

日本語同語反復表現の検出手法について ——トートロジー理解タイプの自動判別——

滝澤 修 井佐原 均
郵政省通信総合研究所 関西支所

1. はじめに

「約束は約束だ」のようなトートロジー（同義循環）は、比喻やアイロニー等と同様に、言語学、心理学、認知科学的に興味ある言語表現であるが、計算機処理の対象としてはまだあまり注目されていない。

筆者らは、トートロジーを含む広い意味での日本語同語反復表現について、計算機による検出／理解手法の検討を進めている（図1）。本稿では、認知心理学の分野で指摘された「トートロジーの理解タイプ」について、それを自動判別するための手法について述べる。

2. トートロジーの「理解タイプ」について

トートロジーは表層的には無意味な言語表現であるが、それが情報伝達機能を持っている以上、理解のされ方はランダムなものではなく、聞き手に依存しない何らかの傾向がなければならない。佐山らは、認知実験（質問紙法）により、トートロジーの理解のされ方（理解タイプ）

には図2の4つがあると指摘している[5]。そこで筆者らは、この自動判別をトートロジー理解の第1歩と位置づけ、この4タイプの自動判別アルゴリズムの構築を試みる。

3. トートロジー理解タイプの自動判別手法

3.1 設計方針

理解タイプの自動判別について、筆者らは既に文献[1]において手法を提案し、オブジェクト指向言語を用いたプロトタイプを作成した。しかしその方法は ad hoc な知識に頼っており、また文脈を十分に考慮していなかった。そこで本稿で提案する方法は、以下の2つの点に留意して設計した。

- (1) 普遍的知識（既存の電子化辞書等）だけで対応でき、本判別のための独自の知識はなるべく必要としないこと。
- (2) 反復語に加えて先行文も分析対象とすること。

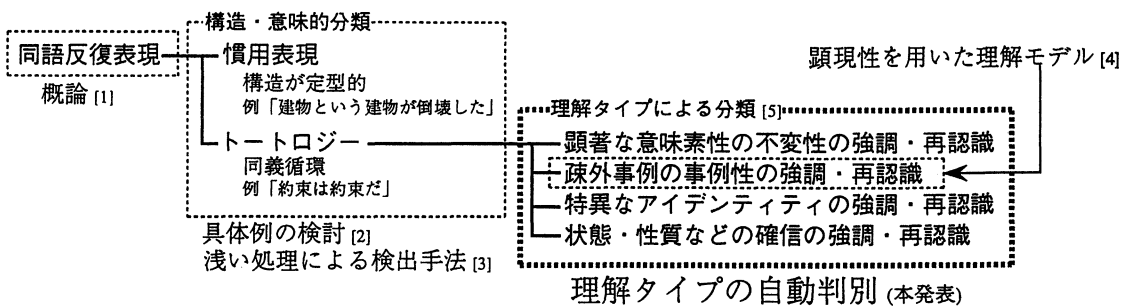


図1 本発表の位置づけ

【タイプ1】

顕著な特徴の不変性を強調／再認識させる

例「(腐っても)鯛は鯛だ」

解釈例「鯛が有している『高価で美味』という顕著な特徴は、腐っても不変である」

【タイプ2】

例外的な性質をもつ概念でも仲間であることを強調／再認識させる

例「(ペンギンだって)鳥は鳥だ」

解釈例「ペンギンは、飛べないという点で、鳥としては例外的な性質を持っているが、それでも鳥の一種であることには変わりはない。」

【タイプ3】

特異なアイデンティティを強調／再認識させる

例「(彼は彼、)私は私だ」

解釈例「私は彼とは異なった性格の人間だから、彼がやったとしても私はしない。」

【タイプ4】

状態・性質などを確信していることを強調／再認識させる

例「楽は楽だ」

解釈例「この仕事は面倒とはいっても、従来に比べたら楽であることは間違いない。」

図2 トートロジーの「理解タイプ」(佐山)

3.2 判別システムにおける前提

提案する判別手法は、以下の2点を前提とする。

【前提1】1つのトートロジーは、どれか一つの理解タイプに属する。

【前提2】どの理解タイプに属するかは、反復語と先行文によって決まる。

前提1の妥当性には議論の余地があろう。人間による理解タイプの判定では、被験者間および被験者内に結果のばらつきがあって当然であり、曖昧な判定結果を出すシステムが、人間の振舞いをよりよく模擬していることになるといえる。しかし本判別システムは、人間の振舞いを模擬することが目的ではなく、機械翻訳などへの応用を想定した意味理解機構における一機能とすることに主眼を置いている。そのためには、前提1のように、理解タイプをdiscreteに決定する機構である必要がある。

また前提2について、反復語や先行文に表現されない情報(状況や社会的背景など)が理解タイプ決定に関与することは十分に考えられる。しかし本研究では、自然言語処理の立場からト

ートロジー理解を目指しているので、このような非言語的情報を現段階では扱わないことにする。

3.3 具体的方法

3.2節の前提は認知実験によって検証する必要があるが、とりあえずこれらの前提が成り立つと仮定する。各理解タイプの反復語および先行文の特徴を考慮すると、判別システムは、それぞれの理解タイプ毎に分けた以下の4つの判定をserialに行うアルゴリズムを用いることが妥当と考えられる。

- (a)反復語が価値評価を持つ概念
→タイプ1
- (b)先行文中に反復語の(疎外事例としての)下位概念が存在 →タイプ2
- (c)先行文中に反復語の対立概念(兄弟概念)が存在 →タイプ3
- (d)反復語が状態・性質を表す概念
→タイプ4

これらのうち、判定(a)(価値評価の判定)は実際上困難である。なぜなら、価値評価は各概

念が絶対的に有するものでなく、文脈に強く依存している場合が多いためである。例えば「鯛は鯛だ」をタイプ1と判定するために、「鯛」が高価で美味（正の価値評価を有する）であることを概念「鯛」だけを手がかりとして検出するのは無理がある（鯛の価値評価に関するad hocな知識が必要）。そこで本手法ではとりあえず、他のどのタイプとも判定されなかった残りをタイプ1と判定するアルゴリズムを用いる。

また効率的な処理のためには、反復語の処理（概念記述項目の検索）を行なった後に、先行文の処理を行なうのが順序として適当である。更に兄弟概念の検索は、共通の上位概念を持つかどうかを検索することにより実現するので、上位概念の検索のみで済む処理のほうを兄弟概念の検索よりも先行させるべきである。

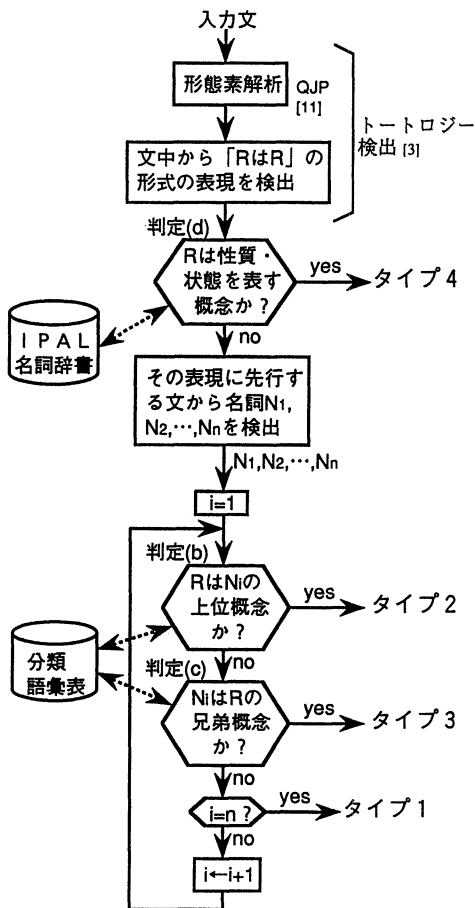


図3 トートロジー理解タイプ判別アルゴリズム

以上の考え方に基づいた判別アルゴリズムを図3に示す。図3の判定(b)と(c)は、分類語彙表[6]の階層関係を用いる。また判定(d)は現段階では、IPAL名詞辞書に登録されている概念のうち、「INC」（ヒトがある対象に対して持つ心理的傾向）の意味素性[7]を持つ概念とする。「INC」の意味素性を持つ概念の例としては、哀れ、安心、嫌、孤独、…、等がある。

なおこの判別アルゴリズムにおいて、先行文に存在する可能性のある「反復語の除外事例としての下位概念」（タイプ2）や「反復語の対立概念」（タイプ3）の検出は、文の構文/意味解析のような深い解析は行わずに、反復語の近傍に出現した名詞であることだけを手がかりとして行なっている。従ってこの先行文処理は、文脈解析とは言い難い。

4. 認知実験

自動判別システムの構築と並行して、トートロジー理解タイプ判別の認知実験を、質問紙法によって現在実施中である。この実験は、25個の名詞を用いて「RはRだ」のトートロジー表現を作って被験者に提示し（文脈は無し）、これがどの理解タイプに属するかを4者択一の強制選択法で選ばせるものである。実験では更に、タイプ1～3と判定されたものについて、以下の事項を記入させる。

タイプ1の場合：不変な顕著な特徴

（例）「鯛は鯛だ」の場合は「高価で美味」

タイプ2の場合：例外的な性質をもつ概念

（例）「鳥は鳥だ」の場合は「ペンギン」

タイプ3の場合：特異なアイデンティティに

関して対比される対象

（例）「私は私だ」の場合は「彼」

実験で用いた25個の刺激語は、佐山が用いた刺激語[8]、ユングの自由連想法の刺激語[9]、及びIPAL名詞辞書の見出し語のうち「INC」の意味素性を持つもの、の中から選択した。被験者数は約20名である。

この認知実験では文脈（先行文）は被験者には提示されない。従ってタイプ判別に関する被験者間のばらつきは、反復語だけで判別されやすいタイプ4については小さく、先行文が関与しやすい他のタイプでは大きいと予想される。また、タイプ1～3と判定したものについて被験者に書かせた事項については今後、実験データにおける先行文に用い、自動判別システムの

客観的な性能評価に利用する予定である。

5. 考察

図3のアルゴリズムに基づいた判別システムは、主にjperl言語によって記述されたプログラムでインプリメントし、簡単な例について動作を確認した。その結果、以下の問題点が明らかになった

●分類語彙表は完全な概念階層構造になっていないため、判定(b)がうまくいかない場合がある。例えば「人間」の上位概念は「動物」ではなく「人間活動の主体」になっているため、「人間だって動物は動物だ」のようなトートロジーの判別を誤ってしまう。また複数の上位概念（マルチインヘリタンス）を持つ場合への対処を検討しなければならないことも明らかになった。

●IPAL名詞辞書は基本名詞辞書であるため、現段階では見出し語数が少ない（806語）。そのため、判定(d)でヒットしてタイプ4と判定されることは実際には稀である。但しこのことは、普遍的な知識だけを手がかりとするという本手法の妥当性を否定するものではない。詳しい概念記述を持ち、かつ多くの見出し語を網羅している辞書が必要といえる。

6. まとめ

本稿では、トートロジーの理解タイプを自動判別する手法について述べた。またトートロジー理解タイプ判別に関する認知実験の概要について述べた。

今後は考察に基づき、概念階層構造の整備と、状態・性質に関する概念を抽出するより適切な方法などを検討する必要がある。またタイプ1について、絶対的な価値評価をもつ概念に限れば、IPAL形容詞辞書の意味情報の項目である「評価」と「快／不快」を利用する方法[1]や、アイロニー検出における「対人評価値」決定法[10]と同様の方法で判定できる可能性があるので、検討する予定である。

【謝辞】

本研究における形態素解析には、(株)リコーの簡易日本語解析系QJP[11]を用いている。使

用を許諾して下さった同社と、開発者の亀田雅之氏に感謝致します。また認知実験でご協力頂いた、流通科学大学対馬輝昭講師並びに同大学の関係者に感謝致します。

【参考文献】

- [1]滝澤:「トートロジーの工学的理解の試み」, 情処研報, 94-NL-104-14 (Nov.1994).
- [2]滝澤:「計算機によるトートロジーの意味理解-検出機構の検討-」, 言語処理学会第1回年次大会, pp.161-164 (Mar.1995).
- [3]滝澤,井佐原:「品詞の並びに関するヒューリスティックスを用いた日本語同語反復表現の検出」, 情処研報, 95-NL-110-3 (Nov.1995).
- [4]O.Takizawa,H.Isahara:"Machine-Understanding Model of Rhetorical Tautology Using Saliency", NLP'95, pp.77-82 (Dec.1995).
- [5]佐山,阿部:「修辞理解の認知過程-同語反復文の場合(隠喩文との比較)-」, 日本認知科学会第5回大会, D-3, pp.72-73 (1988).
- [6]国立国語研究所編:「分類語彙表(フロッキー版)」, 秀英出版(1994).
- [7]青山,橋本:「計算機用日本語基本名詞辞書IPAL(Basic Nouns)--解説編FTP版--意味素性一覧」(Dec.1995).
- [8]佐山,阿部:「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).
- [9]河合:「ユング心理学入門」, 培風館, p.66 (1967).
- [10]滝澤,伊藤:「アイロニー表現検出の一手法」人工知能学会誌, Vol.9, No.6, pp.875-881 (1994).
- [11]亀田:「軽量・高速な日本語解析ツール『簡易日本語解析系QJP』」, 言語処理学会第1回年次大会, pp.349-352 (Mar.1995).