社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

接続助詞「ため」を含む複文から因果関係知識を獲得する

乾　孝司[†]　乾　健太郎[†]　松本　裕治[†]

[†] 奈良先端科学技術大学院大学情報科学研究科
E-mail: †{takash-i,inui,matsu}@is.aist-nara.ac.jp

あらまし　大規模な電子化テキストから因果関係知識を自動的に獲得する方法について論じる．我々の目標は，事態間の因果関係が明示的に表現されている修辞的表現から，それを一貫性のある表現として成り立たせている (広義の) 因果関係に関する背景知識を同定し，蓄積することである．まず，接続助詞「ため」「のに」「たら」を含む複文を分析し，7 つの基本概念を設定した．次に，接続助詞「ため」を含む複文を対象として，主節，従属節の間に内在する関係を規則に基づいて自動判別する実験を行った．その結果，再現率は低いが，*effect* や *cause* などの関係概念では比較的高い適合率を得た．また，同一の知識源からより多くの知識を獲得する方法について検討した．

キーワード　因果関係，因果知識，知識獲得，知識表現，因果標識

Extracting Causal Knowledge from Text

The Case of Resultative Connectives “*tame*”

INUI Takashi[†], INUI Kentaro[†], and Yuji MATSUMOTO[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology
E-mail: †{takash-i,inui,matsu}@is.aist-nara.ac.jp

Abstract　This paper discusses how to extract causal knowledge from a large text collection. Our goal is to develop an automatic method of identifing the causal knowledge that supports the underlying coherence of a given causal expression. We report our ongoing attempts to design the overall problem setting and to explore the feasibility of pursue the goal, forcusing on a class of target sentences including a resultative conjunction “*tame*”. While the recall was not very high, the precision we achieved in an experiment empirically proved that our approach is promising.

Key words　causal relation, causal knowledge, knowledge acquisition, knowledge representation, causal marker

1. はじめに

我々は日常生活の中で，さまざまな因果関係で結びついた無数の事態を経験している．因果関係という概念は，古代ギリシャ哲学の時代から今日にいたるまで心理学，哲学をはじめ，多くの学問領域で研究対象になってきた．人工知能の領域でも，初期の頃から知能の源として常識・因果知識に関心が向けられ，それらの知識を蓄積した知識ベースに関する研究が活発になされてきた [2] [9] [22]．知識ベース構築の方法論については，CYC プロジェクト [15] や OpenMind プロジェクト [23] を含む様々な試みが報告されているが，常識的知識を人手でどこまで書き尽せるかは依然として未知数である．

このような背景から，近年，WWW の普及とともに急増する大量の電子化テキストを知識源と見なし，そこから因果知識・世界知識を獲得する試みがいくつか報告されるようになってきた [5] [7] [20]．試みに，WWW 検索エンジンを用いて「”ため”，”洗濯物がよく乾く”」や「”ので”，” 洗濯物がよく乾く”」などをキーワードとする検索を実行すると，図 1 のような文を含む文書が返ってくる．これらの文は，どういう状況で洗濯物が乾きやすく，またどういう状況で洗濯物が乾きにくいかという因果関係を表している．こうした文をうまく検出し，加工すれば，比較的安いコストで大量の因果関係に関する知識 (以下，因果知識) を獲得できる可能性がある．

図 1 のようなテキストから因果知識を自動的に獲得・蓄積し，深い言語理解の実現に利用することが我々の目標である．その第 1 段階として，明示的に因果関係を表す手掛かり語を含むテキストを対象とし，テキストの背後に潜む因果関係を自動的に判別できるかどうかについて分析を進めてきた．本稿では，これまでに得られた知見や明らかになった課題を報告する．

‥天気が良い ので 、洗濯物がよく乾く。‥
‥梅雨空が続いた ため 、洗濯ものが乾かず、‥
‥湿度はなかった から 、洗濯物が乾く。‥
‥梅雨な のに よく乾く, ‥
‥3月27日水曜日 風が強く て 洗濯物がよく乾く。‥
‥ちゃんとベランダに干して たら 乾くのだろうけれども‥
‥乾燥機かけちゃえ ば 、乾くでしょう。‥
‥カラッと晴れてくれる と 、洗濯物が乾くんだけどな〜。
‥意外にも陽当たりが良 く 、結構洗濯物がよく乾く。‥

図 1　Web 上のテキスト

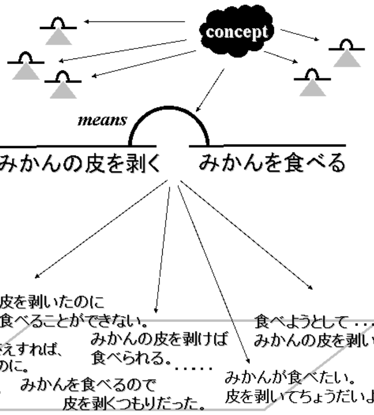
concept
means
みかんの皮を剥く
みかんを食べる
.....
皮を剥いたのに
食べることができない。
食べようとして .....
みかんの皮を剥いている。
.....
皮を剥きさえすれば、
食べれたのに。
みかんの皮を剥けば
食べられる。
.....
みかんが食べたい。
皮を剥いてちょうだいよ。
.....
みかんを食べるので
皮を剥くつもりだった。

図 2　世界の事態関係と修辞関係

2. 目標・問題設定

2.1 目標: どういう知識を獲得したいのか

次のような例を考えてみよう．

(1) a. みかんを食べるためにていねいに皮を剥いた．
b. みかんは皮を剥かないと食べられない．
c. みかんの皮を剥いたのに結局食べることができなかった．

各文の主節と従属節の間にはそれぞれ異なる修辞関係が認められる．たとえば，(1a) の従属節「みかんを食べるために」はふつう目的を表す節と解釈される．(1b) の従属節は仮定，(1c) は逆接である．ここでは，「みかんの皮を剥く」という行為と「みかんを食べる」という行為の関係の仕方がさまざまな修辞的表現を使って言語化されている．我々がこれらの修辞的表現をどれも同様に一貫している (coherent である) と感じるのは，どの文も

(2) 「みかんの皮を剥く」という行為が「みかんを食べる」という行為の〈手段〉である

という我々の世界知識と整合しているからだと考えられる．このことは，(3) のような一貫性のない例を考えると明らかである．

(3) a. みかんの皮を剥いたのに結局食べることができた．
b. みかんの皮を剥いたのに雨が降ってきた．

我々の目標は，(1) のような修辞的表現から，それを一貫性のある表現として成り立たせている (2) のような (広義の) 因果関係に関する背景知識を同定し，蓄積することである．図 2，図 3 に示すように，こうした因果知識は具体的な修辞的表現を一段抽象化したものと考えることもできる．ここで抽象化の際に捨象される情報は，次のようなモーダル情報や情報構造に関する情報である．

- テンス・アスペクトに関する情報: 主節，従属節が表す事態が完了した事態なのか，これから起こることが予測される事態なのか
- 認め方に関する情報: 主節，従属節が表す事態が実際に起こったのか，起こらなかったのか
- 実行可能性に関する情報: 主節や従属節が表す行為が実行可能かそうでないか，
- 情報構造に関する情報: 主節，従属節が表す事態のどちらに焦点が当たっているか．何が新情報か

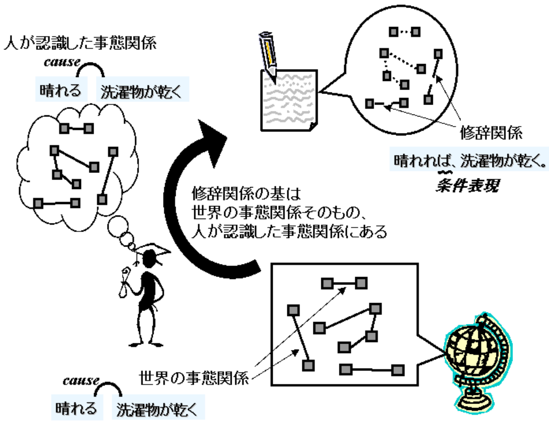
人が認識した事態関係
cause
晴れる
洗濯物が乾く
修辞関係
晴れれば、洗濯物が乾く。
条件表現
修辞関係の基は
世界の事態関係そのもの、
人が認識した事態関係にある
世界の事態関係
cause
晴れる
洗濯物が乾く

図 3　世界の事態関係と修辞関係 (2)

表 1　因果標識

言語標識	構成単位	例文
接続詞	文	ニューロ電算機では性能を示すのに最適なのは学習速度。だから 一秒間にどれだけの学習が出来るかを示す単位をつくったわけだ。
接続助詞	節	同社はコメの販売に弾みをつける ため、今後、欧州で漬物やみそ、しょうゆも生産する。
		山ユリは食用になる ので、将来は特産品づくりも目指す。
格助詞	句	入社後四十年のうち、病気 で 休んだのは三日だけ。
動詞	句	流通分野の規制緩和が新しいビジネスチャンスを 生み出す

こうした知識が十分に蓄積できれば，

(4) みかんの皮，剥いてくれる？

のようなテキストの背後に隠れた発話者の意図を推定するといった深い言語理解が実用規模で実現できる可能性が拓けてくる．

2.2 問題設定

図 4 に示すように，テキストから因果知識を獲得する作業を次の 2 つの部分タスクに分けて考える．

（1） 事態認定課題．テキスト中の文字列から任意の事態を表現している部分形態素列を認定する．おおよそ一つの節がそれぞれ一つの事態の候補となる．

（2） 関係判別課題．複数の事態が与えられた時，どの事態間に関係が存在するか，またその関係はどういう種類の関係かを判別する．文内に限れば，統語構造や係り受け構造は有益な

表 2　接続助詞の出現頻度 (新聞記事 1 年分)

が	7450	たら	336
と	5179	から	325
(れ) ば	3362	なら	134
ため (に)	3252	のに	109
ながら	862	φ、	18886
ので	514	て、	4088

情報である．また，表 1 に例示した言語標識も手掛かり語句として利用できる．

本稿では，表 1 に示したような因果関係を明示する言語標識 (因果標識) を手掛かりとして用い，因果標識で関係付けされた事態のみに着目する．さらに次の理由から，今回は因果標識のうち接続助詞「ため」を含む複文 (タメ複文) に考察対象を絞る．

- 事態は基本的に用言と幾つかの格要素からなる．文，節，句のうち，節はこれらの情報を過不足なく持っている．
- 第一義的に因果関係を明示する接続助詞の中で出現頻度が最も高い (表 2)．

以上を整理すると，本稿で取り扱う課題は次のようになる．

タメ複文を構成する主節，従属節から事態をそれぞれ認定し，そこから言語表現の背後にある因果関係を判別する (図 4)．

2.3　知識表現

表現上の制約により獲得した知識が欠落することがないよう，十分に表現能力の高い知識表現を採用するが大切である．例えば，格フレーム表現は事態の骨格を表現する最適な表現系の一つであると考えられる．ただ

- 十分量の格フレームをあらかじめ用意できるか
- 常に適切な格フレームを選択できるか
- 修飾要素などの骨格以外の情報をどう格納するか

などの問題を解決する必要がある．

本稿では，格フレーム表現での問題点を考慮し，獲得した因果知識を自然言語に基づく緩い形式を持った表現系で記述する (図 5)．

$$\langle rel\ ,e1\ ,e2 \rangle.$$

rel は関係概念を，$e1$ と $e2$ は関係概念 rel で結ばれる各事態を表す．因果関係では前件事態と後件事態の対を把握しておくことが必須である．そこで事態が主節，従属節のどちら側に表出された事態であるかに関 わらず，2 つの事態のうち，先行して生起した事態を $e1$ スロットに格納し，後続して生起した事態を $e2$ スロットに格納する．

各々の事態 e は，p と mas (modal attributes) の組から構成され，p には命題，mas には可能，否定，受動，使役などのモーダル属性が入る．属性空間が閉じているモーダル属性に関してはあらかじめスロットを用意しておく．一方，命題には自然言語を知識表現として採用する．自然言語を知識表現にすると，知識自体に曖昧性が生じる，知識の探索，適用時の負荷が大きいなどの問題もある [8] [13] が，高い表現能力を得ることができるため，少なくとも格フレームで生じる問題には対処できる．以下では，獲得される三つ組 $\langle rel\ ,e1\ ,e2 \rangle$ を便宜的に，因果フレームとして参照する．

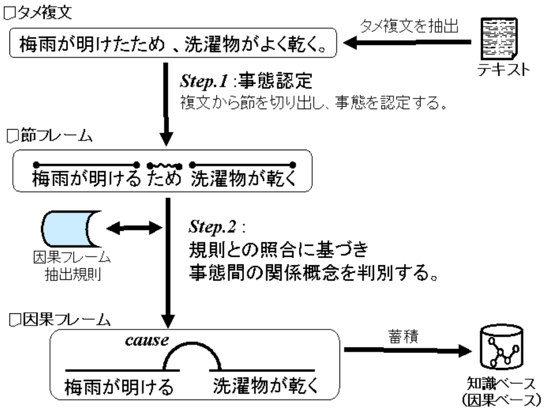
□タメ複文
梅雨が明けたため 、洗濯物がよく乾く。
タメ複文を抽出
テキスト
Step.1 :事態認定
複文から節を切り出し、事態を認定する。
□節フレーム
梅雨が明ける ため 洗濯物が乾く
因果フレーム
抽出規則
Step.2 :
規則との照合に基づき
事態間の関係概念を判別する。
□因果フレーム
cause
梅雨が明ける
洗濯物が乾く
蓄積
知識ベース
(因果ベース)

図 4　知識獲得の処理概要

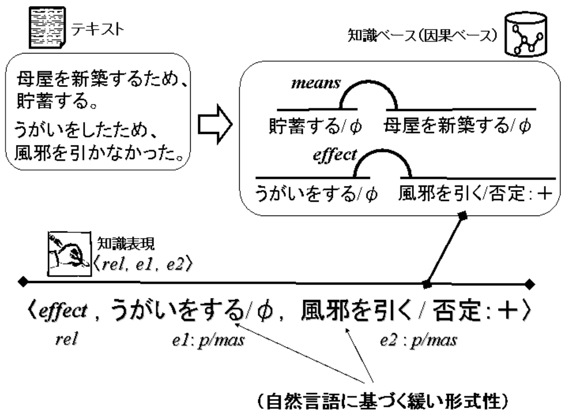
テキスト
母屋を新築するため、
貯蓄する。
うがいをしたため、
風邪を引かなかった。
知識ベース(因果ベース)
means
貯蓄する/φ
母屋を新築する/φ
effect
うがいをする/φ
風邪を引く/否定:＋
知識表現
⟨rel, e1, e2⟩
⟨effect , うがいをする/ φ, 風邪を引く / 否定: ＋⟩
rel
e1: p/mas
e2 : p/mas
(自然言語に基づく緩い形式性)

図 5　知識表現

3.　関係概念の定義

本節では，因果の関係概念を定義する．

3.1　関係概念

安部ら [1] は次のように連接関係における概念設定の難しさについて言及している．

> “連接関係 (coherence relation)” とは，文章を構成する要素である文と文，あるいは複文における節と節との間ばかりでなく，より大きな単位，つまり意味内容上のまとまりをもった文の集合 (文群) の間の意味的なつながり関係をいう．(省略)　文章の読み手が連接関係をどのように理解しているか，その過程を説明することは難しい．そのことは，前節で紹介した連接関係とはどのような関係かという議論がそもそもまだ収束していないということからも推測できるであろう．

また，必要な関係概念の分類体系はおそらくアプリケーションに依存するため，万能な関係概念を設定することは困難である．

そこで，我々は，古典的な知識集約型の談話理解研究 (e.g. [22] [4] [9]) に立ち返り，多くの研究で比較的合意されている分類基準から関係概念の分類を再構築することにした．まず，もっとも基本的な分類は，因果関係に立つ 2 つの事態がそれぞれ意志的 (な行為) であるか，意志的でないかという基準に基づく分類である．以下簡単のため，便宜的に，意志的な行為 (たとえ

ば「薬を飲む」) を表す事態を単に行為と呼び，意志的でない事態 (たとえば「風邪が治る」) をすべて状態と呼ぶことにする．多くの談話理解研究が指摘しているように，言語理解において行為と状態の区別は重要である．行為には主体の意図がその背後にあり，そうした意図を理解することが談話理解の主要なゴールの一つとされてきた．

ここで，因果関係に立つ 2 つの事態のうち，時間的に先行する事態を $e1$，もう一方の事態を $e2$ とし，意志性の有無を $+/-$ で表すとしよう．このとき，$(e1, e2)$ の組み合わせは，$(e1, e2) \in \{(+,+),(-,+),(+,-),(-,-)\}$ のいずれかである．さらに，$e1$，$e2$ がともに有意志行為であるとき，それらの行為の主体が同一かどうかの区別も必要であろう．自分の意志で両方の行為をする場合と，他人が自分の意志とは独立に一方を行為する場合では，意図構造が明らかに異なるためである．

以上の議論から，少なくとも表 3 の左欄に示すような 5 種類の因果関係を区別する必要がある．これら 5 種類の因果関係は，おおよそ従来の談話理解研究で言及されてきた次の 5 つの概念に対応すると予想できる:

means，*enable*，*precondition*，*effect*，*cause*

つぎに，この対応づけがどの程度正しいか，またテキスト中から収集できる因果関係の種類は上の 5 種類で尽くされているかを調べるために，以下の要領で実際のテキストを分析した．

分析対象には，タメ複文の他に接続標識として「のに」「たら」を含む複文を選択した (「ため」200 件，「のに」71 件，「たら」79 件)．表 3 と表 4 に関係概念の定義と各関係の例文を示す．結果として，表 3 に示すように，以下の 3 つ観点によって大きく 7 つの概念に分類することができた．第一の観点は，行為と状態の対立であり，事態の対との組み合わせでまず 4 つに分類する．これにより，狭義の意味での因果関係 (*cause*: 食べなかったら痩せる) や，行為の結果どのような状態なるか (*effet*：薬を飲めば，風邪が治る) であったり，ある行為をするために満たすべき条件 (*precondition*：雨が降ってきたので傘をさす) などが大きく分類できる．

第二に，主体の同一性を観点として採用した．この観点は，上の 4 つの分類のうち主節と従属節が共に意志性:+である事例を 2 つに分類する．このうち，主体が同一である場合は，目標行為のための副次的行為 (*means*：電車に乗るため切符を買う) などが新たに分類される．また，前件事態に対して後件事態の独立性が比較的強い事例が数件あった．これらの事例は「X が A したので X が次に B する」と言い換え可能かどうかという言語テストをパスするが [X が B するためには X が A しなければならない」という言語テストにはパスしないもの (*succession*：食事したため歯をみがく)，絶対時間軸上での事態 (*preparaton* 明日はクリスマスなので，ワインを買った) などである．

3.2　因果関係の連鎖の省略の扱い

テキスト中に明示的に表現される因果関係の中には，次の例のように，より粒度の細い因果関係の連鎖に分解してもよさそうな例も少なくない．

(5)　a.　食事したので，歯をみがく．

b.　食事する-(歯が汚れる)-(虫歯になる)-(虫歯を防ぐ)-歯をみがく

表 3　関 係 概 念

主体意志性				
$e1$	$e2$	主体同一性		(独立性:+)
$+$	$+$	$+$	*means*	*succession*
$+$	$+$	$-$	*enable*	*preparation*
$-$	$+$		*precondition*	
$+$	$-$		*effect*	*preparation*
$-$	$-$		*cause*	

知識を獲得する際，仮にこのような関係連鎖のショートカットを復元しようとすると，どこまで還元するかを決定するという難しい問題に直面することになる [21]．また，ショートカットは，観察者が因果関係の連鎖の一部分に焦点を当て，全体を要約した結果であるともみなせる．このとき，ショートカットされた因果関係は人が認知する因果関係の基本単位となっており，還元をしないそのままの関係に意味があると考えられる．そこで，今回はショートカットの還元のことは考えず，表出されたままの関係を捉えることにする．

3.3　関係概念の再現性

前項で定義した関係概念を判別するためには，主体の意志性，事態の生起順序，主体の同一性の 3 つの素性が被験者内あるいは異なる被験者間を通じて一貫して決定できなければならない．このうち，事態の生起順序，主体の同一性を把握することは人間がテキストを理解するうえで本質的であり，異なる被験者間でも高い再現性が得られると期待できる．しかしながら，主体の意志性に関しては「意志的な行為か」どうかを常に意識しているとは限らない．特に，被験者間では大きな揺れが生じる可能性もある．

そこで，被験者に比較的短い単文を提示し，そこに記述された事態が主体にとって意志的な行為であるかどうかを判定してもらい，κ 統計量を見積もることで被験者間の関係概念の再現性を調査した．κ 統計量 [3] とは，被験者間の判断の一致性を見積もる指標であり，この値を見積もることで，偶然一致の割合を考慮した一致性を見積もることができる．

被験者 4 名 (筆者を含む) による 180 件の判定課題の結果，各被験者間の κ 値は，0.879 ~ 0.929 となった．この結果は意志性の判定に高い一致度が認められたことを示しており，定義した関係概念にもある程度の再現性があることを示唆している．

4.　関係概念の判別の手続き

4.1　判別の手続き

本節では，前節で定義した因果の関係概念を自動判別する手続きを示す．判別の手続きは次の二つのステップからなる (図 4)．

Step.1　事態認定．このステップでは，タメ複文から，主節，従属節のそれぞれの構成要素を切り出し，そこから事態 (表現) を認定する．この出力結果をここでは便宜的に節フレームとして参照する (図 6)．

Step.2　関係判別．節フレームとあらかじめ作成しておいた規則との照合によって因果フレームを獲得する．

以下，それぞれのステップの詳細について述べる．

4.1.1　Step.1　事態認定

事態認定は，次のような処理の流れをとる．

表 4　関係概念 (例)

入力テキスト	rel	$e1$: p/ mas	$e2$: p/ mas
会社に行くのに、団地から最寄りの駅までバスに乗る。	$means$	X が団地から最寄りの駅までバスに乗る/φ	X が会社に行く/φ
その理由を調べるため、もちを顕微鏡で観察した。	$means$	X がもちを顕微鏡で観察する/φ	X がその理由を調べる/φ
大阪地検が特別抗告を断念したため投票用紙は選管に返還されている。	$enable$	大阪地検 (X1) が特別抗告を断念する/φ	X2 が投票用紙を選管に返還する/φ
生産者団体も一斉に反発したため、計画を白紙に戻した。	$enable$	生産者団体 (X1) が一斉に反発する/φ	X2 が計画を白紙に戻す/φ
だから、枝が伸びてきたら切られてしまうの。	$precondition$	枝が伸びる/φ	X が切る/φ
ぎりぎりまで制限が残るため保険会社を保護できる。	$precondition$	ぎりぎりまで制限が残る/φ	X が保険会社を保護する/φ
ただ、それに合わせていたら経営が成り立たない。	$effect$	X がそれに合わせる/φ	経営が成り立つ/[否定:+]
一方公務員を大量解雇したため人材不足の問題も抱えている。	$effect$	X が公務員を大量に解雇する/φ	人材不足の問題を抱える/φ
被害者側が警察に届けたため、成功しなかった。	$effect$	被害者側が警察に届ける/φ	成功する/[否定:+]
一方、底は太陽の光が永遠に届かないため，零下二百度以下に冷やされています。	$cause$	光が届く/[否定:+]	底は零下二百度以下に冷える/φ
十二月になったら新海苔のにおいがぱっーとしますでしょう。	$cause$	十二月になる/φ	新海苔のにおいがぱっーとする/φ
日が差しているのに，雨が降っている。	$cause$	日が差す/φ	雨が降る/[否定:+]
人口は増加しているのに、米の消費料は増えていない。	$cause$	人口が増える/φ	米の消費料が増える/φ

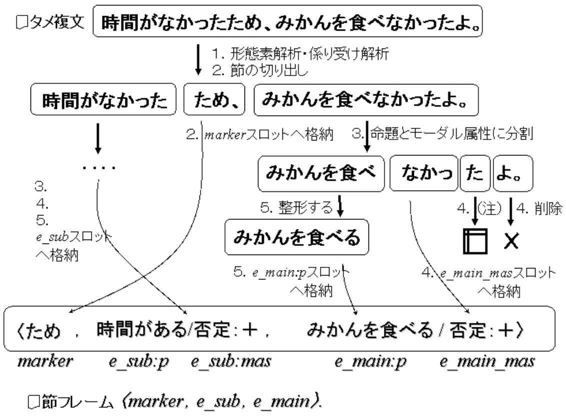
□タメ複文
時間がなかったため、みかんを食べなかったよ。
1. 形態素解析・係り受け解析
2. 節の切り出し
時間がなかった
ため、
みかんを食べなかったよ。
2. markerスロットへ格納
3. 命題とモーダル属性に分割
・・・・
3.
4.
5.
e_subスロットへ格納
みかんを食べ
なかっ
た
よ。
5. 整形する
4. (注)
4. 削除
みかんを食べる
5. e_main:pスロットへ格納
4. e_main_masスロットへ格納
〈ため，時間がある/否定:＋，みかんを食べる / 否定:＋〉
marker
e_sub:p
e_sub:mas
e_main:p
e_main_mas
□節フレーム〈marker, e_sub, e_main〉.

図 6　step1．処理のながれ

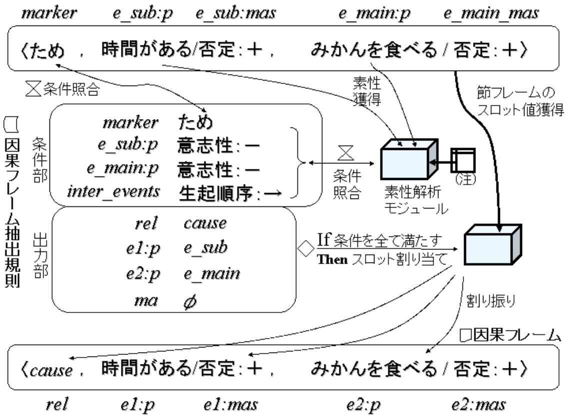
marker
e_sub:p
e_sub:mas
e_main:p
e_main_mas
〈ため，時間がある/否定:＋，みかんを食べる / 否定:＋〉
条件照合
素性獲得
節フレームのスロット値獲得
因果フレーム抽出規則
条件部
marker　ため
e_sub:p　意志性:ー
e_main:p　意志性:ー
inter_events　生起順序:→
条件照合
素性解析モジュール
(注)
出力部
rel　cause
e1:p　e_sub
e2:p　e_main
ma　φ
If 条件を全て満たす
Then スロット割り当て
割り振り
□因果フレーム
〈cause，時間がある/否定:＋，みかんを食べる / 否定:＋〉
rel
e1:p
e1:mas
e2:p
e2:mas

図 7　step2．処理のながれ

1　平文に対して，形態素解析 (chasen [16])，係り受け解析 (KNP [12]) をそれぞれ行なう．

2　接続助詞を探索し，前後の文字列をそれぞれ，従属節，主節として切り出す．また，接続助詞を $marker$ スロットに格納する．

3　従属節，主節のそれぞれの節要素について，正規表現によるマッチングを取り，命題表現，モーダル表現を同定する．

4　以下に示すモーダル表現は，mas スロットに格納し，その他のモーダル形態素は削除する(注1)．

5　命題表現を整形し，p スロットへ格納する．

5. 節の実験では，試験的に，モーダル属性 mas スロットに格納する形態素を次のように設定した．丸括弧内は各属性の属性値情報である．

- 可能表現 (+/−)：帰省 できる
- 否定表現 (+/−)：帰省し ない

4.1.2　Step.2　関係判別

事態認定過程より得られた節フレームとあらかじめ人手で作成しておいた規則 (因果フレーム抽出規則) との照合に基づいて因果フレームを獲得する．因果フレーム抽出規則は，条件部と出力部から構成されている (図 7，表 9)．条件部には，主節と従属節の事態が関係概念 rel となるための条件が記述されており，出力部には，判別された関係概念 rel の他に，節フレームの e_sub，e_main スロットと因果フレームの $e1$，$e2$ スロットとの対応関係が書かれている．また，モーダル属性を付加，削除するためのモーダル演算子 mo (modal operator) も記述されている．規則には優先度が付いており，複数の規則が適用可能な場合は，優先度の高い規則を適用する．

4.2　素性獲得

表層的な手掛かりを除くと，今回作成した関係概念の判別条件は次の 3 つの素性から成る．

- 主体の意志性 (+/−)．主体の意志的な行為であるかどうか．意志的な行為 (+)/意志的な行為でない (−)
- 事態の生起順序 (→/←)．従属節，主節のどちらの事態が先に生起したか，あるいは，どちらが先に生起すると予想されるか．従属節が先に生起または同時 (→)/主節が先に生起 (←)．
- 主体の同一性 (+/−)．主節，従属節の事態が共に意志

(注1): ここで削除したモーダル形態素は抽出規則の条件部で指定された素性を推定する際に利用するため，実際の処理では節フレーム以外の記憶装置にモーダル情報を保存している．

的な行為である場合，その行為主体が同一であるかどうか．同一である $(+)$/異なる $(-)$

以下，3 つの素性の素性値の推定法について簡単に述べる．

4.2.1 主体の意志性

動詞の意志性(注2)に関する情報は，IPAL 動詞辞書 [25] に存在する．しかしながら，IPAL 辞書はその信頼性が高い一方で辞書としての規模は小さく，実用上，その利益を受けるのは難しい．そこで，IPAL に含まれていない動詞については，独自の意志性判定アルゴリズムによって自動推定し，辞書知識の拡大を計った．

具体的に取った手続きは割愛するが，手続きを要約すると，まず，IPAL 辞書，EDR 日本語辞書，同概念辞書 [27] の情報を基に Seed 辞書を作成する．その後，辞書の見出し語と語義文に含まれる動詞の関係を利用して単語間ネットワークを構成し，ネットワークの活性伝播によって Seed 辞書を洗練した．この手続きによって，約 3 万の動詞意志性辞書を獲得した．

また，一般に主体に意志性があるかどうかは静的には決定できず，文脈に依存するケースも少なくない．今回は，ヒューリスティクスとして，次の手がかりが出現する場合は，動詞の意志性とは無関係に決定的に主体の意志性を特定した．

- 受身表現 - 意志性:$+$
- 否定表現/可能表現/アスペクト表現「テアル」 - 意志性:$-$

静的辞書とヒューリスティクスの適切さを見積もるために，3. 節での再現性確認のために用いた意志性データのうち 4 人の被験者の判断が一致した 158 事例 (意志性$+$:102 事例/意志性$-$:56 事例) を用いて，推定精度を評価したところ，約 80%(114 事例) が被験者の判断と一致していた．

4.2.2 事態の生起順序

2 つの事態の生起順序を知る一つの手掛かりとして，従属節テンスがある．従属節のテンスの取り方には大きく発話時を基準として表すか，主節事態を基準として表すかで，絶対テンスと相対テンスの二通りがあるが「ため」の場合は相対テンスの傾向が強い [10] [19]．つまり，従属節のテンス表現がタ形であれば，従属節事態が起こった後，主節事態が起こり，ル形であれば，その逆となる．そこで，今回は一部の例外条件(注3)を除き，デフォルトとして相対テンス解釈をすることにし，事態間の生起順序を定めることにする．意志性の場合と同様の評価をおこなったところ，人による判断との一致度は 88%であった．

4.2.3 主体の同一性

複文における主体の同一性に関する研究は既に幾つかあり [17] [18]，ある程度の精度で推定可能である．ただ，今回は，先行研究で得られている知見を実装するには至らなかったため，明示的に表現されていない場合は，省略表現に注目したヒューリスティック (表 5) のみで主体の同一性を推定した．人による判断との一致精度は 63% であった．

5. 判別実験

本節では，3. 節で定義した関係概念の自動判別実験の結果について述べる．

5.1 セッティング

Web を知識獲得の対象にすることが将来的には考えられるが，Web 上のテキストは崩れた表現も多く，データの質は必ずしも高くない [21]．そこで，今回は新聞記事を対象データとした．

まず，日本経済新聞 [26] から取り出したタメ複文に対して，関係概念を人手で付与した．その内の一部を人手で分析することで，11 個(注4)の因果フレーム抽出規則を作成し (表 9(注5))，その後，作成した抽出規則をタメ複文に適用して因果関係概念を自動判別した．

命題情報の表現に自然言語表現を採用したことにより，入力の自然言語テキストから命題を抽出する操作は，一種の「言い換え」処理となっている．そのため，一連の判別手続きを汎用的な言い換え処理環境 KURA [24] 上で実装した．

5.2 判別結果

表 6 に規則作成の際に用いたタメ複文による判別結果を，表 7 に規則作成の際に用いなかったタメ複文による判別結果をそれぞれ示す．

関係概念ごとに出現頻度が異なっているが，比較的頻度の多かった *means*，*effect*，*cause* では高い適合率が得られた．その一方で，出現頻度の少ない関係概念については良い成績は得られなかった．全体を通して再現率は低いが，知識源となるテキストの規模を考慮すると，再現率の低さはそれほど問題ではなく，むしろ，ある程度の出現頻度を持ちながら適合率の低い関係概念 (*precondition*) に対する精度を向上させる必要がある．

“保留” は，どの抽出規則にも適合しなかった事例の数を表している．保留された事例の多くは，意志性の素性値が決定できないために保留されていた．今回，意志性推定アルゴリズムの適用によって意志性辞書の拡大を計ったが，それ以上に異なる表現が新聞記事上には存在していたことになる．

5.3 議論

テキストから知識を獲得することを仮定した場合，その知識量は基本的に知識源であるテキストの規模に依存することになる．既に相当量の電子化テキストが存在していることから，知識の自動獲得アルゴリズムがあれば，ある程度の規模の知識が蓄積できると考えられる．しかしながら，どれだけの規模の知

表 5 主体の同一性

条件	主体同一性
受動表現，使役表現がある	$-$
従属節，主節の主体が共に省略	$+$
従属節のみ主体が省略	$+$
主節のみ主体が省略	$-$

（注2）: ここで，動詞の意志性とは動詞が静的に持っている意志素性を指し，主体の意志性とは，動詞を含めた事態全体での動的な意志素性を指す．モーダル属性のない事態では，大抵の場合，動詞の意志性と主体の意志性は同じ値となる．以降，単に「意志性」と書いたときは，主体の意志性を指す．

（注3）: 従属節が次の 4 つの素性をもつ場合は，従属節が先に生起 (→)．否定表現．可能表現．意志性:$-$．アスペクト表現「テイル/テクル/テアル」．

（注4）: 今回作成した規則は，概念を定義する際に用いた情報と重複した素性をそのまま用いているため，規則の数はそれほど多くない．その他の因果標識 (ので，から，のに) のための規則を作成する場合も同様に少ない規則数で済むわけではない．

（注5）: 「のに」に対する因果フレーム抽出規則を末尾行に示す．タメでは用いられなかったモーダル演算子が出力部に記述されている．

表 6　関係概念の判別結果 (closed)

	recall	precision
means	0.60(26/43)	**0.93(26/28)**
enable	0.42(5/12)	**0.83(5/6)**
succession	0.00(0/3)	0.00(0/3)
preparation	0.33(1/3)	0.17(1/6)
pcondition	0.48(10/21)	0.43(10/23)
effect	0.41(7/17)	**0.88(7/8)**
cause	0.52(16/31)	**0.73(16/22)**
保留		34 件

表 7　関係概念の判別結果 (open)

	recall	precision
means	0.62(24/39)	**0.86(24/28)**
enable	0.00(0/5)	0.00(0/1)
succession	0.00(0)	0.00(0/2)
preparation	0.67(2/3)	0.50(2/4)
pcondition	0.58(15/26)	0.50(15/30)
effect	0.53(8/15)	**0.89(8/9)**
cause	0.50(13/26)	**0.72(13/18)**
保留		22 件

識があれば十分であるかは不明であるため，利用できる知識源に対し，可能な限り多くの知識を獲得する手法を検討することは重要である．以下では，知識の規模拡張に関する 2 つの手法について考察する．

5.3.1　因果フレーム間の推論

ある条件を満たす因果フレームは，その命題の素性やモーダル属性を変更することで異なる因果フレームに変換することができる．表 8 に例を示す．変換規則 1，2 は，*e1*，*e2* が共に否定のモーダル属性をもつ場合には，否定のモーダル属性を相殺した因果フレームに変換できることを示している．また，変換規則 3 の例は，他動詞/自動詞の対応をもつ動詞が命題として含まれていれば，自他を反転させた因果フレームに変換できることを示している．このように変換可能な条件を事前に知ることができれば，事例としては出現していない因果フレームを自動的に生成できるようになり，因果フレーム抽出規則と組み合わせることで同一のテキストからより多くの知識を獲得できる可能性がある．

5.3.2　精神的活動を含む命題の展開

新聞記事では「検討する」や「図る」「示す」といった，発言や思考などの精神的活動を表す表現が比較的多く出現していた．このような精神的活動を表す命題を含む文は次のように展開文に言い換えることで事例規模を自動的に増やすことができる．

(6)　a.　【元文】予想以上に収益が上回ったため予算配分の調整を図る．
　　 b.　【展開文】予想以上に収益が上回ったため予算の配分を 調整する．

しかしながら，次の例のように精神的活動に焦点が当たっている場合では，展開してしまうと元の文意を損なってしまうため，常に展開が可能なわけではない．

(7)　a.　【元文】米大統領がペルシャ湾岸への 米軍増派を発表したため朝方からドル買いが先行した．
　　 b.　【展開文】米大統領がペルシャ湾岸へ 米軍を増派するため朝方からドル買いが先行した．

現在，展開の可能性をもつ動詞を 100 件ほど集め，事例分析を行なっており，今後，展開の条件について考察を行なう予定である．

6. 関連研究

1990 年代初までは，談話理解などの領域において，テキスト中の事態間の非明示的な因果関係を推定する研究が行なわれていた．知識ベース，あるいはそれに基づく推論を仮定した方法論が採られていたが，この時に必要な知識ベースは人手であらかじめ作成したものを利用していた (例えば [11])．これとは対照的に，近年では，知識ベースの存在は仮定せず，また注目する対象も非明示的なものではなく，言語的な手掛かり語句が含まれるような明示的な因果関係から知識を獲得する手法が積極的に検討されている [7] [6] [14]．本節では，近年の先行研究の中で特に我々の考えと近い，Khoo ら [14] と Girju ら [7] の研究事例を紹介する．

まず，Khoo ら [14] は，医療関係のテキストから構成される Medline database を対象とし，*cause*，*effect* などの因果関係の概念要素をテキストの構成要素に対して割り振る課題を取り扱っている．彼らのアプローチでは，まず対象文から構文構造木を作成する．そして，文内の因果標識を頼りに木構造のどのノードが *cause* や *effect* などの概念要素になりやすいかをパターン化したグラフィカル・パターンを作成し，このパターンとの照合に基づいて因果知識を獲得することを試みている．

Girju ら [7] は, “cause” や”force” といった causative verb を挟む二つの名詞句 (NP1,NP2) がそれぞれ NP1=*cause*，NP2=*effect* となる (NP1 verb NP2) 組のパターンを作成し，それとの照合に基づき因果知識を自動獲得することを試みている．

Khoo ら，Girju らを含め，テキストからの因果知識獲得に関する諸事例は，始動しはじめたばかりであり，その技術は洗練されたものではない．しかし，機械学習，テキストマイニングなどの手法を採り入れることで，今後の進展は十分期待できる．

7. おわりに

大規模な電子化テキストから因果関係知識を自動的に獲得する方法について論じた．我々の目標は，事態間の因果関係が明示的に表現されている修辞的表現から，それを一貫性のある表現として成り立たせている (広義の) 因果関係に関する背景知識を同定し，蓄積することである．まず，接続助詞「ため」「のに」「たら」を含む複文を分析し，7 つの基本概念を設定した．次に，接続助詞「ため」を含む複文を対象として，主節，従属節の間に内在する関係を規則に基づいて自動判別する実験を行った．その結果，再現率は低いが，*effect* や *cause* などの関係概念では比較的高い適合率を得た．また，同一の知識源からより多くの知識を獲得する方法について検討した．

今後の展開としては，以下の事項について検討していく予定である．

- 異なる因果標識による検討．
- 素性 (主体の意志性など) の推定精度の向上．
- 適用時の命題間の照合アルゴリズムの開発．

表 8　因果フレーム間の推論

変換規則 1	$\langle \mathit{cause}, \mathit{p1}/\text{否定}:+, \mathit{p2}/[\text{否定}:+, \text{可能}:+]\rangle.$	$=$	$\langle \mathit{precondition}, \mathit{p1}/\varphi, \mathit{p2}/[\text{可能}:+]\rangle.$
	$\langle \mathit{cause}, \text{見通しがたつ}/\text{否定}:+, X\text{ が数字を示す}/[\text{否定}:+, \text{可能}:+]\rangle.$	$=$	$\langle \mathit{precondition}, \text{見通しがたつ}/\varphi, X\text{ が数字を示す}/[\text{可能}:+]\rangle.$
変換規則 2	$\langle \mathit{cause}, \mathit{p1}/\text{否定}:+, \mathit{p2}/[\text{否定}:+, \text{可能}:+]\rangle.$	$=$	$\langle \mathit{means}, \mathit{p1}/*, \mathit{p2}/[\text{可能}:+]\rangle.$
	$\langle \mathit{cause}, X\text{ がチケットを買う}/\text{否定}:+, X\text{ が観戦する}/[\text{否定}:+, \text{可能}:+]\rangle.$	$=$	$\langle \mathit{means}, X\text{ がチケットを買う}/\varphi, X\text{ が観戦する}/[\text{可能}:+]\rangle.$
変換規則 3	$\langle \mathit{cause}, \mathit{p1}/\varphi, \mathit{p2}/\varphi\ \rangle.$	$=$	$\langle \mathit{effect}, \mathit{p1}/\varphi, \mathit{p2}/\varphi\ \rangle.$
	$\langle \mathit{cause}, \text{値段が下がる}/\varphi, \text{客足が伸びる}/\varphi\ \rangle.$	$=$	$\langle \mathit{effect}, \text{値段を下げる}/\varphi, \text{客足が伸びる}/\varphi\ \rangle.$

表 9　因果フレーム抽出規則

id	優先度	条件部				出力部			
		marker	*e_sub*	*e_main*	*inter_events*	*rel*	*e1*	*e2*	*mo*
1	1	ため	動詞意志性:＋，可能:＋	意志性:＋	生起順序:→	*precondition*	従	主	φ
2	1	ため	意志性:＋	動詞意志性:＋，可能:＋	生起順序:→，主体一致性:＋	*means*	従	主	φ
3	2	ため	意志性:＋	意志性:＋	生起順序:←，主体一致性:＋	*means*	主	従	φ
4	2	ため	意志性:＋	意志性:＋	生起順序:→，主体一致性:＋	*succession*	従	主	φ
5	2	ため	意志性:＋	意志性:＋	生起順序:←，主体一致性:−	*preparation*	主	従	φ
6	2	ため	意志性:＋	意志性:＋	生起順序:→，主体一致性:−	*enable*	従	主	φ
7	2	ため	意志性:＋	意志性:−	生起順序:←	*precondition*	主	従	φ
8	2	ため	意志性:−	意志性:＋	生起順序:→	*precondition*	従	主	φ
9	2	ため	意志性:＋	意志性:−	生起順序:→	*effect*	従	主	φ
10	2	ため	意志性:−	意志性:＋	生起順序:←	*effect*	主	従	φ
11	2	ため	意志性:−	意志性:−	生起順序:→	*cause*	従	主	φ
参考		のに	意志性:＋	意志性:−/否定:＋	生起順序:→	*effect*	従	主	*mo_delete(e2_mas:*否定*)*

文　献

[1] 阿部純一, 桃内佳雄, 金子康朗, 李光五: Cognitive Science and Information Processing(12) 人間の言語情報処理-言語理解の認知科学, サイエンス社 (1994).

[2] Allen, J.: *Natural Language Understanding*, The Benjamin/Cummings Publishing (1995).

[3] Bakeman, R. and Gottman, J.: *Observing Interaction: An introduction to Sequential Analysis*, Cambridge University Press (1986).

[4] Carberry, S.: *PLAN RECOGNITION IN NATURAL LANGUAGE DIALOGUE*, The MIT Press (1990).

[5] Craven, M., Dipasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K. and Slattery, S.: Learning to Extract Symbolic Knowledge from the World Wide Web, *Proc. of the 15th National Conference on Artificial Intelligence(AAAI-98)* (1998).

[6] Garcia, D.: COATIS,an NLP system to locate expressions of actions connected by causality links, *Proc. of the 10th European Knowledge Acquisition Workshop,*, pp. 347–352 (1997).

[7] Girju, R. and Moldovan, D.: Mining Answers for Causation Questions, *Proc. the AAAI Spring Symposium* (2002).

[8] Harabagiu, S. M. and Moldovan, D. I.: TextNet - A Text-Based Intelligent System, *Natural Language Engineering*, Vol. 3, pp. 171–190 (1997).

[9] Hobbs, J. R., Stickel, M., Appelt, D. and Martion, P.: Interpretation as Abduction, *Artificial Intelligence*, Vol. 63, pp. 69–142 (1993).

[10] 岩崎卓: 複文における時制, 月刊言語, Vol. 30, No. 13, pp. 50–55 (2001).

[11] Joskowicz, L., Ksiezyk, T. and Grishman, R.: Deep domain models for discourse analysis, *The Annual AI Systems in Government Conference* (1989).

[12] 黒橋禎夫: 日本語構文解析システム KNP version2.0 b6 使用説明書 (1998).

[13] 黒橋禎夫, 酒井康行: 辞書とコーパスからの世界知識の自動抽出, 「知識発見のための自然言語処理」シンポジウム (1999).

[14] Khoo, C. S. G., Chan, S. and Niu, Y.: Extracting Causal Knowledge from a Medical Database Using Graphical Patterns, *Proc. of the 2000 ACL*, pp. 336–343 (2000).

[15] Lenat, D.: Cyc: A large-scale investment in knowledge infrastructure, *Comunications of the ACM*, Vol. 38, No. 11 (1995).

[16] 松本裕治, 北内啓, 山下達雄, 平野善隆, 浅原 正幸松田寛: 日本語形態素解析システム『茶筌』version 2.2.1 使用説明書 (2000).

[17] 中岩浩巳, 池原悟: 語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析, 自然言語処理, Vol. 3, No. 4, pp. 49–65 (1996).

[18] 西澤信一郎, 木村啓一, 中川裕志: 用言の意味情報を利用した日本語複文の共参照関係の推定法, 情報処理学会論文誌, Vol. 38, No. 3, pp. 472–481 (1997).

[19] 中村ちどり: 日本語の時間表現, くろしお出版 (2001).

[20] 佐藤浩史, 笠原要, 松澤和光: 表層的因果知識ベースによる事象推移予測方式, 情報処理学会第 56 回 (平成 10 年前期) 全国大会, pp. 2–251–2–252 (1998).

[21] 佐藤浩史, 笠原要, 松澤和光: テキスト上の表層的因果知識の獲得とその応用, 信学技報 (TL98-23) (1999).

[22] Schank, R. and Abelson, R.: *Scripts Plans Goals and Understanding*, Lawrence Erlbaum Associates (1977).

[23] Stork, D. G.: Character and Document Research in the Open Mind Initiative, *Proc. of Int. Conf. oon Document Analysis and Recognition(ICDAR99)*, pp. 1–12 (1999).

[24] Takahashi, T., Iwakura, T., Iida, R., Fujita, A. and Inui, K.: KURA: A Transfer-Based Lexico-Structural Paraphrasing Engine, *The 6th Natural Language Processing Pacific Rim Symposium, Workshop on Automatic Paraphrasing: Theories and Applications* (2001).

[25] 情報処理振興事業協会技術センター: 計算機用日本語動詞辞書 IPAL, 日本情報処理振興協会 (1986).

[26] 日本経済新聞社: 日本経済新聞 CD-ROM 版 1990～1996.

[27] 日本電子化辞書研究所: EDR 電子化辞書仕様説明書 (1995).