	14
4.2.1	時間関係の定義	14
4.2.2	語彙アスペクト	16
4.2.3	述語素の組合せによる状態の抽出	17
4.2.4	見出し語と定義動詞の時間関係	19
4.2.5	語釈文における複文の節間の時間関係の解析	21
4.3	動詞間の格要素の一致	23
4.4	使役形、受身形の動詞に対する述語素の拡張	24
4.5	語義ネットワーク	25
5	時間関係の解析方法	27
5.1	述語素が表す名詞句の意味素性	29

5.2	意味素性への代入と例外への対策	30
5.3	イベント、状態の類似性の判定	30
5.4	時間関係の計算	32
5.5	複数のイベント・状態が一致した時の処理	33
5.6	語義ネットワークの再帰的参照	34
6	システムの実装と評価	36
6.1	語釈文の形式化システムの実装	36
6.1.1	入力	38
6.1.2	出力：語義ネットワークのデータ構造	38
6.1.3	KNP の語義決定結果	40
6.1.4	語義決定できなかった原因と対策	42
6.1.5	決定された語義の精度	44
6.2	時間関係解析システムの実装と評価	46
6.2.1	実装	46
6.2.2	実験の方法	46
6.2.3	実験結果	47
6.2.4	考察	48
7	結論	53
7.1	研究内容のまとめ	53
7.2	研究の結論と今後の課題	53
A	日本語基本動詞辞書 IPAL に関する資料	58
A.1	意味素性	58
A.2	意味分類	60
B	システムに関するデータ	61
B.1	語義ネットワークのデータ形式の例	61

第 1 章

序論

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。テキスト理解には、テキスト中の動詞間の時間関係を明らかにすることが必要であるが、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。

本研究では動詞間の時間関係の解析を目的とする。動詞間の時間関係を解析する方法には

手法 1. テンス・アスペクト・時間指示語などの時制情報を利用する

手法 2. イベント間の時間関係に関する一般知識を利用する

などの方法がある。[手法 1] は、複文中の動詞間の時間関係を解析する場合に用いられるが、時制情報だけでは曖昧さが生じることが多い。[手法 2] を用いることの利点は、時制情報に関係なく時間関係を決定できることである。従って、[手法 1] による曖昧さの解消や、2 つの単文における動詞間の時間関係も解析することができる。2 つの単文の時制が一致していて、「その後」、「その前に」のような時間指示語がない場合、[手法 1] で時間関係は解析できない。

東らは時間に関する言語情報と一般知識や記述分野の専門知識の両方を用いている [東 96]。さらにそれらの知識を統合する 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ている。また Alterman は知識として動詞間の時間関係を意味ネットワークとして用いている [Alterman 85]。この研究ではその意味ネットワークから、入力テキスト中のイベント間の関係だけでなく、各イベントに関連して生じるイベントと

その関係も出力する。東条は、一般知識を用意する代わりに各イベントに時間的特性を与える方法を提案している [東条 95]。イベント間の関係は時間特性に基づいて定義された時間区間どうしの関係の付け方によって決定される。

これらの研究と我々の研究の違う点は、一般知識や専門知識を知識ベースとしてあらかじめ人手で作成しないという点である。人手で作成する場合、手間がかかる上、客観性を欠いた知識になる危険性がある。我々はこれまで、時制表現を考慮した解析に加え、動詞間の依存関係を導くために発生や消滅などの状態変化を表す動詞に注目してきた [森部 95]。この手法は動詞の分類と状態変化する対象を調べるだけで依存関係がわかるが、この条件を満足する文は非常に少ない。本研究ではより多くの依存関係に対応するために、一般知識として計算機用日本語基本動詞辞書 IPAL [情処 87] に記述されている語釈文を情報源として用いる。

本研究では、イベント間の依存関係や順序関係に関する知識を計算機用日本語基本動詞辞書 IPAL の語釈文から半自動的に導く。語釈文は、

- 動詞を別の単語で言い替えたイベント・状態 (文の主動詞)
- 動詞の原因、目的、方法などを表すイベント (文の従属節)
- 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、動詞の語義が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。

本研究では、まず語釈文から語義ネットワークを自動的に生成する手法を示し、次に語釈文から抽出した一般知識を用いて動詞間の時間関係を導く方法を示す。

最後に、本研究で提案した手法を計算機上に実装し、実際にどの程度時間関係を決定できるかを検証する。

第 2 章

動詞間の時間関係の解析手法

2.1 時制情報を用いた動詞間の時間関係の解析手法

時制情報を用いた動詞間の時間関係を解析する手法として、[東 96] の例を挙げる。

用いている時制情報は大きく分けて「テンス情報」、「語彙アスペクト情報」、「アスペクト情報」、「時制指示語情報」の四つである。ここで、発話時点を t_0 、会話が注目している状況 (以下関心状況) を R とすると四つの情報は次のような役割を果たす。

(1) 語彙アスペクト情報

動詞が瞬間的な動作 (瞬間動詞) か、継続的な動作 (継続動詞) か、状態を表すかなどを、分類した情報。動詞が表示する状況 (表示状況) とその前提条件となる状況 (前状況)、結果を表す状況 (後状況) からなる (詳しくは 4.2.2 節を参照)。通常、表示状況 = 関心状況 (R) となる。図 2.1 に例を示す。「運転する」は継続的な動詞の例で、表示状況が時間幅を持つ。「停止する」は瞬間的な動詞の例で、表示状況は時間幅を持たない。

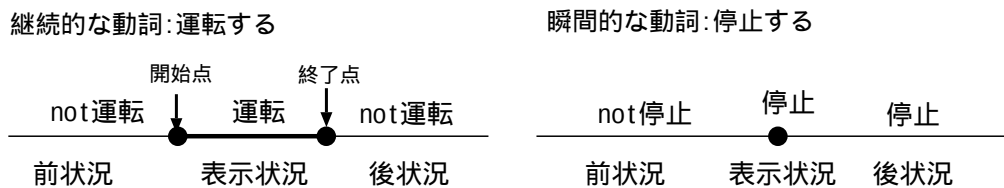


図 2.1: 語彙アスペクトから得られる時間関係

(2) テンス情報

t_0 と R の時間関係を決める。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。ル形なら $t_0 \leq R$ (R より t_0 は時間的に同時か前)、タ形なら $R < t_0$ (R より t_0 は時間的に後) である (図 2.2)。

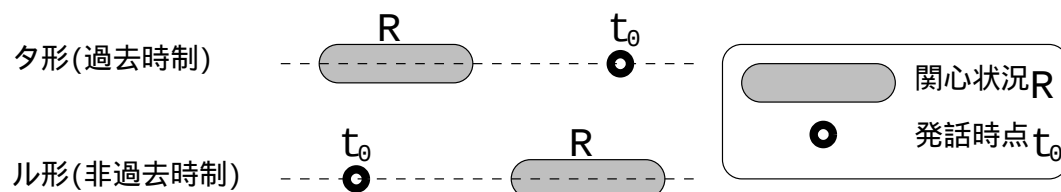


図 2.2: テンス情報による発話時点と関心状況の時間関係

(3) アスペクト情報

固有の語彙アスペクトの動詞と接続することにより関心状況 (R) を表示状況の内部、後状況などに変更する役割を果たす。

図 2.3 中の「運転する + (はじめる)」は継続動詞にアスペクト「はじめる」が接続した場合の例である。この場合、関心状況は表示状況の左端(「運転する」の開始点)に変更される。

「停止する」+「ている」は瞬間動詞にアスペクト「ている」が接続した場合の例である。この場合、関心状況は後状況に変更される。

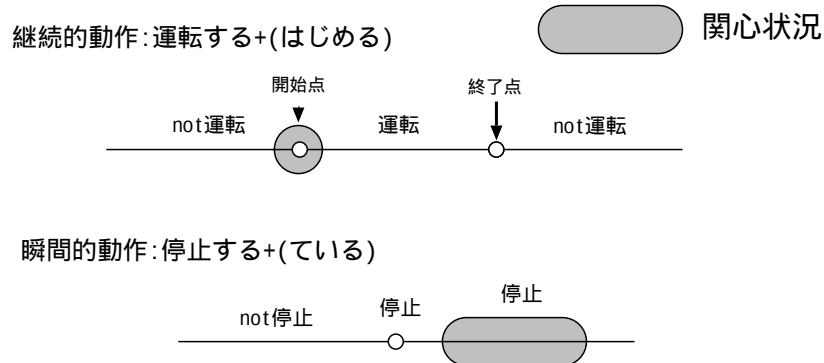


図 2.3: アスペクト情報による関心状況の変更

他にも次のような場合に関心状況は変更される。

- 開始点に変更 (例：図 2.3 「運転する」 + 「はじめる」)
継続動詞 + (はじめる、だす、かける、かかる)
- 終了点に変更
継続動詞 + (おわる、おえる、きる)
- 動作中 (表示状況) の内部に変更
継続動詞 + (ている、つつある、つづける)
- 後状況に変更 (例：図 2.3 「停止する」 + 「ている」)
瞬間動詞 + (ている)、m3,m4 の瞬間動詞 (表 4.3 参照) + (つづける)

(4) 時制指示語情報

テンスより具体的に関心状況 (R) の成立する時間を指示する。

- 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」)
発話時を起点とした関心状況の絶対時間を表す
- 歴法的直示 (例：「9 月 20 日 9 時」)
関心状況の絶対時間を表す

- 述語的指示 (例 : 「前、以降」)
イベント間の相対的時間関係を示す

これらの情報を用いて、2つの単文の発話時を同時としたときの動詞間の時間関係は図2.4のように表される。

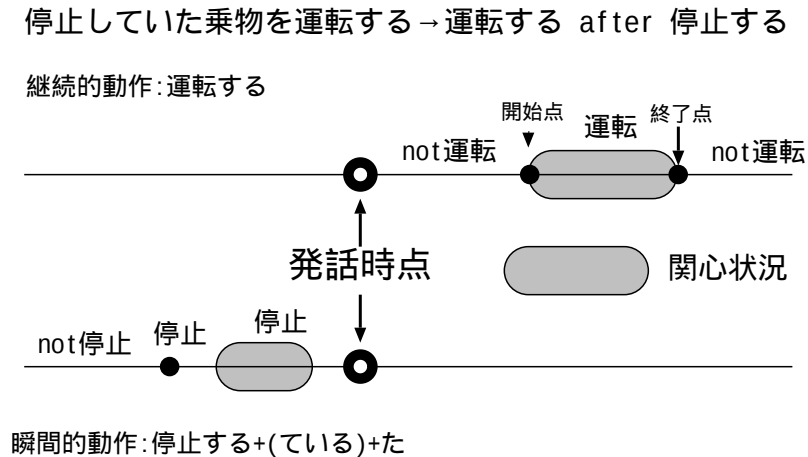


図 2.4: 時制情報を用いた動詞間の時間関係例

図 2.4 場合、「運転する」の関心状況 ($R1$) = 表示状況で、テンスはル形なので $t_0 < R1$ となる。一方、「停止する」の関心状況 ($R2$) = 後状況でテンスはタ系なので $R2 < t_0$ となる。この結果「運転する」と「停止する」の各表示状況の時間関係が図のように決定する。

つまり、東の手法は (2),(3) で文におけるイベントの関心状況を決定し、(1)(4) で、関心状況と発話時点 t_0 との相対的時間関係、関心状況の絶対時間、イベント間の相対的時間関係を決定する。

4.2 節で述べるように、本研究でも、語釈文の解析はこの手法を参考にしている。

2.2 一般知識を用いた時間関係の解析手法

一般知識だけを利用した時間関係の解析を行なっている研究として [Alterman 85] がある。この研究ではあらかじめ動詞間の意味関係を意味ネットワークとして作成しておき、それを利用している。意味ネットワーク中のリンクには分類 (上位下位) 関係、部分的関

係 (共起、順序)、時制関係 (前件、後件) のような関係が付けられている。この研究で使用されている意味ネットワークの一部を図 2.5 に示す。

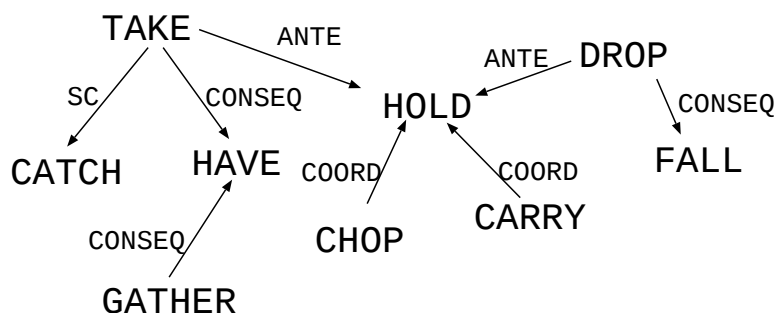


図 2.5: 意味ネットワークの一部

図 2.5 中の sc は上位下位関係、ANTE は前件 (原因)、CONSEQ は後件 (結果)、COORD は共起 (同時) 関係を示す。ここで、'chop' というイベントが生じたとすると、この意味ネットワークから、'hold' というイベントが同時に発生していることが分かる。

一方、前節の [東 96] の手法でも一般知識が使われている。この研究でもあらかじめ作成された知識を用いている。知識ベース中で用いられる動詞間の関係は因果 ($X < Y$)、排除 ($X < Y$ or $X > Y$)、同義 ($X \equiv Y$) の三つである。

本研究の目的は、この手法を用いた時間関係の解析である。しかし、他者の研究で行なわれているように知識を人手で作成することはしないで、機械可読辞書中の情報 (語釈文) を用いて、一般知識に変わりうる知識ベースを作成することが重要な点である。

第 3 章

研究に利用した辞書

3.1 計算機用基本動詞辞書 IPAL

本研究では、計算機用基本動詞辞書 IPAL[情処 87](以下 IPAL) の情報を一般知識を導くための客観的な知識源として利用する。IPAL は日本語の動詞において重要な基本動詞 861 語について、意味および統語的特徴に基づいて下位区分し、それを一つの単位として、意味、形態、統語、文法的カテゴリ、慣用表現などに関わる情報を詳細に記述したものである。

ここでは、この研究で主に用いる辞書中の項目の解説をする。

1. 見出し情報

- (1) エントリ：動詞終止形のひらがなによる表記
- (2) 同音異義番号：同音異義の区別をするための番号
- (3) サブエントリ番号：何番目の下位区分かを示す番号
(意味、統語的特徴に基づいた区分)

IPAL ではサブエントリごとに 1 レコードずつ情報を持っている。つまり語釈文を始めとする辞書中の項目はサブエントリごとに固有である。本研究では「エントリ_同音異義番号_サブエントリ番号」を「見出し語」と呼ぶ。次に例を示す。

(例)

エントリ : なげる
同音異義番号 : 1
サブエントリ番号 : 1
* 見出し語 : なげる_1_1

この場合、見出し語は「なげる_1_1」と表記する。

2. 意味情報

(1) 意味記述：語彙の意味

この情報は人間向けとして、サブエントリの意味を分かりやすく記述したものである（本研究ではこれを語釈文と呼ぶ）。語釈文には単文や複文があるが、単文中の動詞や複文の主節の動詞を定義動詞と呼び、複文の従属節の動詞を従属動詞と呼ぶ。また語釈文には次のよう特徴がある。

- * 見出し語と定義動詞の間に類義関係または上位下位関係がある [富浦 89]
(定義語は見出し語を言い替えたもの)
- * 主節の動詞の時制がすべて現在形
- * 語釈文の定義動詞と従属動詞の間の時間関係の決定しやすい

(2) 意味分類：動詞の意味分類

これは、動詞を意味によって階層的に分類し、サブエントリごとにどのグループに属するかを示したものである。

3. 統語情報

(1) 文型：動詞がとる格形式 (格助詞) のパターン (例：図 3.1 の '格形式')

(2) 述語素：動詞の統語的特徴

動詞と (格要素となる) 名詞句との関係や、名詞句と名詞句の関係を同時に表す情報を示す。

図 3.1 の場合、なげる_1_1 は次のような動詞であることを示す。

A 「なげる」の動作主をガ格に持つ

見出し語	: なげる ₁₁₁
述語素	: A , O 2 a , L S 2 , L G 2
格形式 1	: ガ
意味素性 1	: HUM
格形式 2	: ヲ
意味素性 2	: CON
格形式 3	: カラ
意味素性 3	: LOC
格形式 4	: ニ / ヘ
意味素性 4	: LOC

図 3.1: IPAL における統語情報の例

O 2 a 「なげる」によって影響を受ける対象をヲ格に持つ

L S 2 「なげる」によって影響を受ける対象の空間的起点をカラ格に持つ

L G 2 「なげる」によって影響を受ける対象の空間的着点を二格またはへ格に持つ

(3) 意味素性：名詞句の意味分類

ある動詞が格要素としてどのような分類の名詞をとるかを記述した情報 (詳細は付録 A.1 参照)

4. テンス・アスペクト

(1) テイル形：テイル形の有無とその意味

アスペクト「ている」が接続した時の意味を表す。これはその動詞が継続動詞 (継続的) か結果動詞 (瞬間的) かを区別する重要な情報となる。テイル形には次の4つの分類がある。

単純状態 動きの契機を一切もたず単に状態を表す。

進行 動きが始まって終わるまでの動きの最中の状態に主体 (例えば、「男が走る」の「男」) がある。

結果残存 動きが終わり、その結果生じた状態が主体に存在・継続していることを表す。

経験・完了 動きが終わり、その影響が単に残っているだけで、直接的に現在の主体の状態を規定していない場合を表す。

3.2 分類語彙表 (増補版)

本研究では、名詞句の意味や二つの名詞句の間の類似性を調べるための知識として分類語彙表 (増補版) の電子化データ索引 [佐藤 92] を用いる。電子化データ索引のフォーマットは次の通りである。

表記, 読み, 分類番号, 段落番号, 段落内行番号, 行内語番号

‘分類番号’は、‘表記’の品詞 1 桁 (1=体の類, 2=用の類, 3=相の類) と、意味分類 4 桁で示してある。段落番号以降は、書籍形式の分類語彙表 (増補版) [分類 96] の段落、行、行の何番目かを示す。

本研究では名詞句を表記、読みで索いて、その意味を 5 桁の分類番号から判断する。

(例) 犬, いぬ, 1.5610, 11, 1, 2

‘犬’の分類番号に対する分類項目名を次に示す。

1.5 自然物および自然現象

1.56 動物

1.561 獣

このように分類番号の 1 桁ごとに、階層的に細分類されている。

第4章

語釈文の形式化

語釈文は通常の文で記述されているので、文の内容を人間は理解できても、計算機上ではそのまま利用することは難しい。この章では、IPAL 中の見出し語とその語釈文の情報を形式的に表記する方法を示す。

語釈文の形式化は、語釈文の節単位で表されている内容と見出し語の間を時間関係のリンクで結びつけることにより行なう。見出し語と語釈文中の動詞をリンクで結びつけたものをその見出し語の語義ネットワークと呼ぶ(図 4.1)。

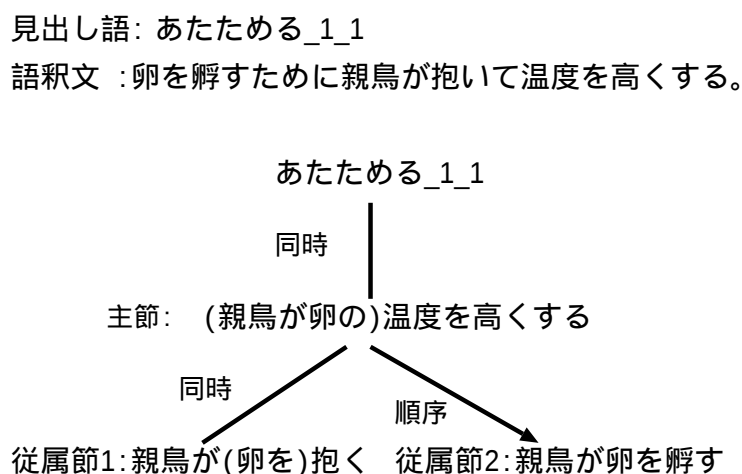


図 4.1: 語義ネットワークの概略

見出し語の語義を表す語義ネットワークを生成するために、まず 4.1 節で単文および複文中の各節) を形式化して表現する方法について説明し、4.2 節で見出し語と定義動詞の

時間関係について定義し、語釈文が複文の場合は節間の時間関係を解析する方法を説明する。

4.1 述語素を用いた単文の形式化

ある見出し語がどのような格を持ち、その格が見出し語に対してどのような関係にあるかを示す情報として「述語素」という項目がIPAL 動詞辞書中にある。本研究では述語素を次の例のような名詞句を表す変数として扱う。

(例)

A : ガ格、かつ動作主である名詞句を表す

O1c : ガ格、かつ変化する対象である名詞句を表す

O2e : ヲ格、かつ出現する対象である名詞句を表す

LG1 : 二格、かつ変化する対象の空間的着点である名詞句を表す
本研究では動詞を

見出し語 (述語素 1, 述語素 2,...)

のように形式化して表す。このようにすれば、括弧内の述語素を見ただけで、その動詞がどのような格構造を持つかが分かる。また、単文は

見出し語 (名詞句 1, 名詞句 2,...)

のように形式化する。このとき名詞句_n($n=0,1,2,\dots$) は、形式化された動詞の述語素_n が表す名詞句でなければならない。次に動詞「あずける」の語義の一つである、見出し語「あずける_{1.3}」の例を挙げる。

- 見出し語の形式化

見出し語 : あずける_{1.3} $\Rightarrow$ あずける_{1.3}(A, LG2, O2)

述語素 : A, LG2, O2

この場合、述語素が A, LG2, O2 であることから、あずける_{1.3}(A, LG2, O2) はガ格の名詞句、カラ格の名詞句、ヲ格の名詞句を持つことが分かる。

- 単文の形式化の例

ϕ が 相手 に 身 を あずける $\Rightarrow$ あずける_{1.3}(A, 相手, 身)

A(省略) LG2 O2 あずける_{1.3}

二格の名詞句「相手」はLG2 が表す名詞句に相当し、ヲ格の名詞句「身」はO2 に相当する。ここで、あずける_{1_3}(A, LG2, O2) のLG2 とO2 を変数として扱い、「相手」と「身」をそれぞれ代入する。形式化された結果が「あずける_{1_3}(A, 相手, 身)」である。

実際の語義ネットワークのデータ構造においてのみ、単文を形式化する時には、名詞句を直接述語素に代入しない。動詞の述語素が表す名詞句が単文中にあった場合、代入する代わりに def(述語素, value, 名詞句) と表す。「相手に身をあずける」という単文は、語義ネットワークのデータ構造上で、次のように表示される。(詳細は 6.1.2 節を参照)

```
あずける1_3(A, LG2, O2)
def(LG2, value, 相手)
def(O2, value, 身)
```

4.2 語義ネットワークにおける動詞間の時間関係の解析

この節では見出し語と、語釈文を構成する節の間の時間関係、つまり見出し語と定義動詞(主節)、定義動詞(主節)と従属動詞(従属節)の時間関係を解析する。

動詞の時間関係を扱うために、まず時間関係の定義をする。次に、動詞の意味に現れる状況を規定することによって動詞を分類する。次に見出し語と定義動詞の時間関係の決定方法を示し、最後にテンスや時制指示語などを利用した複文の節間の時間関係の解析方法を示す。

4.2.1 時間関係の定義

時間幅を持つことを継続的、時間幅を持たないことを瞬間的と呼ぶ。本研究では継続的な二つのイベントの時間関係を [Allen 83] に従い、13 の分類で示す(表 4.1)

表 4.1: 継続的な二つのイベントの時間関係

X before Y	XXX YYY	X after Y	XXX YYY
X meets Y	XXX YYY	X met-by Y	XXX YYY
X overlaps Y	XXXX YYYY	X overlaped-by Y	XXXX YYYY
X starts Y	XXX YYYYYY	X started-by Y	XXXXXX YYY
X finishes Y	XXX YYYYYY	X finished-by Y	XXXXXX YYY
X during Y	XXX YYYYYY	X contains Y	XXXXXX YYY
X equal Y	XXX YYY		

また継続的な動詞 (d) と瞬間的な動詞 (m)(これらの説明は次の小節で行なう) の時間関係は次のように制限される。

{ before,after,equal } 表 4.1 と同様に使用可能。

{ starts,started-by,finishes,finished-by,during,contains } 他方に包含される動詞が必ず m に限定される。(例 : m starts d,d contains m)

{ overlaps,overlaped-by } 両方が継続的な動詞の時間関係を示すため、あり得ない。

さらに瞬間的な動詞 (m) と瞬間的な動詞 (m) の時間関係は次のように制限される。

{ before,after,equal } 表 4.1 と同様に使用可能

{ starts,started-by,finishes,finished-by,during,containsoverlaps,overlaped-by } 片方が両方が継続的な動詞の時間関係を示すため、あり得ない。

4.2.2 語彙アスペクト

動詞は継続的なイベント、瞬間的なイベント、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動詞の意味に現れる状況を規定したものである。これは動詞が表示する状況(表示状況)と、その前提条件を表す(前状況)、結果を表す(後状況)から成り立つ。さらに、動詞と各状況の間には次のような時間関係がある。

- 動詞 met-by 前状況
- 動詞 equal 表示状況
- 動詞 meets 後状況

動詞の表示状況は継続的(継続動詞)なものと瞬間的(結果動詞)なものに分けられることから語彙アスペクトは10種類に分けられる[金子 95, 東ほか 96]。本研究ではこれらのうち9種類の分類を対象とする。そして、IPAL 中の見出し語がどの分類に相当するかを、IPAL の辞書中の情報で決定できるような対応表を表 4.2, 4.3 のように作成した。

表 4.2: 語彙アスペクトの分類 (継続的な動詞)

分類番号	動詞の分類	前状況	表示状況	後状況	意味分類	テイル形
d1	状態動詞	V	V	V	状態	—
d2	継続動詞	not V	V	not V	動作 (not 状態変化)	進行
d3	継続動詞	not V	V	not V(eff)	状態変化、加圧	進行
d4	継続・結果動詞	not V	V	V(res)	動作	進行、結果残存
d5	継続動詞	not V	V	V(eff)	動作	進行、経験・完了

表 4.3: 語彙アスペクトの分類 (瞬間的な動詞)

分類番号	動詞の分類	前状況	表示状況	後状況	意味分類	テイル形
m1	瞬間動詞	not V	V	not V	動作	×
m2	瞬間動詞	not V	V	not V(eff)	動作、状態変化	×
m3	結果動詞	not V	V	V(res)	動作	結果残存
m4	瞬間動詞	not V	V	V(eff)	動作	経験・完了

表中の V(eff) はイベントの結果、対象に何らかの影響が及ぼされた状態を示す。V(res) はイベントの結果として、主体に変化が生じている状態を示す。また、テイル形の「×」はテイル形のない動詞を示す。

この表のうち継続的な動詞 d3 と、瞬間的な動詞 m3 から得られる時間関係の例を図?? に示した。ある見出し語の語彙アスペクトの分類番号を調べたい時は、IPAL で、見出し語の意味分類とテイル形を調べる。例として、意味分類が「動作、状態変化」、テイル形が「進行」である見出し語の分類番号は d3 となる。

4.2.3 述語素の組合せによる状態の抽出

この小節では、動詞の前、表示、後状況における状態を抽出する手法について説明する。状態は、動詞間の時間関係を解析する時に用いられる (5.3 節を参照)。

ここでいう状態とは本研究で形式的に定義したもので、存在状態、所有状態、単純状態のことを示す。それぞれの状態は次のように形式化される。

存在 (人・物, 場所)	= '人・物' が '場所' に存在している状態
所有 (人, 物)	= '人' が '物' を所有している状態 (存在 (物, 人))
状態 (人・物・属性, 状態)	= '人・物 (の属性)' が '状態' である状態

括弧内の第 1 要素には動作主 A が、変化する対象 O が、第 2 要素には動作主や対象の前状況の状態 (=LS1, NS1, LS2, NS2)、後状況の状態 (=LG1, NG1, ...)、表示状況の状態 (=LL1, NL1, ...) があてはまる。

つまり、二つの述語素 (括弧内の第 1 要素と第 2 要素) の組合せにより、前、表示、後状況の状態が抽出される。述語素の組合せによってどのような状態が抽出されるかを表 4.4 に示した。

例として、表 4.4 用いて、はしる₁ (A, LS1, LG1, SP) から状態を抽出する過程を示す。

表 4.4 中の記号 LS1 から時間関係 'met-by' と、状態 '存在 (A|O1c, LS1)' が抽出される。'A|O1c' は述語素 A が O1c のどちらかであることを示す。ここで述語素中には A しかないなので、結果として、'met-by 存在 (A, LS1)' が表から得られる。LG1、SP も同様に、'meets 存在 (A, LG1)', 'during 存在 (A, SP)' が得られる。

表 4.4: 述語素から抽出される状態

記号	訳語	格形式	抽出される状態	動詞との時間関係
A	動作主	[ガ]		
O2a	対象 [受影]	[ガ, ヲ]		
O1c	対象 [変化]	[ガ]		
O2c	"	[ガ, ヲ]		
O1d	対象 [消滅]	[ガ]	存在 (O1d,X)	met-by (前状況)
O2d	"	[ガ, ヲ]	存在 (O2d,X)	met-by (前状況)
O1e	対象 [出現]	[ガ]	存在 (O1e,X)	meets (後状況)
O2e	"	[ガ, ヲ]	存在 (O2e,X)	meets (後状況)
O1	対象 [その他]	[ガ]		
O2	"	[ガ, ヲ]		
LS1	空間的起点	[ガ, カラ]	存在 (A—O1c,LS1)	met-by (前状況)
LS2	"	[ヲ, カラ]	存在 (O2c,LS2)	met-by (前状況)
LG1	空間的着点	[ガ, ニ/ヘ]	存在 (A—O1c,LG1)	meets (後状況)
LG2	"	[ヲ, ニ/ヘ]	存在 (O2c,LG2)	meets (後状況)
LL1	空間的场所	[ガ, ニ/デ]	存在 (A—O1c,LL1)	during (表示状況)
LL2	"	[ヲ, ニ/デ]	存在 (O2c,LL2)	during (表示状況)
NS1	非空間的起点	[ガ, カラ]	状態 (A—O1c,NS1)	met-by (前状況)
NS2	"	[ヲ, カラ]	状態 (O2c,NS2)	met-by (前状況)
NG1	非空間的着点	[ガ, ニ/ヘ]	状態 (A—O1c,NG1)	meets (後状況)
NG2	"	[ヲ, ニ/ヘ]	状態 (O2c,NG2)	meets (後状況)
NL1	非空間的场所	[ガ, ニ]	存在 (A—O1c,NL1)	during (表示状況)
NL2	"	[ヲ, ニ]	存在 (O2c,NL2)	during (表示状況)
SP	空間	[ガ, ヲ]	存在 (A—O1c,SP)	during (表示状況)
IN	手段	[ガ, デ]	使用*(A,IN)	equal (表示状況)

‘使う’から得られるイベント

述語素‘IN’はデ格で、手段を表す名詞句を表す。このとき、表 4.4 から、使用*(A,IN) という状態が抽出される。これは動作主 A が IN を使用していること示す。しかし、‘使う’というイベントは対象となる名詞句によって異なる使い方を表す。そこで、使い方を表すイベントをデ格の名詞句(‘使う’の対象)ごとに分けて図 4.5 のように定義する。

表 4.5: ‘使う’から得られるイベント

ガ格	デ格	発生するイベント	動詞との時間関係
人	乗り物	(人が乗り物に) 乗る	started-by
人	道具	(人が道具を) 持つ	equal
人	行為	(人が行為を) する	equal

‘使う’がどの意味になるかはデ格の名詞句の意味を調べねばならない。そこで、名詞句の分類番号を分類語彙表で調べ、表 4.6 のように決定する。

表 4.6: 名詞句と分類番号の対応表

名詞句	分類番号
乗り物	1.465x
	1.466x
	1.467x
道具	1.45xx
行為	1.3xxx

x は任意の数字を示す。例えば、「自動車」は分類番号 14650 なので、「乗り物」であると決定する。また、「ペン」は分類番号 14530 なので「道具」であると決定する。

4.2.4 見出し語と定義動詞の時間関係

定義動詞は見出し語を言い替えた語であると考えれば、通常は時間関係は一致すると考えられる。しかし、見出し語が継続的な動詞 (m) で定義動詞が瞬間的な動詞である場合

は同時とはいえない。その逆も同様である。この節では、見出し語と定義動詞の語彙アスペクトが異なる時の時間関係について定義する。

見出し語の表示状況を (E1) 後状況を (S1)、定義動詞の表示状況を (E2) 後状態を (S2) とする。以降、「後状況が存在する」とは、後状況が $V(\text{res})$ または $V(\text{eff})$ (not $V(\text{eff})$ を含む) であることを示す。

優先度 1 E1、E2 が共に継続的である時 E1 と E2 は同時とする。(→ 見出し語 equal 定義動詞)

優先度 2 S1、S2 が共に存在する場合 S1 と S2 は同時とする。(→ 見出し語 { finishes | finished by } 定義動詞)

優先度 3 E1 が継続的で、S2 が存在する場合、E1 と S2 を同時とする。またはその逆 (→ 見出し語 { starts | started-by } 定義動詞)

この優先度を用いた、見出し語を定義動詞の時間関係の例を次に示す。

- 見出し語 (m3) 定義動詞 (d3) の例 (結果として、主体が対象の状態が変わる動詞)

見出し語：うつる_{2.4}(m3)

定義動詞：すすむ_{1.1}(d3)

一致する状況：「うつる」の後状況 (S1) = 「すすむ」の後状況 (S2)

時間関係：うつる_{2.4} finished-by すすむ_{1.1}

「うつる」の表示状況 (E1) が瞬間的であるから優先度 1 の条件には合わない。しかし S1 と S2 が存在することから、優先度 2 の時間関係が選択される。この場合、見出し語が瞬間的であるため、時間関係は finished-by となる。(定義動詞が瞬間的な場合は finishes)

- 見出し語 (d2) 定義動詞 (m3) の例

見出し語：あいする_{1.1}(d2)

定義動詞：(好きだという気持ちをもつ)_{1.5}(m3)

一致する状況：「あいする」の表示状況 (E1) = 「もつ」の後状況 (S2)

時間関係：あいする_{1.1} started-by もつ_{1.5}

「もつ」の表示状況 (E2) が瞬間的であるから優先度 1 の条件には合わない。さらに、「あいする」の後状況 (S1) が存在しないため、優先度 2 の条件にも合わない。しかし、E1 が継続的で S2 が存在することから、優先度 3 の時間関係が選択される。

4.2.5 語釈文における複文の節間の時間関係の解析

この節の目的は本研究のテーマである動詞間の時間関係の解析と類似する。しかし、本研究の目的が手法 2(動詞間の依存関係) による時間解析であるのに対して、ここでは語釈文における複文を手法 1(時制情報) と IPAL 中の動詞の意味情報を利用して解析を行なう。つまり、次のような利点を生かすことができる。

- 節間の接続助詞を始めとする時制情報を利用できる。
- 時制情報を用いた動詞の時間関係を解析しやすい。
例えば、従属節が過去形の連体節で、主節が過去形か、進行形のときは曖昧が生じる。しかし、語釈文の性質上、主節動詞が基本形である (時制が現在形で、アスペクトが接続していない)。

前者は複文の特徴で、後者は語釈文の特徴である。

時間関係の解析は主節と従属節の間の接続関係で大まかに時間関係を限定し、従属節のテンス、アスペクト、各節の動詞の分類 (語彙アスペクト) から時間関係をさらに限定する (表 4.7)。

表 4.7: 接続形式による節間の時間関係

接続形式	テンス・アスペクト	語彙アスペクト	主節との関係	時間関係
連体	タ形・なし	d2 ~ d5,m	過去の動作	after
	タ形・なし	d1	過去 (から) の状態	after*2
	タ形・テイル	d2 ~ d5	過去 (から) の動作	after*2
	タ形・テイル	m1 ~ m4	過去の動作	after
	ル形・なし	d1	付帯状況	during*1
	ル形・なし	d2 ~ d5	未来の動作	before*2
	ル形・なし	m1 ~ m4	未来の動作	before
	ル形・テイル	d2 ~ d5	付帯 (動作) 状況	during*1
	ル形・テイル	m3,m4	付帯 (結果) 状況	before
テ形	ル形・なし	d1	付帯状況	during*1
	ル形・なし	d2	手段	equal
	ル形・なし	d3 ~ d5	原因・手段	after
	ル形・なし	m1 ~ m4	原因・手段	after
ために	ル形・なし	m1 ~ m4	目的	before*1
	のに	ル形・なし	目的	during*2
ために	タ形・なし	d,m	原因	after*1
ながら	ル形・なし	d2 ~ d5	付帯 (進行) 状況	during*1
まま	タ形・なし	d4,d5,m3,m4	付帯 (結果) 状況	after
ように	タ形・—	d2	方法を表す様帯	after*1
	ル形・なし	d2	方法を表す様帯	after*1
	ル形・なし	d3 ~ d5,m	目的	before*1

(以下、主節動詞の動作の開始点を s_1 終了点を e_1 、従属節の動作の開始点を s_2 終了点を e_2 とする。)

$\text{before}^*1 = \{ \text{before} \mid \text{meets} \} (s_1 < s_2)$

$\text{before}^*2 = \{ \text{before} \mid \text{meets} \mid \text{overlaps} \mid \text{finished-by} \mid \text{contains} \} (e_1 \quad s_2)$

$\text{after}^*1 = \{ \text{after} \mid \text{met-by} \} (e_2 \quad s_1)$

$\text{after}^*2 = \{ \text{after} \mid \text{met-by} \mid \text{overlapped-by} \mid \text{started-by} \mid \text{contains} \} (s_2 < s_1)$

$\text{during}^*1 = \{ \text{during} \mid \text{started-by} \mid \text{finished-by} \mid \text{equal} \} (s_2 \quad s_1 < e_1 \quad e_2)$

$\text{during}^*2 = \{ \text{during}^*1 \mid \text{before}^*1 \}$

表 4.7 の解説をするために、次の例文をこの表を用いて、節間の時間関係を解析する仮定を示す。

[例文] 止まっていた乗物を動かす (類似例は 2.4 を参照)

とまる 1.1 の語彙アスペクトは m3、テンスはタ形、そして、アスペクト「テイル」が接続している。表 4.7 を参照して、この情報から時間関係を調べると、after であることが分かる。

4.3 動詞間の格要素の一致

動詞が受けえる名詞句を格要素と呼ぶ。この節では、見出し語の格要素が、語釈文中の定義動詞のどの格要素と一致するかを、決定する手法について解説する。

IPAL 中には格要素の意味素性が記述されている。動詞間の格要素の一致を決める手法の一つとして、意味素性が一致する格要素を選択する方法が考えられる。しかし、格要素に複数の人を受ける動詞だった場合は曖昧さが生じる。

もう1つの手法として、動詞間で一致する格の格要素を選択するという方法がある。しかし、これは二つの動詞が、類似した意味と類似した格構造を持っていないなければならない。

見出し語と定義動詞には上位下位関係があるという研究が[富浦 89]で行なわれている。実際に IPAL を調べると、ほとんどの場合、定義動詞は見出し語のレコード中の‘上位語’が‘類義語’に登録されている。

以上より、本研究では見出し語と定義動詞は類似しているとみなし、前述の2つの手法を併用して、動詞間の格要素の一致を決定する。

どのような条件で、格要素が一致するかは、次のような優先度の上から順に従う。

優先度 1 格が一致し、格要素の意味素性も一致する

優先度 2 格は一致しないが、格要素の意味素性が一致する

優先度 3 格が一致するもの (状態変化に影響のない格のみ)

(例)

見出し語 [あいする-1-1]:HUM ガ HUM/ANI/ORG ヲ

語釈文: 誰かや何かに対して、とても好きだという気持ちを持つ。

定義動詞 [もつ-1-5]:HUM ガ HUM/ANI/ORG(=誰か、何か) ニ MEN(=気持ち) ヲ

この場合見出し語の格の一致は次のようになる。

あいする _{1.1}	→	もつ _{1.5}
ガ格	→	ガ格
ヲ格	→	二格
— — —	→	ヲ格 (= 気持ち)

この場合、まず、ガ格の格要素が優先度 1 によって一致すると決定される。他は優先度 1 の条件に合わない。次に、ヲ格の格要素と二格の格要素が優先度 2 によって対応すると決定される。「もつ」のヲ格の格要素はこの時点でどの格要素にも一致しない。

4.4 使役形、受身形の動詞に対する述語素の拡張

使役、受身形である動詞の格構造は使役、受身形でない動詞の格構造とは異なる。述語素は動詞の格構造も表しているが、使役、受身形の格構造は IPAL 中に登録されていない。そこで、使役、受身形でない動詞の述語素から使役、受身形である動詞の述語素が得られるような変換規則を定義する。

• 使役文への変換

- (1) 述語素に動作主 A がある場合 $A \rightarrow PT_n$
- (2) 主格となる対象 $O1_x$ がある場合 $O1_x \rightarrow O2_x$
- (3) 主格の状況を表す述語素がある場合 $Lx1 \rightarrow Lx2, Nx1 \rightarrow Nx2$
- (4) 新たに動作主 A を追加

この変換規則を利用して、使役形の述語素に変換した例を次に示す。

(例)

うまれる_{1.1}($O1_e, LS1, IN$) (ひなが卵から生まれる)

↓

うまれる_{1.1}[使役]($A, O2_e, LS2, IN$) (親鳥がひなを卵から生まれさせる)

のむ_{1.1}(A,O2d) (赤ん坊がミルクを飲む)

↓

のむ_{1.1}[使役](A,PTn,O2d) (母親が赤ん坊にミルクを飲ませる)

前者の例は変換規則 (2),(3),(4) を利用し、後者の例は変換規則 (1),(4) を順に用いて変換した。

- 受身文への変換 (行為を受ける対象が主格となる場合)

(1) 述語素に動作主 A がある場合 $A \rightarrow PTn$

(2) 行為を受ける対象 O2x がある場合 $O2x \rightarrow O1x$

(3) 主格の状況を表す述語素がある場合 $Lx2 \rightarrow Lx1, Nx2 \rightarrow Nx1$

この変換規則を利用して、受身形の述語素に変換した例を次に示す。

(例)

たべる_{1.1}(A,O2d) (ライオンがシマウマを食べる)

↓

たべる_{1.1}[使役](O1d,PTn) (シマウマがライオンに食べられる)

この例は変換規則 (1),(2) を用いて変換した。

4.5 語義ネットワーク

以上の処理で図 4.2 のような語義ネットワークをすべての見出し語に対して作成する。

X after Y のような時間関係は図 4.2 中では

$$X \xrightarrow{\text{after}} Y$$

のように表される。これは X から Y に対する時間関係を表している。矢印の向き逆になる時は時間関係も変わる。表 4.1 中の左右対になっている時間関係に変わる (例: after の逆は before)。equal の場合のみ矢印の方向に関係なく、常に equal である。

見出し語: あたためる_1_1

語釈文: 卵を孵すために親鳥が抱いて温度を高くする。

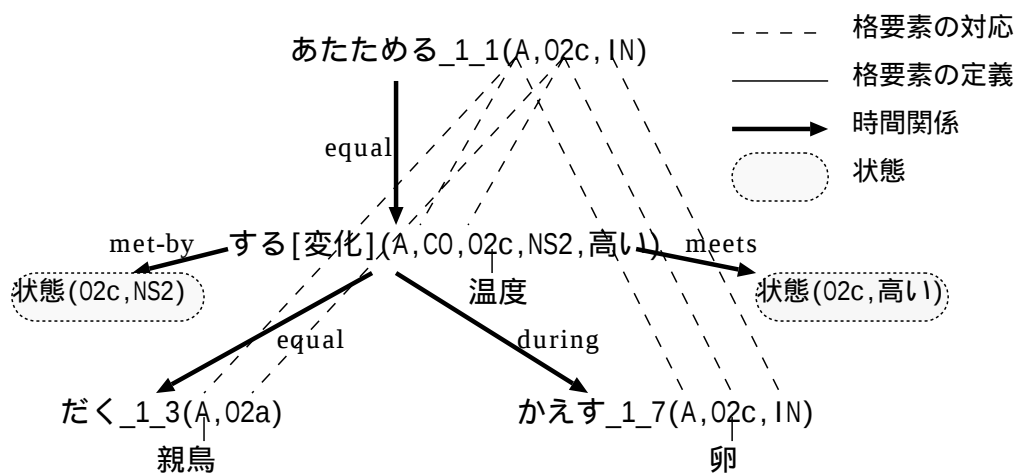


図 4.2: あたためる_1_1 の語義ネットワーク

第 5 章

時間関係の解析方法

ここでは形式化された 2 つの単文を入力し、それらの時間関係を解析する方法を示す。時間関係解析部分のおおまかな流れは次の通りである。

1. それぞれの見出し語で、語義ネットワークを索く
2. 格要素を語義ネットワーク中の述語素に代入する
3. 述語素の組合せから前・表示・後状況の状態を抽出する
4. 語義ネットワーク間のイベントの類似性を調べる (一致するイベントを equal とみなす)
5. 語義ネットワーク間の状態の類似性を調べる (一致する状態を equal とみなす)
6. 入力された見出し語の間の時間関係のつながりから時間関係を計算する
7. 複数のイベントや状態が一致し、矛盾が生じる場合は複数解を出力する

時間関係解析システムの全体的な流れを図 5.1 に示した。図中の前処理で行なう処理に関しては 6.1 節、図 6.1 で説明する。前述の流れは図中の「語義ネットワークの抽出」と「時間関係解析部」を示す。「時間関係解析部」の処理は図中に示した節で説明する。

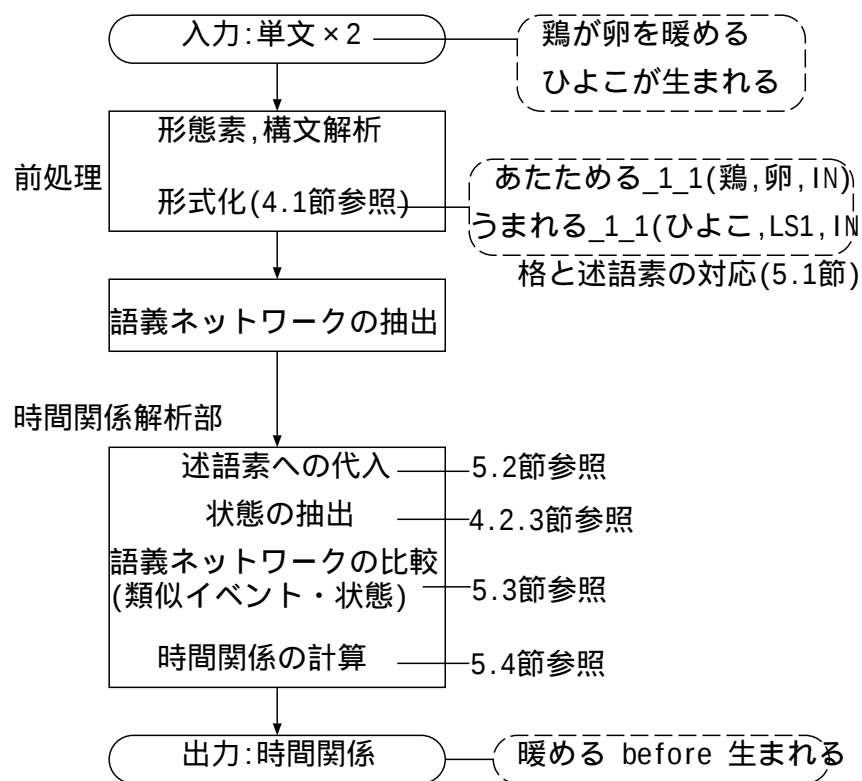


図 5.1: 時間関係解析システム

5.1 述語素が表す名詞句の意味素性

述語素はどのような格を持つ名詞であるかを示す。デ格を持つ名詞句を表す述語素や、二格を持つ名詞句を表す述語素は複数ある。従って、入力文中にデ格を持つ名詞句があった場合、どの述語素になるかは名詞句の意味から判断するしかない。そこでこのシステムでは、入力文中の名詞句の意味素性から、その名詞句がどの述語素に当てはまるかを決定する。述語素がどのような名詞の意味素性を表すかを次に示す。

- 手段を表す述語素 IN が受け付ける名詞句の意味素性 (ACT, PRO, PAR)
- 場所を表す述語素 Lx,SP が受け付ける名詞句の意味素性 (LOC, CONの下位素性)
- 非空間的場所を表す述語素 Nx が受け付ける名詞句の意味素性 (TIM以外)
- 起因を表す述語素 AS が受け付ける名詞句の意味素性 (ACT, MEN, PHE)

入力文中の名詞句が、うまれる_{1.1}(O1e,LS1,IN) のどの述語素に相当するかを決定する例を次に示す。

- 格要素が述語素にあてはまる例

入力： 動詞 (うまれる_{1.1}) 格構造 (太郎 が おなか から 帝王切開 で)

↓

出力： うまれる_{1.1}(太郎, おなか, 帝王切開)

「おなか」はカラ格、かつ意味素性 PAR で述語素 LS1 に、「帝王切開」はデ格、かつ ACT で述語素 IN に当てはまる。

- 格要素が述語素にあてはまらない例

入力： 動詞 (うまれる_{1.1}) 格構造 (太郎 が 朝 から 病院 で)

↓

出力： うまれる_{1.1}(太郎,LS1,IN)

「朝」はカラ格、意味素性「TIM」なので LS1 ではない。「病院」はデ格、かつ意味素性「ORG、LOC」なので IN ではない。

5.2 意味素性への代入と例外への対策

入力文から得られる格要素は語義ネットワーク中動詞の述語素に代入される。さらに動詞間の格要素の一致によって、ネットワーク中の関連イベントのスロットにも代入される。

しかし、動詞間の格要素の一致をたどって関連イベントの述語素を埋める場合、入力された名詞句が関連イベントの述語素に当てはまらない場合が生じる。これは語尺文を解析した時に、すでにある語釈文中の名詞句が代入されている場合に起こりうる。図 4.2 のような語義ネットワークに対して、次のような入力があったとする。

入力：あたためる_{1_1}(飼育係, 卵, IN)

このとき図中の格要素の一致より、名詞句「飼育係」は「だく_{1_3}」の述語素 A に代入される。同様にして「卵」は O_{2a} に代入される。すると、動詞「だく」は次のようになる。

出力：だく_{1_3}(飼育係, 卵) ???

この表現は明らかにおかしい。図中格要素の定義を見ると「だく」の述語素 A に「親鳥」という名詞句が定義されている。これは語釈文中の名詞句を表す。「飼育係」を A に代入するかどうかは、この名詞句と類似しているかどうかで判断する(名詞句の類似性の判断は次の節で行なう)。この場合、「飼育係」と「親鳥」は類似していないので代入されず、次のように出力される。

出力：だく_{1_3}(A, 卵)

5.3 イベント、状態の類似性の判定

2つの語義ネットワーク間の類似状況を調べるために、すべてのイベント間と状態間の類似性を調べ、一致または類似するイベントや状態を同時(equal)とみなす。手順としては、まずイベント間の類似を調べ、失敗した場合にのみ状態間の類似を調べる。

イベント間、状態間の類似性を調べる手法を次に示す。

イベントの一致：イベントが一致しているとみなす条件は、まず先頭の見出し語が一致していること、次に動作主(A)、対象となる名詞句(O)が一致、または類似していることである。残りの格要素となる名詞句は片方か、両方が変数であってもよい(条

件つき一致)。この時二つのイベントは同時に起こるとみなす。

(例)

うまれる_1_1(ひな, 卵, IN)
 ↓一致 ↓類似 ↓条件付き一致 ↓一致
 うまれる_1_1(ひよこ, LS1, IN)

状態の一致：状態の一致する条件は、先頭の状態(存在, 所有, 状態)が一致していることと対象を表す名詞句と、その場所や所有物、状態を表す名詞句が一致または類似していることである。このとき二つの状態は同時に発生するとみなす。

(例)

存在(太郎, 学校)
 ↓一致 ↓一致 ↓一致
 存在(太郎, 学校)

状態の矛盾：対象を表す名詞句が一致または類似しているが、その場所、所有物、状態が異なるとき二つの状態は矛盾しているとみなされ、同時には起こらない。状態は一致するが片方の先頭に「¬」がついている場合も矛盾するとみなす。

(例)

存在(太郎, 学校)
 ↓一致 ↓一致 ↓不一致
 存在(太郎, 家)

名詞と名詞が類似しているかを調べる方法は、分類語彙表(3.2節参照)を用いる。分類語彙表中の名詞(分類番号が1xxxxx)をそれぞれ参照し、5桁の分類番号が一致すれば二つの名詞句は類似しているとみなす。

(例)「ひな」と「ひよこ」の類似性

表記	読み	分類番号	段落番号	段落内行番号	行内語番号
ひな	ひな	14570	3	2	4
		<u>15620</u>	6	2	1
		12540	3	2	2
ひよこ	ひよこ	<u>15620</u>	6	2	2

この場合、「ひな」に関して、複数の分類番号が出力されるが、分類番号の最初の5桁が一致する組合せが一つでもあれば類似するとみなす。

5.4 時間関係の計算

見出し語 (X) ともう一つの見出し語 (Y) の時間関係がほかの複数の動詞 (A,B,C) との時間関係を介して

X after A, A starts B, B equal C, C before Y

のように決定されているとする。この時、X と Y の時間関係を自動的に決定するためには、あらかじめ、2 つの隣接する時間関係に対して、時間関係を決定しておく必要がある。例えば、'X after A, A starts B' の時の X と B の時間関係。このように全ての組合せに対して時間関係を決定しておけば、2 つの隣接する時間関係を一つにすることができる。この手法は [Allen 96] でも用いられている。

[共通](x:任意の時間関係)

A equal B, B x C = A x C

A x B, B equal C = A x C

[規則の一部]

A finishes B ,	B before C	= A before C
	B after C	= A after C
	B meets C	= A meets C
	B met-by C	= A after C
	B starts C	= A during C
	B started-by C	= A { after,met-by,orverlaped-by } B
	B finishes C	= A finishes C
	B finished-by C	= A { finished-by,eqaul,finishes } B

この表は A の語彙アスペクトが d か m によって、かなり制限される (曖昧さが解消される)。

片方の動詞が m: overlap はすべて無効、equal,after,before 以外条件付きで無効。

両方 m: equal,after,before 以外無効

[例] A finishes B, B started-by C, C finishes D のとき A と D の時間関係
(A ~ D の動詞はすべて継続的な動詞とする)

A finishes B , B started-by C = A { after, met-by, overlaped-by } C

A after C , C finishes D = A { after } D

A met-by C , C finishes D = A { met-by } D

A overlaped-by C , C finishes D = A { overlaped-by } D

A. A { after, met-by, overlaped-by } D

C が瞬間的な動詞の時は A finishes B , B started-by C = A { after } C
A after C , C finishes D = A { after } D
より A after D

5.5 複数のイベント・状態が一致した時の処理

複数のイベント・状態が一致した場合、二つの見出し語間は複数の経路でつながると考えられる。このときは、図 5.2 のように各経路ごとに時間関係を計算し複数の時間関係を出力する。

図 5.2 の場合、「おきる」と「ねる」の語義ネットワーク間のイベント「たつ」と「たおれる」が一致している。それぞれを時間関係 equal のリンク (点線) で結ぶと、見出し語「おきる」と「ねる」を結ぶ経路は

- 「おきる」 $\xrightarrow{\text{after}(1)}$ $\bullet$ $\xrightarrow{\text{equal}(5)}$ $\bullet$ $\xleftarrow{\text{starts}(4)}$ 「ねる」
- 「おきる」 $\xrightarrow{\text{starts}(2)}$ $\bullet$ $\xrightarrow{\text{equal}(6)}$ $\bullet$ $\xleftarrow{\text{after}(3)}$ 「ねる」

の 2 通りである。「おきる」と「ねる」の関係を求めるには「おきる」→「ねる」の形にしなければならない。4.5 節で述べたように、矢印の方向を変える場合は時間関係が次のように変わる。

- 「おきる」 $\xrightarrow{\text{after}(1)}$ $\bullet$ $\xrightarrow{\text{equal}(5)}$ $\bullet$ $\xrightarrow{\text{started-by}}$ 「ねる」

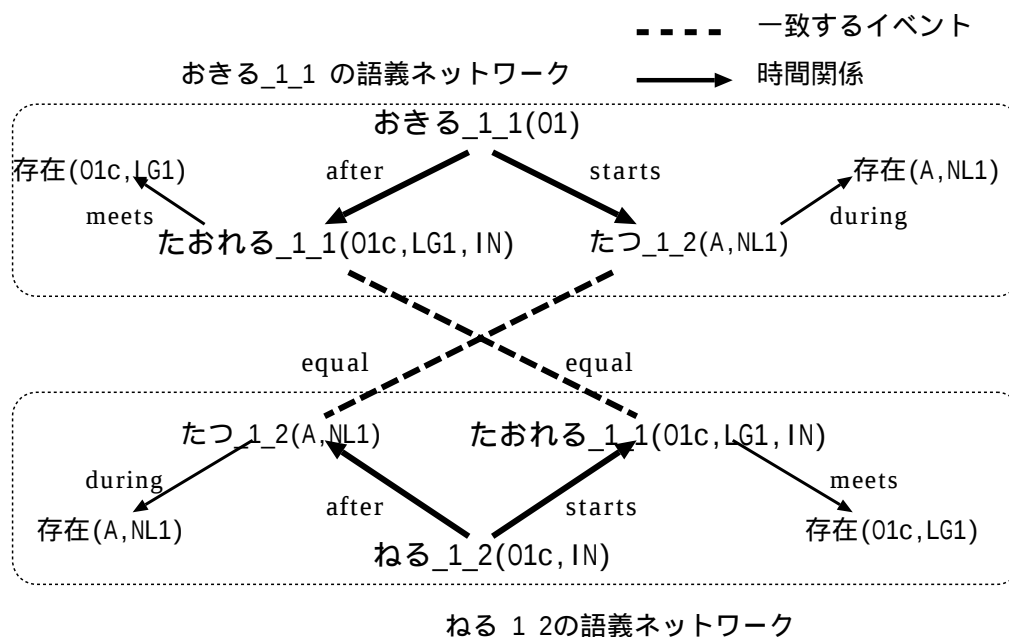


図 5.2: 複数のイベントが一致した時の時間関係の解析

- 「おきる」 $\xrightarrow{starts(2)}$ $\bullet$ $\xrightarrow{equal(6)}$ $\bullet$ $\xrightarrow{before}$ 「ねる」

後は前節の定義に従って時間関係を計算する。

$A \text{ after } B \text{ started-by } C = A \text{ after } C$

$A \text{ starts } B \text{ before } C = A \text{ before } C$

以上より「おきる」 $\{ \text{after} \mid \text{before} \}$ 「ねる」という時間関係が決定する。

5.6 語義ネットワークの再帰的参照

見出し語の語義ネットワークだけでは情報が足りない場合がある。この場合は語義ネットワーク中の動詞を見出し語とした語義ネットを参照することで、より多くの情報を利用することができる (図 5.3)。本研究ではこの再帰的参照を 1 回行なった場合の実験も行う。

見出し語: あたためる_1_1

語釈文: 卵を孵すために親鳥が抱いて温度を高くする。

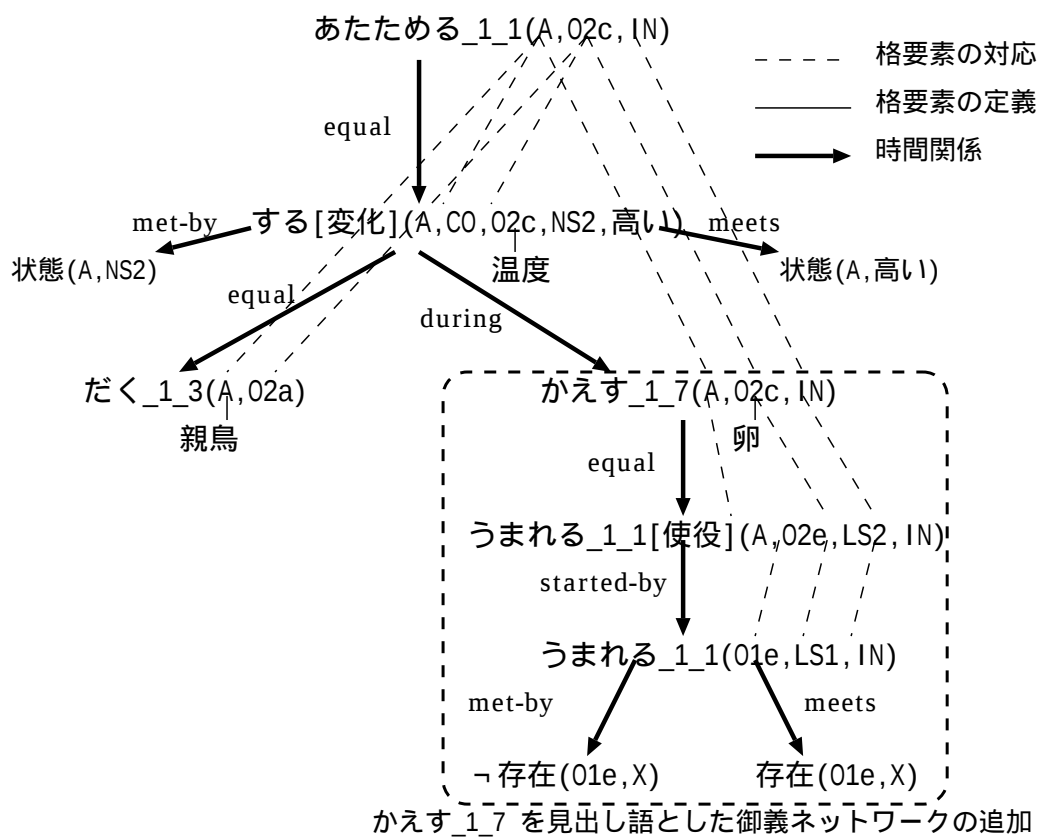


図 5.3: あたためる-1-1 の語義ネットワークの再帰的参照

第 6 章

システムの実装と評価

6.1 語釈文の形式化システムの実装

この節では語釈文から語義ネットワークを作成する過程に関する解説を行なう。日本語基本動詞辞書 IPAL の見出し語に対する語義ネットワークを語釈文から自動的に作成するためのシステムを作成した。そのシステムの流れを表 6.1 に示す。

語義ネットワークを作成するには前処理として、語釈文を形態素解析、構文解析して動詞の語義をする必要がある。形態素解析には日本語形態素解析システム JUMAN version3.0beta [松本他 94] を用いた。構文解析と動詞の語義 (IPAL 動詞辞書中の語義) の決定には KN parser [Kurohashi 92] を用いた。前処理の終わりに、KNP の結果から、語義ネットワークを作成するのに必要な情報を抜き出す作業を行なう (表 6.1 中の 'KNP 出力の加工' で行なう)。

前処理の後は、4 章で解説した手法を用いて、「単文の形式化」、「状態の抽出」、「節間の時間関係の解析」を順に行なう。

以降の小節ではまず、入力した語釈文に関すると、出力する語義ネットワークのデータ構造について解説する。次に KNP でどの程度、語義が決定できたかについて解説する。

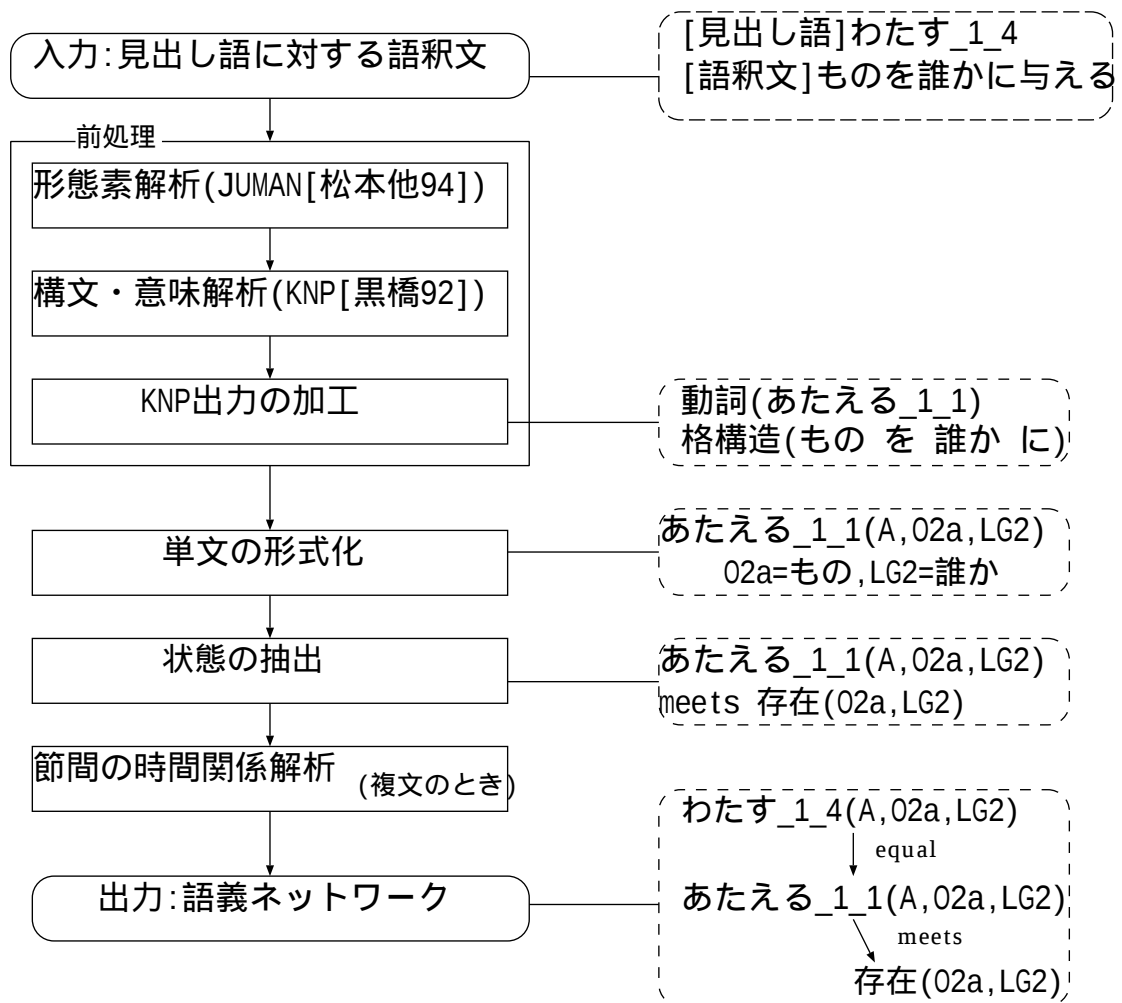


図 6.1: 語義ネットワーク作成システムの流れ

6.1.1 入力

入力とする語釈文は日本語基本動詞辞書 IPAL の 3379 レコード中の意味記述である。意味記述の中で文になっていないものは対象としないか、部分的に削除する。

対象としない記述 名詞で終わっている記述

例 1. 「やる」「与える」丁寧な表現。

例 2. 「聞く」の謙譲語。

削除する記述 括弧で囲まれた記述のうち、脚注を表すもの

例 1. 足を交互に動かして移動する(普通、一方の足は地に付いている)

例 2. 自分の所有物を(一時的に)使わせる(具体物は返却を前提とする)

括弧を外した時、文として成り立つ記述はすべて利用する。

6.1.2 出力：語義ネットワークのデータ構造

語義ネットワークは、見出し語と語釈文中の定義動詞や従属動詞の時間関係のリンクで構成される。構造規則を図 6.2 に示す。

データ構造の構成要素は、見出し語を始めとする動詞(= イベント、状態)、動詞間の時間関係、動詞間の格の一致 (rel)、語釈文中の格要素 (def) から構成される。

イベント イベント (述語素 1, 述語素 2,...)

(例)

はしる_1_1(A,LS1,LG1,SP)

{ A=動作主 (ガ格),LS1=空間的起点 (カラ格),LG1=空間的着点 (ニ/ヘ格),SP=空間 (ヲ格) }

状態 存在 (述語素 1, 述語素 2) または 状態 (CO, 述語素 1, 述語素 2)

(例)

存在 (O1c,LS1) { O1c=対象 (変化)(ガ格) }

状態 (CO,O1c,NG1) { CO=対象の所属 (ノ格),NG1=非空間的着点 (二格) }

```

- ー ー 見出し語
  |   語彙アスペクト分類
  L- ー ー [(見出し語との) 時間関係] 定義動詞 (語釈文の主動詞)
    |   語彙アスペクト分類
    |   rel 見出し語との述語素 (格) の一致
    |   def 述語素の値 (語釈文中の格要素)
    L- ー ー [(定義動詞との) 時間関係] 従属節 1 の動詞
      |   語彙アスペクト分類
      |   rel 見出し語との述語素 (格) の一致
      |   def 述語素の値 (語釈文中の格要素)
      トー [(定義動詞との) 時間関係] 従属節 2 の動詞
      ...

```

図 6.2: 語義ネットワークのデータ構造

rel 見出し語の述語素 (格) との対応

ある動詞の述語素 (格) が見出し語のどの述語素と一致するかを示す。これは、4.3 節で述べたように、名詞の意味素性と格の一致により決定している。

(例)

‘あたためる_1_1(A,O2c,IN)’ と ‘だく_1_3(A,O2a)’ の間には次のような格の一致がある。

rel(A,value,A) #動詞「だく」の述語素‘A’は「あたためる」の‘A’と一致する

rel(O2a,value,O2c) #動詞「だく」の述語素‘O2a’は「あたためる」の‘O2c’に一致する

def 述語素の値

述語素が表す名詞句を示す。名詞句は語釈文中の格要素を表す。

(例)

あたためる_1_1 の語釈文「卵を孵すために親鳥が抱いて温度を高くする」である場合、述語素の値は次のように表される。

だく_1_3(A,O2a)

def(A,value, 親鳥)

(def(O2c,value, 卵) は「かえす」に係る)

実際に出力される実際のデータ形式は図 B.1 のようになる

6.1.3 KNP の語義決定結果

6.1 節で解説したように、語釈文中の動詞の語義は KNP で決定される (動詞の語義が決定するとは、動詞が IPAL 中のどの見出し語を示すかを決定すること)。この語義が決定できるか、また決定できた語義があっているかどうかは、語義ネットワークの精度 (ネットワーク中の語義がどの程度正解なのか) に大きく影響する。従って、この節では語釈文中の動詞の中で、KNP で語義が決定できる動詞はどれくらいあるかを調べる。そして、決定できた語義の正解率を調べる。

全語釈文中の述語の解析結果を表 6.1 に示す。

表 6.1: 語釈文中の述語の分類

品詞	数	備考
普通動詞	4341	動詞辞書の見出し語に対応する語義を決める
サ変名詞 + する	951	サ変動詞辞書の見出し語に対応する語義を決める
サ変名詞 + できる	19	サ変動詞辞書の見出し語に対応する語義を決める
形容詞 + する, なる	210	すべて状態変化を表す動詞として扱う
形容詞	565	状態として扱う
名詞 + だ	32	状態として扱う
合計 (全述語数)	6118	

KNP は表 6.1 の「普通動詞」と「サ変名詞 + する」の語義を決定することができる。それ以外の品詞は、表中の「備考」欄のとおり扱う。

普通動詞の語義決定率と決定できなかった動詞の分類を表 6.2 に示す。

表 6.2: 普通動詞の語義決定結果

普通動詞 4341				
語義決定可能 3095	語義決定不可能 1246			
	(A) エントリ 504	(B) 類義語 336	(C) 未定義語 391	(D) エントリ + 助動詞 35

表 6.2 中の (A) ~ (D) は次のような動詞を表す。

(A) 基本動詞辞書の項目‘ エントリ’ に登録されている動詞

語釈文の構造が複雑で、文の係り受けを導くことができなかつたり、格要素がないため語義が決定できなかった動詞。

(B) 見出し語には登録されていないが、別の見出し語の項目‘ 類義語’ に登録されている動詞

(C) 基本動詞辞書の見出し語にも、別の見出し語の‘ 類義語’ にも登録されていない動詞

(D) 動詞 + 助動詞 (使役・受身) が 1 つの複合動詞として解析された動詞のうち、動詞が基本動詞辞書の見出し語に登録されている動詞

語義を決定できたのは、普通動詞 4341 語中、3095 語 (71.3 %) だった。解析できなかった (A) ~ (D) の動詞のうち、(A)(B)(D) は前処理を行なってから、再度解析を行なう (6.1.4 節参照)。(C) は語義が決定できなかった動詞とみなす。

次にサ変動詞の語義決定結果を表 6.3 に示す。

表 6.3: サ変名詞 + 「する」の語義決定結果

サ変名詞 + 「する」 951			
名詞辞書の語義に決定可能 142		名詞辞書の語義に決定不可能 809	
(E) サ変動詞辞書に登録 37	(F) サ変動詞辞書に未登録 105	(G) 動詞辞書の類義語に登録 617	(H) 動詞辞書の類義語に未登録 192

表 6.3 中 (E) ~ (H) の動詞について解説する。

(E) サ変名詞の語義がサ変動詞辞書の語義に変換できる

サ変動詞として語義が決定できたもの。

(F) サ変名詞の語義がサ変動詞辞書の語義に変換できない

サ変名詞としては語義が決定できたが、サ変動詞としては語義が決定できなかったもの。

(G) 見出し語には登録されていないが、別の見出し語の項目‘類義語’に登録されている

(H) 見出し語にも、別の見出し語の項目‘類義語’にも登録されていない

述語がサ変名詞+「する」の構造を持つ時、KNP ではサ変動詞として語義を決定しないで、サ変名詞として名詞辞書の語義を決定する。これは、サ変動詞に関する情報が、基本動詞辞書の付録としてしか登録されていないからである。表中(E),(F) はKNP でサ変名詞として語義が決定できた動詞を表す。本研究では動詞辞書の情報を対象とするので、名詞辞書の語義をサ変動詞辞書の語義に対応させておく必要がある。サ変動詞辞書に対応しない動詞は別の方法で解析を行なう。

6.1.4 語義決定できなかった原因と対策

語義が決定できなかった原因を考え、その対策として、語釈文が動詞を加工してからもう一度語釈文を KNP で解析する。次に原因と対策について説明する。

- 基本動詞辞書に登録されている動詞

(原因) 依存関係(係り受け)が導けない。(対象(A))

(対策) 単文化して再解析する。(結果は表 6.5)

(原因) 格情報の不足

(対策) 格を補うことで解消できるが、自動的に格を補うことは困難なので、本研究では再解析は行なわない。

- 基本動詞辞書に登録されていない動詞

(原因) 構造が動詞+助動詞の複合動詞である(対象(D))

(対策) 助動詞を削除し、受身文、使役文を通常の文になるように格を変換して再解析する。(結果は表 6.5)

(原因) 単に登録されていない (対象 (B)(F)(G))

(対策) 動詞辞書の項目「類義語」に登録されている場合、その見出し語に置き換えて再解析する。複数の候補がある場合は、KNP の語義出力でもっとも評価の高い動詞を選択する。この処理の例を次に挙げる。

(例)(汚れを)取り去る

「取り去る」の類義語 = { おとす_{1.6}, とりのぞく_{1.1}, とる_{1.11} } とする

「取り去る」を類義語に置き換えた語釈文を KNP で構文解析した結果を表 6.4 に示す。

表 6.4: 類義語に置換した場合の再解析結果例

	入力文	再解析後の語義		評価値
(1)	汚れを <u>おとす</u>	おとす	=おとす _{1.6}	100
(2)	汚れを <u>とりのぞく</u>	とりのぞく	=とりのぞく _{1.1}	10
(3)	汚れを <u>とる</u>	とる	=とる _{1.9}	77

表 6.4 中の評価値でいえば (1) > (3) > (2) の順だが、「とる」は「取り去る」の類義語としての語義と解析結果の語義が異なるので適切でない。ここで、(1),(2) のような例を「解析後の語義が一致する類義語」、(3) のような例を「解析後の語義が一致しない類義語」とする。

(B)(F)(G) の類義語を 1 つ決定する際にどのように決定したかを次のように分類する

(類義 A) 解析後の語義が一致する類義語が 1 つ以上あり、その中で最も評価値の高い類義語を選んだもの

(類義 B) 「類義 A」解析語の語義が一致せず、類義語が 1 つしかない場合

(類義 C) 「類義 A」でも「類義 B」でもなく、解析後の語義が一致しない類義語がある場合、その中でもっとも高い評価値の解析前の語義を選ぶ (前述の例で (3) しかない場合は「とる_{1.9}」を選択する)

(類義 D) どの類義語に置き換えて語釈文を解析しても、語義が決定しない場合、先頭の類義語を選択する (前述の例ではおとす_{1.6})

これらの対策を実行し、語義決定できなかった文に対して再解析を行なった結果が表 6.5 である。

表 6.5: 再解析結果

再解析方法	対象動詞数	語義決定数
単文化	504	100
複合動詞の分離	35	24
類義語への変換	1058	972

6.1.5 決定された語義の精度

KNP で決定された動詞の語義がどの程度正しいかどうかの精度を調べる。語釈文中の動詞のうち、動詞辞書に登録されている動詞のうち、270 語に対して人間が語義を決定する。これを正解として KNP の出力結果と比較したものが表 6.6 である。

表 6.6: KNP の出力における動詞の語義の精度

対象とする動詞 (A)	: 270 語
システムが語義を出力できた数 (B)	: 218
システムの出力が正解と一致した数 ($A \cap B$)	: 130
語義決定率 ($\frac{B}{A}$)	: 80.4 %
$precision(\frac{A \cap B}{B})$	: 60.0 %
$recall(\frac{A \cap B}{A})$	: 48.1 %

次に、語義が決定できなかった動詞のうち類義語に置き換えた動詞がどの程度正しいかを調べる。類義語に置き換えた 972 の動詞を前節で分類した類義 A ~ D ごとに精度を調べ、表 6.7 に示した。表中の精度とは語義決定した動詞の中で、正しい語義だった動詞の割合を示す。

表 6.7: 類義語へ変換した動詞の語義の正確さ

再解析方法	動詞数	精度
類義 A	254	85.5 % (94/110)
類義 B	295	67.6 % (71/105)
類義 C	214	57.5 % (61/106)
類義 D	209	52.4 % (44/84)
合計	972	66.8 % (649/972)

KNP で語義が決定し、最も高い評価値を選んだ分類 A は普通動詞の精度 (表 6.3 中の precision) をはるかに上回る 85.5 % であった。類義 B は KNP で語義が決定していないにもかかわらず、67.8 % の精度であった。この結果から類義 B は、サ変動詞の類義語が一つしかないので、語義の曖昧さが少ない動詞が多いと推測される。

最後に、KNP が語釈文中の動詞の語義をどの程度決定し、その語義がどの程度正しいかをまとめて表にしたものを表 6.8 に示す。

表 6.8: KNP の語義決定結果

動詞数	語義決定可能 (4191)				語義決定不可能
	前処理なし	前処理あり			1101
	3095	単文化	複合動詞	類義語	
		100	24	972	
精度	60.0 %			66.8 %	

表 6.8 より、最終的に語義が決定できた動詞は 4191 語 (79.2 %) である。精度は、類義語に置き換えていない動詞の精度が 60.0 %、類義語に置き換えた動詞の精度が 66.8 % となった。

6.2 時間関係解析システムの実装と評価

6.2.1 実装

関係を調べたい動詞で構成される 2 つの単文を入力とし、動詞の時間関係を出力するシステムを作成した (図 5.1 参照)。

6.2.2 実験の方法

入力 入力する単文は格の省略がない単文とする。省略がある場合は人間の手で補う。対象とする動詞は IPAL の基本動詞辞書の見出し語に登録されているものに限る。動詞の時制は情報として利用しない。

再帰 見出し語の語義ネットワークだけを利用して時間関係を求める場合と、見出し語の語義ネットワーク中の動詞の語義ネットワークも利用する場合の 2 通りで実験を行なう。(前者を「再帰なし」、後者を「再帰あり」とする)

出力の評価 次のような手順で評価を行なう

- (1) 人間が時間関係の正解を決めておき、システムの出力と比較する。
- (2) 出力できなかった動詞や間違った動詞の語義ネットワークを調べ、語義ネットワーク中に間違いがないかを調べる (語義や格対応間違い)。
- (3) 修正した語義ネットワークでもう一度解析を行なう。
- (4) 時制情報だけを用いた手法 (2.1 節参照) で、同様に時間関係の解析を行なう。
- (5) (4) で正解でなかった時間関係のうち、(3) でどの程度正解が得られたかを調べる。

実験の対象 入力として用いた文は以下の通りである。

- 1992 年度毎日新聞記事

基本動詞辞書 IPAL に存在する動詞だけで構成されている複文の中から 100 文抽出する。この複文を手で単文化する。ここで格の省略も補う。単文化した 100 組 (200 文) の単文を入力とする

- 物語文 (ごんぎつね)

158 文の中から、新聞記事の場合と同様に複文を抽出する。抽出された 35 文を単文化し、格を補う。単文化した 35 組 (70 文) を入力とする。

- 人手で作成した単文 32 組 (64 文)

各種文献から何らかの依存関係があると思われる動詞の組を選び、人手で単文にしたもの (32 組 (64 文)) を入力とする。

6.2.3 実験結果

出力結果を評価するため、次のような尺度を用いる。

適合率 (precision) システムの解答数のうち、何割が正解だったかを示す。

再現率 (recall) 入力した単文の組の数のうち、何割が正解だったかを示す。

各入力文に対する時間関係の解析結果を表 6.9 に示す。(A: 動詞が一致して時間関係が決定したもの B: 状態が一致して時間関係が決定したもの)

表 6.9: 時間関係解析結果

入力文	入力数 (組)	再帰なし				再帰あり			
		解答数		正解		解答数		正解	
		A	B	A	B	A	B	A	B
人手で作成	32	6	0	5	0	7	2	6	1
新聞記事	100	2	10	2	7	4	10	3	7
物語文	35	0	11	0	9	1	11	0	9

語義や、格の対応を修正した語義ネットワークで同様に解析した結果を表 6.10 に示す。

表 6.10: (語義ネットワークの修正後の) 時間関係解析結果

入力文	入力数 (組)	再帰なし				再帰あり			
		解答数		正解		解答数		正解	
		A	B	A	B	A	B	A	B
人手で作成	32	12	2	9	1	14	2	11	1
新聞記事	100	2	11	2	8	4	11	3	8
物語文	35	0	21	0	19	0	21	0	19

語義ネットワークの修正後の最終的な(再帰ありの)時間関係解析結果の再現率(recall)と適合率(precision)を表 6.11 に示す。

表 6.11: 時間関係解析結果の精度

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

6.11から言えることは、再現率はどの入力に対しても高く、適合率は、物語文以外はかなり低いという点である。適合率だけから考えると、作成した時間関係解析システムは、新聞のような記事に対してはあまり有効でないと考えられる。一方、物語文は両方の尺度で高い数値を示しているので、有効であると考えられる。

6.2.4 考察

この節ではまず、うまくいかなかった入力文に関する考察をする。次に、時制情報だけを用いて時間関係を解析する手法と比較を行なう。

入力文の内容と解析結果の関係に関する考察

結果を見て分かることは新聞記事からはほとんど依存関係を導くことができなかった。語義ネットワークの修正後で見ても、正解は約 1 割しか出力されない。一方、物語文では約半分の入力に対して時間関係を導くことができた。これは入力された文の性質による差であると考えられる。

新聞記事は、2 つの単文の主体が全く異なる場合や、2 つの単文の主体が一致していても主体が影響を及ぼす対象が異なっている場合が多い。語義ネットワーク中の動詞から抽出される状態は主体か、主体が影響を及ぼす対象の状態を示す。従って、語義ネットワーク間の状態が一致することは少なくなるため、時間関係の決定数が低かった。

例 1. 大統領夫妻がタラップ上に現れる。礼砲が響く。

例 2. 看護婦が患者の傷に薬を塗る。看護婦が患者の傷に包帯を巻く。

例 1 は主体が異なる場合である。この場合それぞれの語義ネットワーク中に、一致するイベントや状態は全くない。例 2 は影響を及ぼす対象が異なる場合である。「塗る」の文からは「薬」に関する状態「存在 (薬, 傷)」が、「巻く」の文からは「包帯」に関する状態「存在 (包帯, 傷)」が抽出される。この二つの状態が一致しないことから、状態の一致では時間関係を導くことはできないことが分かる。

一方、物語文では、ある主体の位置変化、移動先で行為などを表している場合が多いため、イベント前後の状態が一致することが多い。

例 1. 兵十が川でびくを持つ。兵十が川から土手にあがる。兵十が土手にびくを置く。...

例 2. ごんが森の中に穴を掘る。ごんが穴にすむ。

また、人手で作成した文は動詞間の依存関係が強いせいか、ほとんどが語義ネットワーク間に一致する動詞があったので時間関係を導くことができた。

時制情報のみを用いる手法との比較

時制情報のみを用いる手法を手法 1 と呼ぶ(1.1 節参照)。システムの出力と、手法 1 の解析結果とを比較する。表 6.11 中の正解の多くが、手法 1 で解析できない時間関係ならば、手法 1 の欠点を補うという、システムの利点を見つけられる。これを調べる為に、入

力文を単文化する前の原文に対して、手法 1 で時間関係を解析し、決定できない時間関係を比べる。

対象 語義ネットワークの修正後にシステムが時間関係を出力できた単文の組 (表 6.11 における新聞記事の解答数 + 物語文の解答数) の原文 (36 文) とする。人手で作成した文は原文がないので対象としない。

時間関係解析方法 基本的に語釈文を解析するのに用いた方法 (表 4.7) に基づく。しかし、接続形式が表中にないものや、主節の時制がタ形だったり、アスペクトが接続している場合があるので、表 6.12 を表 4.7 に加えて時間関係の解析を行なう。

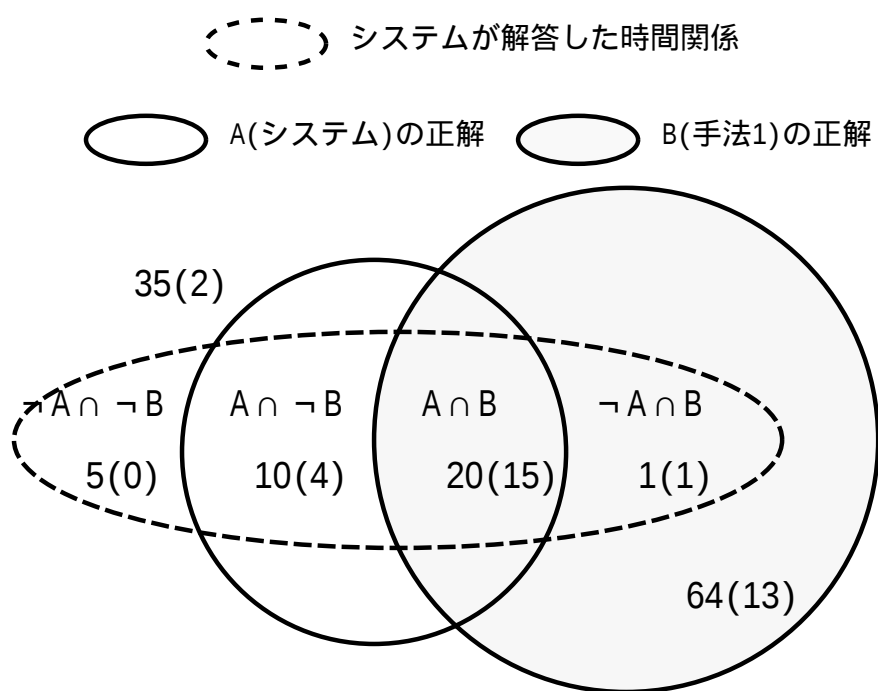
表 6.12: 接続形式による節間の時間関係 (追加)

接続形式	従属節の時制	主節の時制	主節との関係	時間関係
連体	タ形・なし	タ形・なし	発話時 or 主節の時制	after*1,before*1
	タ形・なし	ル形・テイル	より過去のイベント	after*1,contains
ト形	ル形・なし	—・なし	条件 (偶有的依存関係)	after*1

手法 1 との比較 手法 1 とシステムの時間関係の解析結果を図 6.3 に示す。

図 6.3 は全ての新聞記事と物語文の入力文に対して、出力した時間関係が正解だった数と正解でなかった数を、弁図で表したものである。図中の A はシステムの正解を表し、B が手法 1 の正解を表す。点線で描かれた楕円の内部は全体の入力のうち、システムが解答した時間関係 (合計 36) の数を表す。括弧内の数字は物語文を入力として得られた時間関係の数を示す。

新聞記事と物語文の両方を入力として考えた場合、システムだけが正解をだした時間関係は 10 ある。これはシステムの解答した時間関係の数からみれば、手法 1 で解析できなかった時間関係の数の 66.7 % である。入力文全体から見れば、手法 1 で解析できなかった時間関係の 20 % にあたる。新聞記事と物語文を分けて考えた場合、新聞記事を入力としてシステムだけが正解を出した時間関係は、入力文全体から見て、手法 1 で解析できなかった時間関係の 13.6 % にあたる。同様に、物語文の場合は 66.7 % にあたる。



括弧内は物語文だけを入力とした時の結果を表す

図 6.3: 手法 1 とシステムの出力結果の比較

新聞記事の場合 [手法 1] だけでは半数近く決定できなかった。システムはそのうちの 13.6 % しかカバーしていない。一方物語文では、[手法 1] で失敗した時間関係が少ないうえ、失敗した 6 つの時間関係のうちシステムが 4 つの時間関係を出力している。物語文は、この二つの手法を合わせるにより入力文のほとんどをカバーする。

第 7 章

結論

7.1 研究内容のまとめ

日本語基本動詞辞書 IPAL の語釈文を形式化することにより、見出し語の意味を表す語義ネットワークを作成した。その知識を利用して、依存関係のある動詞の時間関係を解析する手法を示した。実際に時間関係を解析するシステムを実装し、新聞記事、物語文を対象に時間関係を解析した。結果として、新聞記事では約 1 割、物語文では 6 割の時間関係が導かれた。新聞記事では再現率は低い、時制情報のみを用いて時間関係を解析する方法で解析できない部分の時間関係をカバーしていた。

7.2 研究の結論と今後の課題

まず語釈文を形式化するにはかなり精度の高い意味解析ができていないといけないうことが言える。しかし、語義の決定は人間が行なっても一意に決められないような場合があることから、意味解析に精度を期待するのは難しいかもしれない。

本研究では語釈文を見出し語の意味を表す語義ネットワークとして複数のイベントと状態に分けた。しかし、今回の実験の結果からも分かるように、語義ネットワーク間のイベントが一致することは非常に限られていた。つまりこの形式化の方法では、限られた動詞間の関係 (強い依存関係を持つ場合) でしか有効でないといえる。

問題は形式化の方法にもあるが、対象を基本動詞辞書に限った点にもある。動詞は具体的な概念を表すものほど、語義が少なく、格要素が省略されても、動詞単体でおおよそ

の意味を表す。逆に抽象的な概念を表す動詞ほど広い範囲で使われるため、語義を多く含み、重要な意味を持つ名詞を補わねば意味を表すことができない。

入力された動詞が抽象的な概念を表すとき、その語義を決定することが難しい上、展開される語義ネットワークは名詞句に重要な情報が移ってしまうため、本システムではあまり有効な情報が得られない。名詞句に動作を表す名詞が使われた場合、本システムでは名詞として類似度を比較する程度しか利用できない。この問題点に対する対策として、名詞の語義を利用することが考えられる。

また、IPAL の語釈文の定義動詞にはサ変動詞が使われていることが非常に多かったことも問題であった。これを基本動詞辞書中の動詞に置き換えることはやはり多少なりとも意味を変えてしまっている。語義ネットワークの対象とする動詞の数を増やすということも、今後のシステムの発展に必要であると考えられる。

謝辞

本研究を進めるにあたり、熱心な御指導をして下さった奥村学助教授に心から感謝致します。

また、國藤進教授、東条敏助教授には適切な御助言を頂き、深く感謝致します。

さらに、タナラック・ティラマヌコン助手にも、研究内容に関する御指導をしていただき、深く御礼を申し上げます。

その他、数々の貴重な御意見、御討論、システム作成に関するアドバイスをいただきました奥村研究室の皆様に心から御礼を申し上げます。

参考文献

- [西田 96] 西田元樹, 松本 裕治. 「テンス・アスペクトを考慮した現代日本語の複文の生成」情報処理学会 116-22, 1996, pp151-156.
- [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.
- [富浦 91] 富浦洋一. 「語義文に表現されている動詞間の上位-下位関係」情報処理学会 81-5, 1991, pp33-40.
- [Alterman 85] Richard Alterman. A Dictionary Based on Concept Coherence. Artificial Intelligence, 1985, pp153-186.
- [玉野 96] 玉野健一 「制約条件を用いた事象の時間構造の記述」 情報処理学会 115-2, 1996, pp9-14.
- [東条 95] 東条敏 「状況理論による法的推論の形式化」 情報処理学会論文誌 vol.36,no,1, pp.51-60, 1995.
- [吉川 89] 吉川武時. 「日本語文法入門」アルク, 1989.
- [田中 90] 田中茂範 「認知意味論」三友社出版, 1990.
- [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 96] James Allen. Natural Language Understanding. second edition, The Benjamin/Cummings Publishing Company, Inc., 1996.
- [Allen 83] Allen,J.F. “Maintain knowledge about temporal intervals”,Commun. ACM, Vol26, No.11, pp.832-843, 1983.
- [佐藤 92] 佐藤理史. 「BGHKit v0.1 マニュアル」北陸先端科学技術大学院大学佐藤研究室,1992.
- [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.
- [森部 95] 森部正二郎「動詞の意味関係に注目した複文における節間の時間関係の抽出」情報処理学会第50回全国大会, 文冊3,pp133-134,1995.

付録 A

日本語基本動詞辞書 IPAL に関する資料

A.1 意味素性

「いる」という動詞の対象 (ガ格) が人間や動物に限られており、生命のない物体 (本、箱等) は「いる」とは言えず、「ある」を用いなければならない。このような、名詞句に対する制限を「意味素性」と呼ぶ。次に IPAL で採用した意味素性を表 A.1 に示す。

表 A.1: 意味素性

略号	FULL NAME	素性名	例
CON	concrete	具体名詞	
ANI	animal <+con>	動物	犬・馬
HUM	human <+con,+ani>	人間	姉・先生
ORG	organization <+con>	組織・機関	国・企業
PLA	plant <+con>	植物	花・桜
PAR	parts <+con>	生物の部分	頭・足
NAT	natural <+con>	自然物	山・空
PRO	products <+con>	生産物・道具	紙・車
PHE	phenomenon <+con>	現象名詞	光・音
ABS	abstract	抽象名詞	
ACT	action <+abs>	動作・作用	勉強・練習
MEN	mental <+abs>	精神	心・意識
LIN	linguistic products <+abs>	言語作品	名前・ニュース
CHA	character <+abs>	性質	美・欠点
REL	relation <+abs>	関係	緑・原因
LOC	location <+abs>	空間・方角	外・公園
TIM	time <+abs>	時間	昨日・日曜日
QUA	quantity <+abs>	数量	3日・3人
DIV	diverse	(制限緩やか)	

A.2 意味分類

IPAL 中で扱われている動詞の意味分類を次に示す。

1. 状態

- (a) 存在・所有
- (b) 関係認定
- (c) 単純状態

2. 動作（動き）

- (a) （抽象的）関係
- (b) 時間
- (c) （状態）変化
 - i. 移動（位置変化）
 - A. 出発・帰着
 - B. 授受
 - ii. 出現・発生
 - iii. 消滅
 - iv. 生産
 - v. もようがえ
- (d) 設置（とりつけ）
- (e) 離脱（とりはずし）
 - i. （衣服）着脱
- (f) 接触
- (g) 加力
- (h) 生理・心理
- (i) 知覚・思考
- (j) 発見

付録 B

システムに関するデータ

B.1 語義ネットワークのデータ形式の例

```

あたためる_1_1(A,O2c,IN)
goiaspect(d3)
equal する [変化](A,O2c,NS2,NG2)
    goiaspect(d3)
    rel(A,value,A)
    rel(CO,value,O2c)
    def(O2c,value, 温度)
    def(NG2,value, 高い)
equal だく_1_3(A,O2a)
    goiaspect(d3)
    rel(A,value,A)
    rel(O2a,value,O2c)
    def(A,value, 親鳥)
before*1 かえす_1_7(A,O2c,IN)
    goiaspect(d3)
    rel(A,value,A)
    rel(IN,value,IN)
    rel(O2c,value,O2c)
    def(O2c,value, 卵)

```

図 B.1: あたためる_1_1 の語義ネットのデータ構造の例