

# 計算機によるトートロジーの意味理解 ～検出機構の検討～

滝澤 修

taki@crl.go.jp

郵政省通信総合研究所 関西先端研究センター  
(651-24 神戸市西区岩岡町岩岡588-2)

神戸っ子は  
くじけません

## 1. はじめに

言外の意味を含む修辞を計算機に理解させる手法を確立することは、自然言語処理における重要な課題である。そのような修辞の一つに、「美しいものは美しい」「ゴミはゴミだ」などのトートロジー(tautology 同語反復, 同義循環)がある。トートロジーは、自然な表現として日常の自然言語に多く出現する一方、文法に沿わない省略や暗黙の指示関係などを伴う場合が多いため、通常の自然言語処理では手に負えない場合が多い(注1)。そのため、自然言語理解研究の一つとして、トートロジーを検出(自動抽出)および理解する手法を検討する必要がある。しかしトートロジーは従来、言語学、心理学の分野における興味の対象であったにとどまり、自然言語処理の観点からあまり注目されていなかった。

本稿では、計算機によるトートロジー理解へのアプローチとして、検出機構の実現のために必要な事柄を検討する。

## 2. トートロジー検出の研究

### 2.1 検出と理解とを分けることの妥当性

トートロジー以外の代表的な修辞として、比喩(隠喩)がある。比喩的な表現とリテラルな表現とは本来連続的なものであり、理解して初めて比喩であることがわかる(検出される)。従って比喩の場合、検出と理解とを切り離すことは困難である。それに対してトートロジーは、同じ語の反復という明確な表層構造をしていることから、意味理解を行わずに検出だけを行うことが比較的容易と考えられる。従って検出と理解とをある程度切り分けて扱うことができる。

トートロジーを計算機によって意味理解するためには、トートロジーをまず検出し、次にその検出部分だけを独立した意味理解機構において別処理する、という処理の流れが考えられる。つまりトートロジー検出機構は、高度な自然言語理解機構における前処理に位置づけられる。

### 2.2 トートロジーの定義と、検出の研究方法について

トートロジーの厳密な定義は言語学の分野においても見あたらない。そのため「トートロジー」という呼称が指す言語表現の範囲は、言語研究者によってマチマチなのが現状である。例えば中村[1]は、「一切絶対ない」のような「不注意な重複表現」のこととしており、佐山ら[2]は、「ゴミはゴミだ」のような、中村がいう同義循環に相当する表現のこととしている。そこでまず、自然言語理解におけるトートロジーの定義を明確に与える必要がある。

筆者らがトートロジーを自然言語理解の対象にしている理由は、通常の意味理解手法では手に負えないような言語表現を減らすことにある。従って自然言語理解の立場からトートロジーを概念的に定義すると、「同語が反復される修辞のうち、通常の意味理解手法では手に負えないもの」のような(注2)。従って、トートロジーの定義を決めるためには、「通常の意味理解手法」を想定しておく必要があることになる。

以上に基づき、筆者らはトートロジー検出の研究の進め方として、表1に示した4つのステップを踏むことを考えた。すなわち、自然言語理解におけるトートロジーの定義を与えることを第3ステップまでに行い、それに基づいて第4ステップにおいてトートロジー検出機構を構築する(注3)。本章で、第1ステップの検出機構について検討した結果を述べる。

表1 トートロジー検出機構の研究の進め方

<p>【第1ステップ】 トートロジーである可能性のある言語表現を、極力漏らさず検出する機構を作る。</p> <p>【第2ステップ】 その機構を用いて、ある程度の規模のコーパスからトートロジー(である可能性のある言語表現)を自動抽出する。</p> <p>【第3ステップ】 その抽出結果を検討し、想定した意味理解手法を参考にして、トートロジーの範囲(定義)を決める。</p> <p>【第4ステップ】 その定義に従ったトートロジーのみを検出する機構を構築する。</p>
---

### 3. トートロジー検出機構の検討

#### 3.1 第1ステップにおけるトートロジーの条件

第1ステップにおけるトートロジーが満たす条件として、以下の3つを設定する。

##### (1) 反復語の存在

一文中に同じ語が近接して2つ出現していることを、トートロジーであるための最低条件とする(注4)。この語を「反復語」と呼ぶ。反復語を記号「R」で表すと、トートロジーは、「R~R」のように定式化できる。この「~」は、反復語にはさまれた語句を表す。この「R~R」をトートロジーの「出現形式」とする。

##### (2) 反復語の品詞

反復語は、名詞、形容詞、動詞など、それだけで意味をもつ語、すなわち自立語に限定される。助詞などの付属語は反復語にはならない。

##### (3) 反復語の基本形の一致

反復語は、例えば「送るだけ送る」において前出と後出の反復語が共に「送る」であるように、形態的に完全に一致する場合と、例えば「送るだけ送って下さい」において前出が「送る」で後出が「送って」であるように、基本形が一致しても活用変化などのため形態的には不一致の場合とがある。トートロジーは同義を循環させることによって含意するという意味上の効果を狙った言語表現であるので、活用変化などの意味的に無関係な要因は、トートロジーであるか否かには無関係と考えられる。そこで、反復語は基本形が一致することを条件とする。

これらの条件を満たす言語表現を、第1ステップにおけるトートロジーとする。

#### 3.2 トートロジーの出現形式

3.1の条件に従うトートロジーの例をいくつか検討した結果、表2に示す出現形式が得られた。これらの出現形式の中には、反復語の間に副詞などの挿入が許容される場合がある(例「勉強する時はさすがに勉強するよ」、「法律はあくまでも法律だ」)。

表2 トートロジーの出現形式の例

出現形式	Rの品詞	文例
RはR	名詞など	ゴミはゴミだ 契約は契約なんだから それはそうだ それはそれとして おまじないはおまじないでも、よく効くの うちはうちで勝手にやる 大学は大学なりの問題を抱えている
RがR	名詞など	歳が歳だから
RもR	名詞など	歳も歳だから
Rにnot R	動詞など	寝るに寝られない
RにR	名詞、動詞など *1	走りに走る 乗りに乗る
RにはR	動詞など	あるにはあった
R(な)らR	名詞	駄目だった(な)ら駄目(だった)でいいや 駄目なら駄目でいいや
RというR	名詞など	建物という建物が倒壊した
RといえばR	名詞、動詞、形 容詞など	あるといえばある、ないといえはない *2 当たり前といえば当たり前だ
RといえるR	名詞など	成果といえる成果が無い
RらしいR	名詞など	成果らしい成果が無い
RことはR	動詞、形容詞な ど	あることはある、ないことはない *3 知らせることは知らせる *4
RものはR	形容詞など	やっぱり、美しいものは美しい
R時はR	動詞など	私だって、勉強する時は勉強するよ
R場合はR	動詞など	落ちる場合は落ちる
RだけR	動詞など	送るだけ送って下さい 聞くだけ聞いたら?

[備考]

\*1 反復語は前出が名詞、後出が動詞

\*2 反復語「ない」は、両方とも形容詞

\*3 反復語「ない」は、前出は形容詞、後出は助動詞 → \*2との違いに注意

\*4 この文の解釈は以下の通り少なくとも2つ考えられる

解釈① 知らせるべきことは知らせるようにしろよ

解釈①の例：「知らせるのを忘れていました」に対する返答  
「知らせることは知らせろよ」

解釈② 一応知らせるが、多分無駄だろう

解釈②の例：「彼にも知らせておいてくれ」に対する返答  
「知らせることは知らせるけどね」

### 3.3 構想中の検出機構（第1ステップ）の概要

検出機構には、語の同定、すなわち形態素解析がまず必要となる。次に、トートロジーが「R～R」の出現形式であることを利用し、予め定めておいた出現形式と入力文とを照合する。その際、出現形式毎に反復語の品詞に関する制約を設ける（例えば表2に例示したRの品詞に限定するとか）。また反復語の照合は、3.1の(3)で述べたように、活用形などでなく基本形で行う。

現在、出現形式の増強と制約条件等の分析、および照合方式の検討を進めている。なお、形態素解析はJUMAN[3]を用いている。

### 4. 考察

入力文と出現形式との単純な照合だけでは、明らかにトートロジーでない言語表現まで検出してしまう場合がある。例えば出現形式「RはR」との照合をとると、「彼は彼の妻に電話した」も検出されてしまうことになる。これを避けるためには、単語間の構文上の関係なども条件に加えた機構が必要と思われる。

### 5. おわりに

計算機によるトートロジー理解へのアプローチとして、検出機構の実現のために必要な事柄を検討した。そしてトートロジーの出現形式を調査した。今後、出現形式の増強と制約条件等の分析、および照合方式の検討を進めていく。

#### 【参考文献】

- [1] 中村明:「日本語レトリックの体系」, p.144, 岩波書店(1991).
- [2] 佐山公一, 阿部純一:「日本語同語反復文の意味解釈」, 心理學研究, Vol.65, No.1, pp.25-33 (1994).
- [3] 松本裕治他:「日本語形態素解析システムJUMAN使用説明書」, 京都大学工学部長尾研究室 (1993).
- [4] R.W.Gibbs, Jr. and N.S.McCarrell: "Why Boys Will Be Boys and Girls Will Be Girls: Understanding Colloquial Tautologies", J. Psycholinguistic Research, Vol.19, No.2, pp.125-145 (1990).

(注1) 機械翻訳などでは、名詞述語文のように構文構造の単純なトートロジーに対しては、構文レベルでの逐語訳によってある程度の対応が可能である。例えば「男の子は男の子だ」に "A boy is a boy" という訳を与えて処理を終了しても問題は無いように思える。但しこの逐語訳として "A boy is a boy", "Boys will be boys", "Boys are boys" などが考えられるが、これらはそれぞれ理解し易さや受け取られる意味が異なると指摘されている[4]。このような異言語間のトートロジー理解の違いに対応するためには、逐語訳では限界があるといえる。

(注2) この定義自体がトートロジーかも知れない<sup>^^</sup>;

(注3) 表1の第2ステップにおいて、トートロジーがコーパス中にどれ位の頻度で出現するかを知ることでもでき、トートロジーの計算機処理研究の意義を示すデータも得られることが期待できる。

(注4) 「男はオスだ」のような同義循環も意味的にはトートロジーに含めるべきであるが、現段階では対象外とする。