

小説における論理構造理解のための因果・含意関係の抽出と分析

神山文子 西島恵介 藤田米春

大分大学 工学部 知能情報システム工学科
870-11 大分市旦野原 700 番地

我々は、より柔軟な自然言語処理を実現するために情緒的要素を含む文書である小説や物語などの計算機による自動理解の研究を進めている。人が情緒的要素を含む文書を理解する場合には明示されている言語情報から作成される論理的構造とそれを補完する様々な命題、常識、登場人物の心理などが読者の知識や感情に整合的に体系づけられなければならない。現在、我々は、短編推理小説を例にとり記述されている文章から小説の論理的構造の自動構築を目指して研究を進めている。本報告では、小説の記述から論理構造を構築するための方法として文中から因果含意を表す語を手掛かりとして、因果含意関係にある文の自動抽出およびその抽出結果の分析について報告する。

Extraction and analysis of the causality / implication relation for understandings of a logical structure in a short detective story

Fumiko KAMIYAMA, Keisuke NISIJIMA and Yoneharu FUJITA

Department of Computer Science and Intelligent Systems,
Faculty of Engineering, Oita University
DannoHaru 700, Oita-shi, Oita 870-11, Japan

The aim of our study is an automatic understanding of documents such as novels and stories which include emotions, because understandings of such documents are indispensable for flexible natural language understandings. In order to understand such a document, a logical structure obtained from the original text has to be supplemented by various propositions representing commonsense and mental states of a characters in the text, and consistently systematized into the logical structure of a reader's knowledge. Currently, we examined an automatic extraction of the cause and effect relations from an original text for construction of its logical structure, utilizing clue words representing those relations. We also describe some analyses of the extraction results.

1. はじめに

現在、自然言語文書処理は、教科書、マニュアル、新聞記事などを対象とした自動要約や抄録、ハイパーテキスト化によるユーザ支援などの方向で研究が進められていることが多い。これらの文章は、読み手に客観的な情報を伝達する目的を持っているため、事実を対象とすることが多く、文章も比較的簡潔かつ論理的に書かれている。このような点からこれらの文書は計算機処理に向いていると考えられる。一方、物語や小説などの文学書と呼ばれるものは、登場人物の感情の描写や会話文の挿入など多くの情緒的要素を含み、また修飾表現も多く事実ではない事の記述があったりと様々な情報提示や表現形態があり、さらにそれらの記述も、複数の文によって表現される事が多いため、これまでは計算機処理の対象としては除外されてきた。しかし、より人間に近い自然言語処理を実現するためにはこれらの文書が処理できるような枠組は必要であり、特に文中に現れる情緒的要素の処理を含んだ処理の枠組はマンマシンインタフェースへの応用など様々な利用が期待できる。

近年、脳科学や認知科学の発展により、人間の情緒的要素を物理化学的な現象をしてとらえ計算機上に乗せるいくつかのモデルが提案されている[1]。我々も、感情は広い意味で論理的にとらえ得るとの立場から感情のモデル化を進めてきた[2]。そこで、我々は短編推理小説を例にとり、感情も含めた論理に観点をおきながら計算機による小説の自動理解の研究を進めている。推理小説は登場人物などの情緒的要素を含みながら、ある程度論理的に物語が展開していくため我々の研究に適していると考えられるからである。計算機によって小説などの文書を理解させる場合には、明示されている言語情報から作成される論理的構造を、読者が持っている

常識、登場人物の心理等の様々な命題によって補充し、統合的な論理体系を構築する必要がある、と我々は考えている。そこで現在、ある短編推理小説に記述されている文章からその小説の論理的構造の自動構築を目指して研究を進めている。言語情報から作成される論理的構造は、文の因果・含意関係によって表現されることが多く、これらを文章間で連鎖することで、文章全体の論理体系の構築がある程度可能になると考えられる。本報告では、文中から因果・含意を表す語を手掛かりとした、因果・含意関係（以下、このような関係を「因果的關係」と呼ぶ）にある文の自動抽出と、その抽出された文の分析結果について述べ、小説の論理的構造の自動構築するための検討を行う。

2. 文章中の因果・含意表現

因果関係とはある前件が原因となって後件の結果を引き起こすことであり、筋道の通った論理的な文章を書く上で非常に重要である。技術文章等では、この因果関係が明確に記述されているので、文中の論理的構造をとらえるのは比較的容易である。文学書の中でも推理小説などは、文章中に様々な情報の記述があるが、推理の骨格では、ある事象が原因となって次の事象を引き起こし、またその事象が次の事象を引き起こす原因・結果が比較的明確な関係を繰り返しながら結末に至るという特徴をもっている。そこでこの原因・結果の連鎖を抽出することにより論理構造を明らかに出来るものとする。

従来、複文において因果関係を表す語として、接続助詞によるものがある。この接続助詞を使用し、従属節と主節の因果関係を成立させている。このことに着目し、まず、文献[3](871文)より因果的關係にある文を接続助詞に着目しながら意味的な見地から探し出し、その際手掛かりとなる語を見つけてい

く。次に文中から、その手掛かり語をもとに計算機による因果的關係にある文の抽出を行なっていく。ここで、因果的關係とは、原因とそれによって生じる結果（事実）との關係だけではなく、ある前提から推論によって導かれる結論、仮説から導かれる結論等も含む。

2.1 因果的關係を表わす手掛かり語

推理小説の文中から、複文において従属節と主節が因果的關係を示すものを抽出する。この時、接続助詞等に注意し、文法、品詞にとらわれることなく意味的な見地より、人の手で因果關係にある文の抽出を行った。抽出した文は、211文になり、この中から従属節と主節を接続する語に着目し、分類を行った。これらの語を「手掛かり語」と呼び、それぞれの語の役割、抽出文数を示す。これら手掛かり語は、大部分が接続助詞によるものだが、『たら』は、用言の連用形に接続する過去完了の助動詞である。

因果的關係を示す文の中でも従属節の述語の意味や、従属節や主節の動作主が主人公なのか、それとも主人公が観察者となって推測した事なのかにより因果關係が強くなり易いものと分かり難いものがある。例えば、従属節の述語の意味が動作を表すもので、その動作主が主人公自身の場合は、因果關係が明確である。

また従属節、主節の主語が省略されている場合が多い。主語がなくても因果的關係は成立しているが、前後の文を参照し主語を補足していくことにより、他の文との因果關係も明確になると思われる。

以下、例文は全て文献[3]からの引用である。

【手掛かり語の種類】

<主な意味> <手掛かり語>
原因・理由 … 『から(51文)』

『ので(8文)』 『で(8文)』
確定、仮定の順節… 『ば(18文)』

『と(70文)』

確定、仮定の逆接 … 『のに(3文)』

『けれど(3文)』 『が(44文)』 『ても(1文)』

『でも(1文)』

過去完了、仮定形 … 『たら(14文)』

【手掛かり語を持つ文の例】

- ・どうしても穴埋めをしなければならぬというので、私を狙ったんです。
- ・若い女の声で電話すれば、何とか会ってくれるだろう。
- ・他家に養子に行き性は変わっていますが、血のつながった兄です。
- ・私に対してそのような評価をしてくれたら、私は彼を作品持参で訪ねてみようと思っていた。

2.2 文と文を因果的關係で結ぶ手掛かり語

推理小説の文中では、因果的關係を表すものとして2.1で抽出した複文だけでなく文と文にまたがって因果的關係を表すものがある。接続詞によるものと助動詞によるもので、特に助動詞を使用して文と文の因果的關係を表すものは、前文で先に結果を述べ、後文でその原因を述べるというような倒置の形式をとっているものがほとんどであった。

2.2.1 接続詞によるもの

接続詞を使用することによって文と文の因果的關係を表すものの抽出を行う。接続詞の付属している後文は、結果として認めることが容易であるが、その原因となる前文は、接続詞の位置する前の文だけとは限らず、複数の文がその原因を述べていることがある。

その複数の文は、接続する語を用いることなく羅列されているが、読者は、推論によって原因と結果の関係を明確にしている。

推理小説の文中では、12種類の接続詞が因果的関係を成立させるものとしてあった。

<主な意味> <手掛かり語>

順接 …『だから(9文)』『それで(8文)』
『すると(1文)』『したがって(1文)』
『そうしたら(1文)』『そこで(7文)』
『そして(20文)』
逆接 …『だが(7文)』『しかし(32文)』
『ところが(6文)』『でも(11文)』
『それでも(1文)』

【手掛かり語を持つ文の例】

・最初は、もちろん入選を夢みた。しかし、編集部からの連絡がなく、駄目らしいとなったとき私は、佳作になることを考えた。

2.2.2 接続助詞によるもの

推理小説の文中には、文と文の因果的関係を表す際の手掛かり語として、接続詞の他に接続助詞を用いたものがある。これは、前文で先に結果を述べ、後文で原因を述べるという倒置法を用いた小説特有の表現によるものである。このような表現は、主に手掛かり語「から」を使用しているものが大部分である。

[例]私はまた赤くなった。美人教師といわれたからである。

2.3 手掛かり語を持たないもの

推理小説の文中から手掛かり語を使用して因果的関係を表すものを分類し、抽出を行ってきたが、因果的関係を表す際に、手掛かり語を使用しないで表現するものが文中に

は多数存在する。

[例]私は、階段をかけたのぼった。……Ⅰ
半ミニのスカートがずり上がったが、気にしないことにした。……Ⅱ

上の例文にあるように、Ⅰ、Ⅱを因果関係で結びつける手掛かり語は何も存在せず、文を羅列しているだけだが、Ⅰの文が原因となってⅡの結果が起こったことは明白である。このような因果的関係にある文を抽出するのは、表層の言語情報だけでは、困難である。他に読み手の知識や常識を備えたユーザモデルの存在が必要で、その知識や常識を使用して推論する機構により上記の例のような文も因果的関係としてとらえることで小説文章全体の論理体系の構築が可能になると思われる。また、逆に上記の例文のような文が因果的関係にあると仮定した場合にそれをつなぐ知識を推定することにより著者が読者に期待している常識が抽出されると考えられる。

3. 計算機による因果的関係の抽出

以上の分類結果より、複文における手掛かり語を使用して計算機による因果的関係にある文の抽出を試みた。まず、手掛かり語を探すために、文章の形態素解析を行う。これは奈良先端大の開発した形態素解析システム JUMAN[文獻[5]]を利用した。このJUMANの解析アルゴリズムは、全ての可能な形態素を辞書引きによって得、その個々の形態素について、その直前に位置する全ての形態素との接続可能性をチェックし、コストの低いものが優先される。また、利用者が文法(品詞名、接続関係など)を自由に定義できるという特徴をもっている。形態素解析の結果を、Perl言語で書かれたツールを使用して因果的関係にある文の抽出を行う。

[表 1] 計算機による因果的關係にある文の抽出実験

手掛かり語	形態素解析の結果	人の手による抽出文	抽出文	適合率	人の手による抽出文	修正後抽出文	修正後適合率
ば	動詞(基本条件形) 助動詞(ダ列基本条件形) 動詞性接尾辞(基本条件)	18文	31文	58.1%	18文	24文	75%
たら	動詞(タ系条件形) 動詞性接尾辞(タ系条件形) 判定詞(ダ列タ系条件形)	14文	20文	70%	15文	18文	83.3%
と	述語接続助詞	70文	93文	75%	70文	87文	80.4%
が	述語接続助詞	44文	53文	83%	49文	52文	94%
ので	助動詞(ナ形容詞, ダ列タ形連用)	8文	8文	100%	8文	8文	100%
から	述語接続助詞	51文	56文	91%	51文	54文	94%
のに	助動詞(ダ列基本連用)	3文	4文	75%	4文	4文	100%
けれど	述語接続助詞	3文	12文	25%	6文	9文	66.6%

【実験結果】

- 文献(かわいそうなパパ)の総文数 : 871文
- 因果的關係を表わす文 : 211文
- 計算機による抽出した文 : 277文
- 誤抽出文 : 66文
- 適合率 : 76%

それぞれの手掛かり語の抽出結果を表 1 に示す。表 1 より計算機で抽出を行った文は、人が抽出した文より多く抽出してきている。その抽出文には、人が抽出した文は全て含んでいるが、特に「ば」「たら」「と」「けれど」の手掛かり語を含むものに関しては、余分な文もかなり抽出してきている。その原因として大きく分けて次の 4 つの要因があった。

①今回は、複文における因果的關係にある文を抽出しているため、例 1 にあるような後の句を省略し、結論を読者の知識で補うものは因果的關係にある文として扱ってなかった

が「のに」の手掛かり語があるためこのような文も抽出してきてしまう。

[例 1] 腹をたてなければいけない相手なのに…

②単文の中に手掛かり語が存在し、結論が前の文に書かれているもの

[例 2] 間違いないと思ったのだ。あまりにも、英子に似ていたからである。

③文中に手掛かり語を含んでいるのだが、語の用法が因果關係の機能をしないもの

[例 3] 教育問題だったら、お間違いですよ。

④形態素解析の結果が間違っているもの

[例 4] 非常にスマートだと思ふ。

例にあるような「と」は用言に接続されているため本来は格助詞なのに接続助詞と解析されてしまう。

①～④の要因を解決するために以下の規則を用いて修正を行った。

I) ①②も因果的關係の文として取り扱う。

①は結論を補足すれば因果的關係にある文と認めることができる。しかし、どの

様な方法を用いて結論を補足するかは、検討中である。

Ⅱ) 会話文の中で手掛かり語の後に終助詞がつき、結論が省略されている場合は、明示的な結論がないので含めない。

【例5】単なる憶測だからね。

会話文における終助詞の影響は大きく、例にある「ね」の用法として確認、同意要求、自己確認の3つがあり結論を補ってよいかどうかの検討が必要である。

Ⅲ) この推理小説の中では、手掛かり語の後に動詞、形容詞がある場合は、仮定の順接を表現しているのだが、因果的関係を認める事ができないので含めない。

【例6】犯罪ルポでも書けばいいのだ。

4. 検討

I)～Ⅲ)を規則として用い計算機による因果的関係の文の抽出を行った。

【修正後実験結果】

-文献(かわいそうなパパ)の総文数	: 871文
-因果的関係を表わす文	: 221文
-計算機による抽出した文	: 256文
-誤抽出文	: 35文
-適合率	: 86%

以上の抽出結果より全体的に適合率は14.5%上がった。余分な文も抽出してくる問題として、特に『ば』『と』で多数みられ、手掛かり語の用法自体が因果的関係を表さない③の要因が大きく占めこれは、表層の情報だけで区別することが困難である。

5. まとめ

今回、因果的関係を表す文の手掛かり語の分類と手掛かり語を基にした計算機による抽出実験を行った。

複文で因果的関係にある文は、手掛かり語を使用することで、計算機による抽出でも高い精度で抽出でき論理構造構築には有効であると考えられる。またある手掛かり語により文と文の間に因果的関係が成立するものも、原因となる文を探す際、読み手が持っている常識のような知識を利用して推論する機構が必要である。小説文章の大部分をしめる、手掛かり語によらずに因果関係を表すものは、表層の言語情報だけでは因果的関係を見つけることが困難である。読者は自分の知識や常識に基づき推論することにより原因と結果を結びつけることができる。今後、より人間に近い自然言語処理をおこなうために、小説等の文章処理では人間の知識や常識をもち感情や意図を理解できるユーザモデルの存在は必須である。

参考文献

- [1] 戸田正直:認知科学選書24「感情」人を動かしている適応プログラム, 東京大学出版会, (1992)
- [2] 藤田, 西島:問題解決過程における感情の発生と解消の論理的メカニズムの提案, 認知科学, Vol.1, No2, pp59-63, (1994)
- [3] 佐野洋:かわいそうなパパ, 集英社文庫24「かわい目撃者」より, 集英社, (1979)
- [4] 藤田他:短編推理小説の論理構造の抽出. 日本認知科学会第14回全国大会, June(1997)
- [5] 神山他:小説における論理構造理解のための因果. 含意関係の抽出と分析, 情報処理第55回全国大会論文集(1997)
- [6] 神山他:推理小説における暗黙情報獲得について. 平成8年電関九支連大, (1996)
- [7] 西島他:短編推理小説における明示情報の論理的関係の抽出人文コン研究会, Nov(1996)
- [8] 松本裕次他:日本語形態素解析システム ChaSen Ver1.0 使用説明書, 奈良先端大学院