

## 冗長表現の認識と言い換え

鍛冶 伸裕<sup>†</sup>

黒橋 禎夫<sup>†‡</sup>

<sup>†</sup> 東京大学大学院情報理工学系研究科 <sup>‡</sup> 科技団さきかけ研究 21

{kaji, kuro}@kc.t.u-tokyo.ac.jp

### 1 はじめに

冗長表現は、言い換え表現を生み出す原因の一つである。下の例のように、冗長な語を加える、または取り除くことによって言い換えを生成できる。

- (1) 返事を返す ↔ 返事する
- (2) 港に寄港する ↔ 寄港する

このような言い換えの判定を行うためには、冗長表現の認識が不可欠であると言える。言い換えの判定の重要性は多くの研究者によって指摘されており、多数の先行研究 [2, 3] が存在するが、冗長表現との関係について議論を行っている研究は存在しない。

本論文では、辞書定義文を利用して冗長表現の認識と言い換えを行う手法を提案する。例えば「貯金」の辞書定義文は「お金をためること」なので、「貯金をためる」という表現が冗長であると認識できる。そして、冗長な「ためる」を削除することによって「貯金する」と言い換えることができる。ただし本論文で扱うのは「体言+格助詞+用言」形の表現のみである。以下では、表現中の体言を**体言要素**、用言を用言要素と呼ぶ。そして、それらをまとめて**要素**と呼ぶ。

### 2 冗長表現

まず、どのような表現が冗長表現であるか、冗長表現にはどのような種類が存在するのかを整理する。

#### 2.1 冗長表現の定義

「体言+格助詞+用言」形の表現は、要素間の意味的な関係によって、次の三つに分類できる。

**独立型** 二つの要素の意味が独立している。(例) パソコンを買う

**包含型** 一方の要素の意味が、残りの要素の意味を包含している。前者の要素を**主要素**、後者を**被包含要素**と呼ぶ。(例) 貯金をためる、賞を受賞する

**付属型** 一方の要素は述語として働く(**主要素**と呼ぶ)。主要素は、それが体言であるか用言であるかに関わらず、用言的な述語として働く。残りの要素は、副詞や副詞的に働く連用節(まとめて副詞と呼ぶ)、またはヴォイスやアスペクトを表す付属語に相当する意味をもつ(**付属要素**と呼ぶ)。(例) 改革を断行する

冗長表現を「一方の要素を取り除いても、基本的には元の意味を保持できる表現」と考えると、次のように議論できる。まず独立型を考える。「頭をうなだれる」のように選択制限が強い表現は包含型と解釈すると、独立型は、一方の要素がなくなると元の意味が変化するので、冗長表現でない。次に包含型は、主要素だけでも元の意味を保持できるので、冗長表現である。最後に付属型は、一方の要素を取り除いてもある程度は元の意味を保持できるので、冗長表現である。(3)と(4)は、それぞれ独立型と付属型から用言要素を取り除いた例である。

- (3) 改革を妨害する → 改革する
- (4) 改革を断行する → 改革する

以上の議論より、包含型と付属型を冗長表現と定義する。以下では各々の冗長表現をさらに細かく分類する。

#### 2.2 包含型の分類

**動作包含型** 体言要素が主要素で、用言要素が被包含要素である包含型。主要素の意味は「ある動作の結果や状態など」であり、被包含要素の意味はその「ある動作」である。(5)では、貯金の意味は「ためた結果(=お金)」である。

- (5) 貯金をためる

**機能包含型** 動作包含型と同じく、体言要素が主要素である包含型。主要素は道具や人などの具体物であり、被包含要素はその機能(典型的な性質や使い方や役割など)である。(6)では、ズボンの典型的な使い方は「はく」ことである。

- (6) ズボンをはく

**対象包含型** 用言要素が主要素で、体言要素が被包含要素である包含型。主要素の意味は「ある物に何らかの動作を行う」ことであり、被包含要素は、動作の対象となる「ある物」である。(7)の受賞の意味は「賞をもらうこと」である。

(7) 賞を受賞する

ただし、一部の包含型では、主要素は被包含要素の意味の一部分しか包含していない。例えば、(8)では「即答」の意味の一部分だけが「解答」の意味に包含されている。同様に(9)では、「ノーベル賞」の意味の一部分だけが、「受賞」の意味に包含されている。これらをまとめて**非完全包含型**と呼ぶ。

(8) 解答を即答する

(9) ノーベル賞を受賞する

### 2.3 付属型の分類

**用言付属型** 体言要素が主要素で、用言要素が付属要素である付属型。主要素は用言的な述語として働くので、サ変名詞や動詞の派生語（「守り」など）に限られる。このような体言を**動作名詞**と呼ぶ。

(10) 改革を断行する

(11) 寄付を強制する

**体言付属型** 用言要素が主要素で、体言要素が付属要素であるような付属型。付属要素は副詞相当の意味しか持たない。

(11) 故意に遅れる

(12) 続けざまに勝利する

以上の分類の結果を、図1に示す。

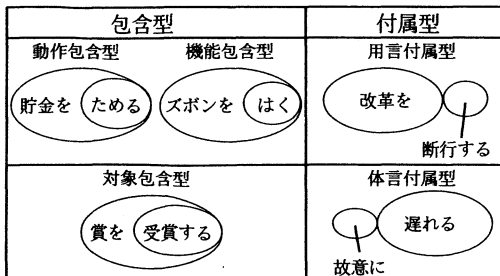


図1: 冗長表現の分類

## 3 辞書定義文による冗長表現の認識

これらの冗長表現を認識するためには、計算機上で要素の意味をどう表現するかを決める必要がある。本論文では「要素の辞書定義文が、その要素の意味表現

である」という考えに基づいた冗長表現の認識手法を提案する。このように考えると、冗長表現と辞書定義文の間には、次のような関係が成り立つことになる。

包含型は、(13)のように、主要素の辞書定義文が被包含要素を使って記述されている。そして付属型は、付属要素の辞書定義文が「物事をする」「行う」「やる」などの動作一般を表す表現（**動作一般表現**と呼ぶ）と、副詞、ヴォイスやアスペクトを表す付属語から構成されている。(14)は付属型の例である。

(13) 貯金をためる

貯金 お金をためる

(14) 寄付を強制する

強制 無理にやらせる

## 4 包含型の認識

包含型の認識を行うには、まず主要素の定義文を取り出し、次にその定義文が被包含要素を使って記述されているかを調べればよい。前者の処理では、要素の多義性が問題となる。また後者では、辞書定義文中で使われている表現と、被包含要素とのずれが問題となる。以下では、認識手法の詳細と、これら二つの問題の解決方法について順に述べる。

### 4.1 認識手法

**動作包含型** 体言要素の辞書定義文の主辞が用言要素と同じであれば、動作包含型と認識する。例えば(13)では、体言要素「貯金」の辞書定義文の主辞が用言要素「ためる」と等しい。

**機能包含型** 体言要素の辞書定義文の主辞を修飾する句が、用言要素と同じ用言を含んでいれば、機能包含型と認識する。例(15)では、体言要素「ズボン」の定義文の主辞は「もの」であり、ズボンの典型的な使い方「腰から下にはく」がそれを修飾している。

(15) ズボンをはく

ズボン 洋服で、腰から下にはくもの

**対象包含型** 用言要素の定義文の主辞の格要素が体言要素と同じであれば、対象包含型であると認識する。しかし主辞がとりうる格要素は「ある物」に限定されるので、その格要素が並列構造をとる、または接尾辞「など」を持っている場合は、対象包含型と認識しない。例(16)では、用言要素「受賞」の定義文の主辞は「もらう」で、その格要素は「賞」なので、「賞を受賞する」は対象包含型と認識できる。

(16) 賞を受賞する

受賞 賞をもらうこと

## 4.2 要素の多義性

多義語には各語義ごとに定義文が与えられており、どの定義文を使って認識を行うかが問題となる。体言要素の多義性を解消するためには、広い文脈を参照した処理が必要と考えられる。しかし本論文は「体言+格助詞+用言」形の表現のみを扱っているため、文脈情報はほとんど与えられておらず、体言要素の多義性解消は非常に困難である。そのため体言要素が多義語の場合は多義性解消の処理を行わない。一方、用言要素の場合は「体言+格助詞+用言」形の表現を与えられただけでも、ある程度の多義性解消が可能なので、[1]の手法を使って多義性解消を行う。

## 4.3 表現のずれ

主要素の辞書定義文には、必ずしも被包含要素と同一の語が使われているわけではなく、(17)や(18)のように、その上位語や同義語が使われている場合がある。

- (17) ノーベル賞を受賞する  
受賞 賞をもらうこと
- (18) 独り言を話す  
独り言 独りでものを言うこと

辞書定義文の主辞は、その見出し語の上位語や同義語に対応しているため、被包含要素の辞書定義文を使って表現のずれを吸収する。例えば「ノーベル賞」と「話す」の辞書定義文はそれぞれ「科学や文学などの偉大な業績に与えられる賞」「言うこと」となっている。

## 5 付属型の認識

付属要素の辞書定義文は「副詞」「動作一般表現」「(ヴォイスやアスペクトを表す)付属語」で構成されているため、要素の辞書定義文が「副詞\*動作一般表現付属語\*」というパターンで記述されていれば、その表現は付属型であると言える。\*や+などの記号は、通常の正規表現で使われるのと同じ意味で使う。

### 5.1 用言付属型の認識

問題は、「物事をする」「行う」といった動作一般表現の認識方法である。動作一般表現は次の二つで構成されている。

- 0個以上の、非常に抽象的な事象を表す体言(基本体言と呼ぶ)。例)物事、何か

- 1個の、用言付属型の付属要素(基本用言と呼ぶ)。例)する、行う

基本体言と基本用言の認識は、それらのリストを人手で用意することによって実現できる。辞書定義文に出現する表現は限られているので、この方法でも十分なカバレッジが期待できる。体言要素が動作名詞であり、なおかつ以下の条件のどちらかを満たす表現を、用言付属型と認識する。(19)は認識例である。

- 用言要素が基本用言の一つである。
- 用言要素の辞書定義文が「(副詞|基本体言)\*基本用言付属語\*」というパターンにマッチする。ただし多義性解消には4.2節の方法を使う。(副詞|基本体言)\*となっているのは、副詞と基本体言の語順は任意だからである。

- (19) 寄付を強制する  
強制 わりに さ せる こと  
副詞 基本用言 付属語

### 5.2 体言付属型の認識

用言付属型と比べて、体言付属型の付属要素やその辞書定義文には、二つの特徴がある。まず、付属要素の意味は副詞に相当するので、付属要素の辞書定義文は必ず副詞を含む。次に、付属要素は体言要素なので格助詞を持っていて、その格助詞に制約が存在する。これらを考慮して、前述の用言付属型の認識方法に以下の変更を加えた手法で体言付属型の認識を行う。(20)は認識例である。

- マッチさせるパターンは「副詞+動作一般表現付属語\*」を用いる。
  - 付属要素の格助詞は、任意的な意味を持つものに限定する。ここでは「で」「に」「と」を任意的な意味を表す格助詞と定義する。
- (20) 故意に遅れる  
故意 わざと する こと  
副詞 基本用言

## 6 言い換え手法

基本的には、包含型と付属型はそれぞれ「被包含要素を取り除く」「付属要素を副詞や付属語に変換する」ことによって言い換えることができる。いずれの場合も、主要素が体言要素であれば、それを用言化する必要がある。この手法で付属型は全て言い換え可能だが、以下の三つの包含型には特殊な扱いが必要となる。

**主要素が用言化不可能** 「表情をあらわす」のように、体言要素が主要素で、それが用言化ができない包含型は言い換えできない。

**主要素が多義語** 主要素が多義語のとき、その多義性を解消するために被包含要素が必要な場合がある。このような場合、被包含要素を取り除くと、人間でも主要素の多義性を解消できなくなるので、言い換えない。(21)の主要素「運動」は「体を動かす」「人々に働きかける」という二つの語義をもつが、被包含要素「働きかける」を取り除くと、「運動」の多義性を正しく解消できなくなる。

(21) 運動を働きかける → 運動する

**非完全包含型** 非完全包含型に前述の言い換え方法を適用すると、元の表現の意味が変わってしまうので、言い換えには特別な処理が必要となる。非完全包含型の扱いは二種類ある。まず(22)のように、主要素に包含されていない意味が副詞に相当すれば、被包含要素を取り除いたあとに、副詞を付け加えることによって言い換えることができる。一方(23)のように、それ以外の場合には言い換えることができない。

(22) 守りを固める → しっかりと守る

守り 守ること

固める しっかりと守ること

(23) ノーベル賞を受賞する

ノーベル賞 科学などのすぐれた業績に

与えられる賞

受賞 賞をもらう

## 7 実験と考察

冗長表現の認識実験を行い、適合率と再現率による評価を行った。使用した辞書は例解小学国語辞典 [4] である。テストセットは、新聞記事から 600 表現をランダムに取り出して作成した。テストセット中の全表現を筆者らが調べた結果、88 の冗長表現を含んでいた。テストセットの中から、提案手法が冗長表現と判断した表現は 79 個であった。14 個が包含型、65 個が付属型として認識されていた。79 個の表現のうち、冗長表現が正しく認識されていたのは 62 表現 (包含型 12 個、付属型 50 個) であった。適合率と再現率は、それぞれ、 $62/79=0.78$ 、 $62/88=0.70$  であった (表 1)。

認識例を以下に示す。(24)と(25)は、それぞれ用言付属型として認識できた。「浴びる」は二つの定義文を持っていたが、多義性は正しく解消された。(26)は動作包含型として認識することができた。

表 1: 適合率と再現率

	適合率	再現率
包含型	0.86(12/14)	0.52(12/23)
付属型	0.77(50/65)	0.77(50/65)
トータル	<b>0.78(62/79)</b>	<b>0.70(62/88)</b>

- (24) 確保に努める  
**努める** 一生懸命にする
- (25) 注目を浴びる  
**浴びる** (1) 水などをかける  
 (2) 受ける
- (26) 構想をまとめる  
**構想** 考えをまとめること

認識失敗の原因の一つは辞書定義文の不足であった。要素の多義性が強いとき、その意味を適切に記述した辞書定義文が存在しないことがある。これが原因で認識に失敗した例が見られた。例えば「競売にかける」は動作包含型と考えられるが、この「かける」の意味を適切に記述した辞書定義文がなく、認識に失敗した。

## 8 まとめ

言い換えと冗長表現の間には非常に密接な関係が存在するにも関わらず、言い換えと冗長表現の関係について議論を行った研究はほとんどない。これに対して本論文では、辞書定義文を利用して冗長表現の認識と言い換えを行う手法を提案した。そして実験の結果、適合率 0.78、再現率 0.70 という認識精度を得た。

## 参考文献

- [1] Nobuhiro Kaji, Daisuke Kawahara, Sadao Kurohashi, and Satoshi Sato. Verb paraphrase based on case frame alignment. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002.
- [2] Dekang Lin and Patrick Pantel. Discovery of inference rules for question answering. *Journal of Natural Language Engineering*, Vol. 7, No. 4, pp. 343–360, 2001.
- [3] Kentaro Torisawa. An unsupervised learning method for associative relationships between verb phrases. In *Proceedings of the 19th International Conference on Computational Linguistics*, pp. 1009–1015, 2002.
- [4] 田近洵一 (編). 例解小学国語辞典. 三省堂, 1997.