

迂言表現と重複表現の認識と言い換え

鍛治 伸裕[†] 黒橋 禎夫^{†,††}

言い換え処理は、様々な自然言語処理アプリケーションで必要とされている非常に重要な技術である。言い換え処理の一つとして、本論文では「名詞 + 格助詞 + 動詞」という形の迂言表現と重複表現を国語辞典を用いて認識し、さらにそれらを言い換える手法を提案する。迂言表現とは、動詞が動作を表していない表現や、名詞が動作の主体や対象を表わずに動作の状態を表している表現のことである。そして重複表現とは、動詞と名詞の間に意味の重複がある表現のことである。これらの表現には、多くの場合、同じ意味をより簡潔な形であらわした表現が存在する。提案手法の認識処理と言い換え処理の精度を二人の被験者が判断したところ、認識処理の精度は、平均して適合率 78%、再現率 52%であった。また、言い換え処理の精度は平均して 91%であった。

キーワード: 迂言表現, 重複表現, 言い換え, 国語辞典

Recognition and Paraphrasing of Periphrastic Phrases and Overlapping Phrases

NOBUHIRO KAJI[†] and SADAO KUROHASHI^{†,††}

Paraphrasing is a very important technique which is required by a lot of NLP applications. This paper represents a dictionary-based method of recognizing and paraphrasing *periphrastic phrases* and *overlapping phrases*. A periphrastic phrase is a phrase in which a verb does not represent an action, or a noun does not represent an agent/object of an action. An overlapping phrase is a phrase in which there is overlapping meaning between a noun and a verb. Most of these phrases have equivalent phrases which are expressed in a simpler form. The result of recognition and paraphrasing was assessed by two human judges. The result of recognition was evaluated through precision and recall, and the average precision and recall were 78% and 52% respectively. The average accuracy of paraphrasing was 91%.

KeyWords: *periphrastic phrase, overlapping phrase, paraphrasing, dictionary*

1 はじめに

近年、多くの研究者によって、言い換え処理の重要性が指摘されている。その一つの理由は、多くのアプリケーションでは、テキスト情報を利用目的やユーザのニーズに応じた形態に変換する必要があるが、それには言い換え処理が必須の技術だからである。例えば、テキスト情報を音声化してユーザに提供するというアプリケーションでは、書きことばを話しことばに変換

[†] 東京大学大学院情報理工学系研究科, Graduate School of Information Science and Technology, the University of Tokyo

^{††} 科学技術振興機構 さきがけ, PRESTO, JST

する処理が必要となる。なぜなら、音声化されるテキスト情報はいわゆる話しことばで書かれているほうが望ましいが、現在、大部分のテキスト情報は音声化に適さない書きことばで記述されているからである。他にも、子供などのユーザを対象としたアプリケーションでは、テキスト情報中の難解な表現は、あらかじめ平易な表現に言い換えておくべきである。

言い換え処理が重要視されているもう一つの理由は、言い換え処理は多くのアプリケーションの高精度化に必要であり、自然言語処理の基礎技術の一つになると考えられるからである。自然言語は、様々な言い換え表現を使って同一の意味内容を表すことができるという特徴を持っているが、これは、多くのアプリケーションの高精度化を困難にしている一因である。例えば質問応答では、ユーザの質問と文書の内容が同じ意味内容を表しているかどうかを判定しなくてはならない。言い換え処理は、このような問題の解決に利用できると考えられる。

本論文では次のような言い換えを扱う。ここでは「パソコンを買う」のような「名詞 + 格助詞 + 動詞」という形をした表現を考える。このような表現の場合、一般に、その動詞は動作(または状態など)を表し、名詞はその動作の主体や対象などを表している。そして、その二つの意味の組み合わせとして表現の意味が構成されている。これに対して、下の(1a)と(2a)に含まれる名詞や動詞は、このような性質を持っていない。

- (1) a. 改革を断行する
 - b. 思いきって改革する
- (2) a. 貯金をためる
 - b. 貯金する
 - c. お金をためる

(1a)では、名詞「改革」が動作を表していて動詞「断行」は副詞的に¹働いている。このような表現を迂言表現と呼ぶ。(2a)では、名詞「貯金」と動詞「ためる」の間に意味の重複があるので、名詞と動詞の意味の組み合わせとして表現の意味が構成されているわけではない。このような表現を重複表現と呼ぶ。このような迂言表現や重複表現には、同じ意味をより簡潔にあらわした言い換え表現が存在する。例えば(1a)は(1b)に、(2a)は(2b)や(2c)に言い換えることができる。(1b)は、動詞句「思いきって」が副詞表現で、動詞「改革する」が動作を表している一般的な表現である。また(2b)と(2c)は、(2a)から「貯金」と「ためる」の間の意味の重複を取り除いた表現である。

本論文では、議論の対象を「名詞 + 格助詞 + 動詞」という形の表現に限定して、辞書定義文を用いて迂言表現と重複表現を認識し、さらに、それらを(1b)、(2b)、(2c)のような一般

¹ 副詞のように、動作の程度や状態を表す表現を、副詞的に働く表現と呼ぶ。副詞的に働く表現の中でも、動作を表す動詞を連用修飾しているものを副詞表現と呼ぶ。例えば「思いきって改革する」の「思いきって」や、「きれいに掃除する」の「きれいに」は副詞表現である。

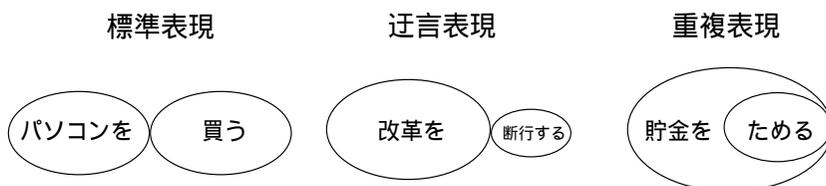


図 1 「名詞 + 格助詞 + 動詞」形の表現の分類

的な表現に言い換える手法を提案する。例えば (1a) では、「断行する」の辞書定義文に具体的な動作の記述がなく、副詞表現しか含まれていないことから、「断行する」が動作を表しているのではなく、副詞的に働いていることがわかる。そして、その辞書定義文中の副詞表現を利用して、(1b) のように言い換えることができる。また (2a) では、「貯金」の辞書定義文から「貯金」と「ためる」の意味の重複を見つけることができ、その結果、(2a) を (2b) や (2c) に言い換え可能である。

以下では、表現中の名詞を名詞要素、動詞を動詞要素と呼び、それらをまとめて要素と呼ぶ。

2 迂言表現と重複表現の分類

「名詞 + 格助詞 + 動詞」形の表現は、その名詞要素と動詞要素の関係にもとづき、標準表現、迂言表現、重複表現の三つに分類できる (図 1)。標準表現とは、「パソコンを買う」のような一般的な表現のことである。

「名詞 + 格助詞 + 動詞」形の表現を考えた場合、名詞要素と動詞要素の間には選択制限がある。名詞要素と動詞要素の間に選択制限が存在するということは、それらの意味にある種の冗長性があるということだが、選択制限は一般的に、(2a) などに比べると、意味の重複と呼べるほどのものではない。そのため、2.2節の重複表現の説明のところでも述べるような例外的な場合を除いて、選択制限は意味の重複とは考えない。

2.1 迂言表現の分類

次の二つの表現を迂言表現と呼ぶ²。

- 動詞要素が動作 (または状態など) を表していない表現
- 動詞要素が動作を表しているが、名詞要素が副詞的な働きをしている表現

前者の迂言表現を動詞付属型、後者を名詞付属型と呼ぶ。いずれの迂言表現でも、一方の要素が動作を表し、残りの要素が副詞的な働き、ヴォイスやアスペクトの働き、名詞要素を述語化

² 迂言表現や迂言的という言葉は、例えば (山梨 1995) などの他の文献では、本論文と異なる意味で使われている場合もある。

する働きなどを持っている。前者の要素を主要素と呼び、後者の要素を付属要素と呼ぶ。図 1 では、主要素を普通の大きさの円、付属要素を小さな円で表している。以下に示す迂言表現の例では、付属要素に下線をひく。

動詞付属型 動詞要素が動作を表していない迂言表現を動詞付属型と呼ぶ。動詞付属型では、動詞要素の代わりに名詞要素が動作を表しているので、名詞要素が主要素、動詞要素が付属要素となる。名詞要素 (=主要素) は動作を表すので、サ変名詞か動詞の派生語 (「守り」など) に限られる。このような名詞を動作名詞と呼ぶ。以下に動詞付属型の例を、その付属要素の働きとともに示す。

- (3) a. 改革を 断行する
副詞的
- b. 守りを 固める
副詞的
- c. 攻撃を 受ける
ヴォイス (受身)
- d. 寄付を 強制する
副詞的 + ヴォイス (使役)
- e. 研究を 行う
述語化
- f. 世話を やく
述語化

動詞付属型では、基本的に全ての付属要素が名詞要素を述語化する働きを持っている。しかし上では、名詞要素を述語化する働きの他に、副詞的な働きやヴォイス、アスペクトの働きも持っている付属要素の場合は、その付属要素が名詞要素を述語化する働きを持っていることを省略している。

名詞付属型 動詞要素が動作を表しているが、名詞要素が副詞的な働きをしている迂言表現を名詞付属型と呼ぶ。動詞要素が動作を表す主要素、名詞要素が付属要素となる。付属要素はすべて副詞的な働きのみを持っている。以下に名詞付属型の例を示す。

- (4) a. 故意に 遅れる
- b. 死に物狂いで 走る
- c. 粉微塵に 砕ける

d. 一夜漬けで 勉強する

「故意に」や「死に物狂いで」といった表現の品詞は「名詞 + 格助詞」ではなく「副詞」であるとも考えることもできる。しかし、そもそも、このような表現は、その品詞を「名詞 + 格助詞」か「副詞」のどちらに決めても間違いではない。さらに、このような表現を形態素解析システムが解析するとき、システムは、各単語の品詞情報を一意に定めた単語辞書を用いるので、このような表現の全てを副詞と解析することは難しい。例えば「故意に」を解析する場合なら、「故意に」が単語辞書に副詞として登録されていれば、それは「副詞」と解析されるが、登録されていなければ「名詞 + 格助詞」と解析される。このような理由で、システムが「名詞 + 格助詞」と解析した表現には、副詞的に働くものが混在している。そのため、上記では「故意に」や「続けざまに」といった表現を「副詞」でなく「名詞 + 格助詞」と考えて議論している。

以上では、名詞要素と動詞要素のどちらが主要素であるか（動作を表しているか）に注目して、迂言表現の分類を行った。これとは別の分類として、主要素と付属要素の結び付きの強さに注目して、迂言表現を次の二つに分けることもできる。一つめは、(3a) のような、付属要素が不特定の主要素と結び付くことができる迂言表現である。このような迂言表現を不特定付属型と呼ぶ。(3a) 以外にも、(3c) , (3d) , (3e) , (4a) , (4b) はいずれも不特定付属型である。二つめは、(3b) , (3f) , (4c) , (4d) のような、付属要素が、特定の主要素としか結び付くことができない迂言表現である。例えば「固める」は、「守りを固める」や「守備を固める」のように、「守り」の同義語と結び付いたときだけに付属要素として働く。同様に (4d) の「一夜漬けで」は「勉強する」のような、限られた動詞と結び付いたときのみ、副詞的な働きをもつ付属要素として働く。「一夜漬けで失敗した」のように、それ以外の動詞と結び付いた場合の「一夜漬けで」は、副詞的な働きの他に「勉強する」という意味を持つので付属要素として働かない。このような迂言表現を特定付属型と呼ぶ。

2.2 重複表現の分類

名詞要素と動詞要素の間に意味の重複がある表現を重複表現と呼ぶ。このとき「貯金をためる」の「貯金」のように、他方を包含している要素を主要素、「ためる」のように包含されている要素を被包含要素と呼ぶ。図 1 では、主要素を大きな円で、被包含要素をその円に含まれる小さな円で表している。重複表現は、対象包含型、動作包含型、機能包含型の三つに分類できる。以下に示す重複表現の例では、主要素に下線をひく。

対象包含型 動詞要素が「ある対象に何らかの動作を行うこと」を表していて、名詞要素が「その対象」を表している重複表現を対象包含型と呼ぶ。動詞要素が名詞要素を包含しているので、動詞要素が主要素で名詞要素が被包含要素となる。以下に対象包含型の例を示す。

(5) a. 賞を 受賞する

b. 馬から 落馬する

c. 馬が いなく

(5c) で、動詞要素「いなく」が名詞要素「馬」を包含しているかどうかは次のように考える。 (5c) の動詞要素「いなく」と名詞要素「馬」の間には、「パソコンを買う」といった表現に比べて、非常に強い選択制限が存在する。そのため、「パソコンを買う」のような表現とは違い、(5c) から名詞要素「馬」を取り除いて「いなく」とだけ言っても情報はほとんど欠落しない。したがって (5c) のように選択制限が強い表現は、動詞要素が名詞要素を包含していると考えられる。

動作包含型 名詞要素が「ある動作、またはその対象や結果」を表し、動詞要素が「その動作」を表している重複表現を動作包含型と呼ぶ。名詞要素が動詞要素を包含しているので、名詞要素が主要素で動詞要素が被包含要素である。

(6) a. 貯金をためる

b. 食事を食べる

c. 歌を歌う

(6a) の名詞要素「貯金」は動詞要素「ためる」を包含しており、(6b) の名詞要素「食事」は動詞要素「食べる」を包含している。(6c) の名詞要素「歌」と動詞要素「歌う」はほぼ同じ意味である。もし、動詞要素が名詞要素を包含していると解釈すれば、主要素は動詞要素であり、(6c) は対象包含型といえる。しかし本論文では、(6c) のように、名詞要素と動詞要素がほぼ同じ意味を持っている表現は動作包含型であるとしておく。

機能包含型 動詞要素が、名詞要素の持つ本質的な機能や役割や目的など(まとめて機能と呼ぶ)を表している重複表現を機能包含型と呼ぶ。この場合も、名詞要素が動詞要素を包含していると考えられることができるので、名詞要素が主要素で動詞要素が被包含要素である。

(7) a. はさみで切る

b. 先生が教える

c. FAXで送る

「はさみ」の機能は「切る」ことで、「先生」の役割は「教える」ことである。また「FAX」の機能は「送る」ことである。

重複表現には、名詞要素と動詞要素の意味の間に完全な包含関係が成立しない表現も含める。そのような重複表現を非完全包含型と呼ぶ。例えば下の (8a) と (8b) は、対象包含型の非完全包含型である。

迂言表現	重複表現
<p>動詞付属型</p>	<p>動作包含型 機能包含型</p>
<p>名詞付属型</p>	<p>対象包含型</p>

図 2 迂言表現と重複表現の分類

- (8) a. ノーベル賞を 受賞する
 b. ハイセイコーから 落馬する

これらは、上の (5a) と (5b) の被包含要素を、それに修飾的な意味が加わった要素に変えたものである。(8a) は (5a) の被包含要素を「ノーベル賞」に変えたもので、(8b) は (5b) の被包含要素を「ハイセイコー」に変えたものである。同様に、動作包含型と機能包含型の非完全包含型の例を以下に示す。

- (9) a. 貯金を ためこむ
 b. 食事を 食べつくす
 (10) a. はさみで 切り刻む
 b. FAX で 送りつける

以上の分類結果をまとめると、図 2 のようになる。

3 辞書定義文に基づく認識

認識手法の概要と、認識全般で必要となる多義性解消の手法について述べる。

3.1 認識手法の概要

これまでで説明してきたような迂言表現と重複表現を、辞書定義文を用いて認識する。迂言表現と重複表現の要素の辞書定義文には以下のような特徴があるので、それらを利用すること

ができる。

迂言表現 (不特定付属型) 付属要素は、副詞表現やヴォイスやアスペクトの働きを持っているので、付属要素の辞書定義文は、副詞表現、ヴォイスやアスペクトを表す付属語を含む。また、辞書定義文は具体的な動作の記述を含まない。すなわち辞書定義文は以下の三つから構成されている³。

- 副詞表現
- ヴォイスやアスペクトの働きをもつ付属語
- 「する」「物事を行う」などの動作一般を表す表現 (動作一般表現と呼ぶ)

不特定付属型は、付属要素の辞書定義文に対するパターンを用いて認識できる。

迂言表現 (特定付属型) 付属要素は特定の主要素としか結び付くことができない。付属要素の辞書定義文には、どのような主要素と結び付くことができるのか、という情報が記述されている。付属要素の辞書定義文は以下の三つから構成されている。

- 副詞表現
- ヴォイスやアスペクトの働きをもつ付属語
- 付属要素が結び付くことができる主要素 (または、その派生語)

つまり、付属要素の辞書定義文は主要素 (または、その派生語) を含んでいるので、主要素を付属要素の辞書定義文に対応付けることによって認識できる。

重複表現 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。包含型も特定付属型と同様に、被包含要素を主要素の辞書定義文に対応付けることによって認識できる。

これらをまとめると図3のようになる。不特定付属型の認識手法をパターンマッチング方式と呼ぶ。そして、特定付属型と重複表現の認識手法を要素-辞書定義文マッチング方式と呼ぶ。

3.2 要素の多義性解消

パターンマッチング方式と要素-辞書定義文マッチング方式のいずれの認識手法においても、まず問題となるのが多義語の扱いである。多義語は、各語義ごとに定義文が与えられているので、どの定義文を使って認識を行うかを決める必要がある。

名詞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である。しかし本論文では「名詞 + 格助詞 + 動詞」形の表現のみを扱っているので、文脈情報はほとんど与えられておらず、名詞要素の多義性解消は非常に困難である。一方、動詞要素の多義性を解消するためには、名詞要素の多義性解消の場合ほど広い文脈は必要ではない。そのため「名詞 + 格助詞 + 動詞」形の表現を与えられただけでも、かなりの程度の多義性解消が可能である。以上の理由により、名詞要素が多義語の場合は多義性解消の処理を行わず、どれか一つでも辞書定義文が認識条件にあてはまれば、その表現を迂言表現/重複表現であると判定する。

³ 辞書定義文は、文末表現が「～すること」という記述になっている場合があるが、以下では文末の「こと」を無視する。

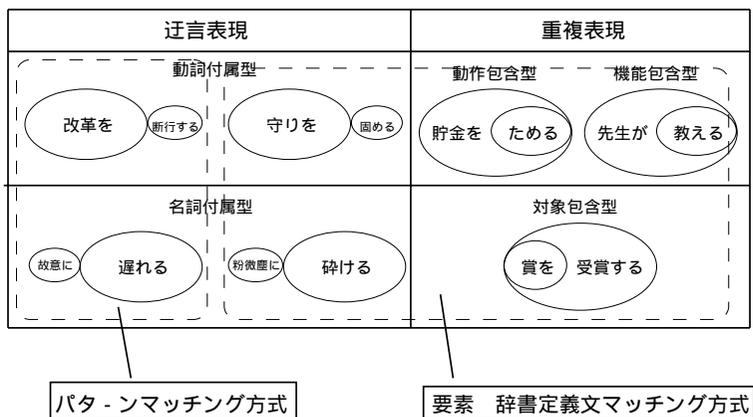


図 3 迂言表現と重複表現の認識手法

動詞要素が多義語の場合は, Kaji らの提案した手法を用いて, 多義性解消の処理を行う (Kaji, Kawahara, Kurohashi, and Sato 2002). この手法は, 教師無し学習によって動詞要素の多義性解消を行うものである. 教師無し学習によって多義性解消を行う場合, 多義性解消の手がかりとして使えるのは辞書定義文だけであるが, 辞書定義文は動詞要素の多義性を解消するために十分な情報を必ずしも含んでいない. そこで Kaji らは, 大規模コーパスから自動構築された動詞の格フレーム辞書 (Kawahara and Kurohashi 2001) を利用して, 辞書定義文に不足している情報を補う手法を提案している. この手法の概要を以下に示す (図 4). なお, Kaji らは, 日本語語彙大系 (NTT コミュニケーション科学研究所 1997) を用いて単語の類似度を定義し, その単語の類似度に基づいて, 格フレームと入力表現の類似度, 格フレームと定義文の類似度, 格フレームと格フレームの類似度を定義しているが, 以下ではそれらの計算方法の詳細な説明は省略している.

1. 入力格フレームの選択 格フレーム辞書から, 入力表現との類似度が最も高くなる格フレームを選択する. これを入力格フレームと呼ぶ. 図 4 の例では, 入力表現「教えをあおぐ」に対して, 下の格フレームが格フレーム辞書から選択される.

{ 本部, 幹部... } ガ { 外部, 専門家... } ニ { 教え } ヲ あおぐ

2. 定義文格フレームの選択 動詞要素の辞書定義文の主辞となる動詞 (この例では「うやまう」と「もとめる」) の格フレームの中から, 定義文との類似度が閾値を越えるような格フレームだけを選択する. 選択された格フレームを定義文格フレームと呼ぶ.
3. 格フレームの対応付け 入力格フレームと最も類似度が高い定義文格フレームを選択することによって, 多義性解消を行う. この例では, 下の定義文格フレームが選択される.

{ 当局, 学生... } ガ { 会社, 企業... } ニ { 教え, 指示... } ヲ もとめる

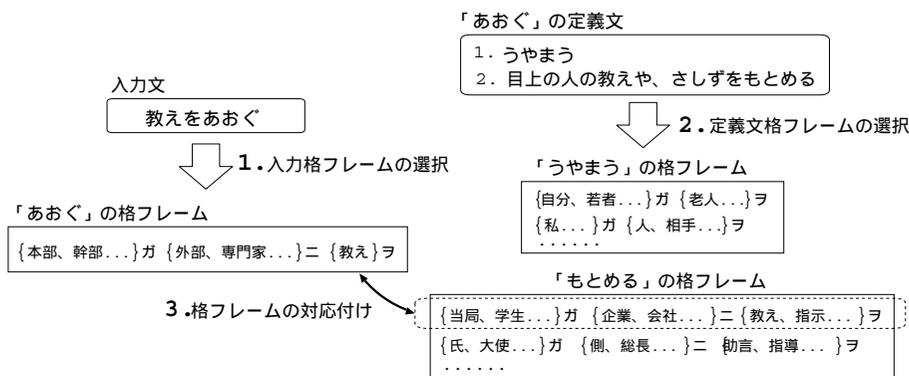


図 4 多義性の解消

したがって、入力表現の「あおぐ」は、二番目の辞書定義文「もとめる」に対応する意味を持っているとわかる。

4 パターンマッチング方式

不特定付属型の付属要素の辞書定義文は、あるパターンで記述されているので、それを手がかりにして不特定付属型を認識する。

4.1 動作一般表現の認識

不特定付属型を認識するためには、まず辞書定義文中の動作一般表現を認識することが必要となる。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名詞 (基本名詞と呼ぶ)。
例) 物事, 何か
- 1 個の、非常に抽象的な動作一般を表す動詞 (基本動詞と呼ぶ)。
例) する, 行う

そこで、辞書定義文を網羅的に観察して基本名詞と基本動詞のリストを作成し、それを用いて動作一般表現を認識する。表 1 に作成したリストを示す。基本名詞は 4 種類、基本動詞は 17 種類がリストに登録されている。基本動詞は、それが持つヴォイスやアスペクトの働きに応じて「無し」「使役」「受身」「可能」「開始」「継続」「終結」の 7 種類に分類されている。「無し」というのは、基本動詞がヴォイスやアスペクトの働きをしないことを表す。

表 1 基本名詞と基本動詞のリスト

基本名詞	物事，こと，何か，状態	
基本動詞	無し	する，やる，行う
	使役	強いる，やらせる
	受身	受ける，行われる
	可能	できる
	開始	始める，始まる，起こす
	継続	進める，進む，続ける，続く
	終結	終わる，終える

4.2 不特定動詞付属型の認識

以上のリストと辞書定義文を用いて不特定動詞付属型の認識を行う．基本動詞が動作名詞と結び付いたときには，その基本動詞は不特定動詞付属型の付属要素として働く．したがって，以下の条件を満たしている入力表現を，不特定動詞付属型であると認識する．

- 名詞要素が動作名詞である．
- 以下の二つの条件うち，いずれかを満たす
 - － 動詞要素が基本動詞の一つである．
 - － 動詞要素の辞書定義文が下のパターンにマッチする⁴．

副詞表現 * 基本名詞 * 基本動詞 付属語 *

動詞要素が複数の辞書定義文を持っている場合は，3.2節の手法を用いて多義性解消を行う．

不特定動詞付属型の付属要素の辞書定義文であると判定された例を以下に示す．

(11) 寄付を強制する

強制 むりに さ せる こと
 副詞表現 基本動詞 付属語

(12) 改革を断行する

断行 思い切って やる
 副詞表現 基本動詞

(13) 戦いを繰り広げる

繰り広げる 次から次へと 続ける
 副詞表現 基本動詞

⁴ 以下では，* と + を，正規表現で使われるのと同じ意味で使う．

固める しっかりと守る

(17) 考えをめぐらす

めぐらす いろいろ考える

(16) では、動詞要素「固める」の辞書定義文の主辞は「守る」で、それは名詞要素「守り」を動詞化したものと同じである。同様に、(17) の「めぐらす」の辞書定義文の主辞は「考える」で、名詞要素「考え」を用言化したものと同じである。

名詞付属型 名詞要素 (= 付属要素) の辞書定義文の主辞が、動詞要素 (= 主要素) と同じであり、なおかつ、4.3節で述べたのと同様の以下の条件を満たしていれば、名詞付属型であると認識する。

- 名詞要素の辞書定義文は、副詞表現を必ず含んでいる。
- 名詞要素の辞書定義文の主辞が、ヴォイスやアスペクトを表す機能語を持っていない。
- 入力表現の格助詞が「に」「で」「と」のいずれかである。

以下に認識例を示す。

(18) 粉微塵に砕ける

粉微塵 ひじょうに細かくくだけること

(19) 小出しに出す

小出し 少しずつ出すこと

(19) の名詞要素「粉微塵」の定義文の主辞は「くだける」で、(20) の名詞要素「小出し」の定義文の主辞は「出す」である。

5.2 重複表現の認識

重複表現は次のように認識する。

対象包含型 動詞要素 (= 主要素) の辞書定義文の主辞が格要素を持っていて、それが名詞要素 (= 被包含要素) と同じであれば、対象包含型であると認識する。

(20) 賞を受賞する

受賞 賞 をもらうこと

(21) 馬がいなく

いなく 馬 が、声高く鳴く

動詞要素の辞書定義文中の格要素が、並列構造であったり、接尾辞「など」を持っている場合、いずれも対象包含型であると認識しない。なぜなら、そのような格要素は動詞要素の選択制限を表している、しかも、その選択制限は(21)のような強い選択制限ではないからである。以下に例を示す。

- (22) 遺体を安置する
 安置 仏像や遺体 を、たいせつに置いておくこと
- (23) 建物を取り壊す
 取り壊す 建物など をこわして、取りのぞく

動作包含型 名詞要素 (= 主要素) の辞書定義文の主辞が動詞要素 (= 被包含要素) と同じであれば、動作包含型と認識する。

- (24) 貯金をためる
 貯金 お金を ためる こと
- (25) 食事を食べる
 食事 物を 食べる こと

機能包含型 名詞要素 (= 主要素) の辞書定義文から、名詞要素の機能を表す動詞を抽出し、その動詞と動詞要素 (= 被包含要素) が同じであれば、入力表現を機能包含型と認識する。以下では、辞書定義文から機能を表す動詞を抽出する方法を説明する。

名詞要素の機能を表す動詞 (以下では、たんに機能と呼ぶ) は、次のような形で辞書定義文に記述されている。

- (26) 寝台 寝る ために使う台
 寝台の機能
- (27) はさみ 物をはさんで 切る 道具
 はさみの機能

(26) では、機能「寝る」の直後に「ために使う」という表現がある。この表現によって「寝る」が機能を表していることが明示化されている。この「ために使う」のような表現を手がかり表現と呼ぶ。辞書定義文が手がかり表現を含んでいる場合、その手がかり表現を利用して機能を抽出できる。これに対して(27)では、辞書定義文は手がかり表現を持っていない。手がかり表現を含まない辞書定義文から機能を抽出する難しさは、主辞にかかる動詞がどれでも機能を表しているとは限らない、ということである。例えば、以下の辞書定義文では、下線部がひかれている動詞は全て主辞にかかっているが、いずれも機能を表していない。

- (28) うどん 小麦粉を水でこねてうすくのばし、細長く切ってゆでた 食べ物
- (29) 難病 なおりにくい 病気
- (30) 桜 山地にはえ、公園や庭にも 植える 木
- (31) 銀杏 葉は扇形で、秋になると 黄葉する 木

そこで、手がかり表現を含まない辞書定義文から機能を抽出するために、次の二つのことを仮定した。

- 動詞が動作の結果や、状態を表している場合、その動詞は機能を表さないとする。例えば(28)の「ゆでた」は動作の結果を、(29)の「なおりにくい」は状態を表しているので、これらは機能ではないと判断できる。
- 機能を持つ名詞は、職業や道具などの、限られた意味カテゴリーに属する名詞だけと仮定する(例えば「先生」「金槌」など)。例えば(30)の「桜」と(31)の「銀杏」は、ともに植物のカテゴリーに属していると考えられるが、「植物」カテゴリーの単語が機能をもつことは殆んどないので、「植える」と「黄葉する」は機能を表しているのではないと判断できる。

動詞が状態、動作の結果を表しているかどうかは、その動詞の活用形や、その動詞に続く接尾辞などから判断する。また、一般に、辞書定義文の主辞は見出し語の上位概念を表しているので、ある見出し語が特定の意味カテゴリーに属するかどうかは、その主辞((30)と(31)の例なら「木」)から判断可能である。

辞書定義文から機能を抽出する手順は以下のようなになる。はじめに、辞書定義文が手がかり表現を含んでいるかどうかを判定する。辞書定義文で使われる手がかり表現の種類は限られているので、手がかり表現のパターンを用意して、辞書定義文をそのパターンにマッチさせることによって判定を行う。パターンは全部で16種類作成した。表2にパターンの例を示す。そして、辞書定義文が手がかり表現を含んでいれば、その手がかり表現の直前の動詞を、機能を表す動詞として抽出する。辞書定義文が手がかり表現を含んでいなかった場合、主辞とそれに係る動詞が以下の条件を全て満たしていれば、その動詞を機能として抽出する。

- 主辞に係る動詞の活用形がタ形活用でない。
- 主辞に係る動詞の後に、「にくい」「いる」といった状態を表す接尾辞が接続していない。状態を表す接尾辞は11種類用意した。例)にくい、やすい、いる...
- 主辞が、道具や人など、あらかじめ用意した単語リストに含まれている。リストには18個の単語が含まれている。例)道具、装置、人、場所...

表 2 手がかり表現パターン

のに(使う 使われる 利用する 利用される) (仕事を 職業を 役を 役目を)している
--

5.3 表現のずれの吸収

以上の認識手法を実現するためには、要素と辞書定義文の間の表現のずれが問題となる。本論文では、この問題を辞書定義文を用いて解決する。以下では、重複表現の場合について、表現のずれの吸収方法を説明する。認識対象が迂言表現の場合でも、全く同様の方法で表現のずれを吸収できる。

重複表現の主要素の辞書定義文は、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い。この問題の主な原因は、被包含要素の同義語が使われていることである。また、特に非完全包含型の場合は、被包含要素の上位語が使われている。

(32) 貯金をたくわえる
貯金 お金を ためる こと

(33) ノーベル賞を受賞する
受賞 賞 をもらうこと

例えば(32)の主要素「貯金」の辞書定義文には、「たくわえる」ではなくて、その同義語の「ためる」が使われている。また、非完全包含型(33)の主要素「受賞」の辞書定義文には、「ノーベル賞」ではなく、その上位語の「賞」が使われている。

辞書定義文の主辞は、見出し語の同義語や上位語に対応しているので、被包含要素の辞書定義文を使えば表現のずれを吸収できる。例えば「話す」と「ノーベル賞」の辞書定義文は次のようになっているので、その主辞から同義語や上位語の情報を取り出すことができる。

たくわえる お金や品物を ためる こと
ノーベル賞 科学や文学などの偉大な業績に与えられる 賞

表現のずれを考慮せずに認識処理を行って、入力表現が重複表現と認識されなかった場合に、被包含要素の辞書定義文を利用して表現のずれを吸収する処理を行う(図5)。被包含要素が多義語の場合には、3.2節で用いたのと同様の手法で多義性解消を行う。

6 迂言表現と重複表現の言い換え

認識処理によって、迂言表現の付属要素がどのような働きを持っているか、また、重複表現においてどの要素が被包含要素であるかが分かる。このような情報を利用すれば、迂言表現と

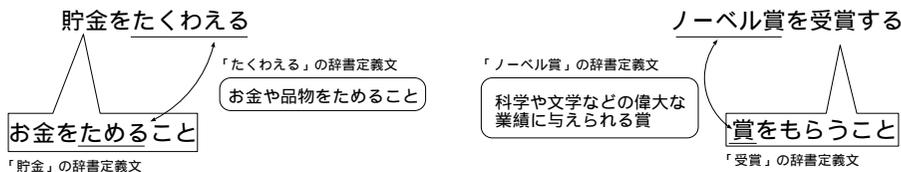


図 5 辞書定義文を用いた表現のずれの吸収

重複表現をその同義表現に言い換えることができる。

6.1 迂言表現の言い換え

認識に利用した付属要素の辞書定義文を利用することによって，付属要素を副詞表現や付属語に変換することができる。

- (34) a. 寄付を強制する → むりに寄付させる
- | | | | | |
|----|------------|----------|-----------|----|
| 強制 | <u>むりに</u> | <u>さ</u> | <u>せる</u> | こと |
| | 副詞表現 | 基本動詞 | 付属語 | |
- b. 非難を浴びる → 非難される
- | | | |
|-----|------------|----|
| 浴びる | <u>受ける</u> | こと |
| | 基本動詞 | |
- c. 守りを固める → しっかりと守る
- | | | | |
|-----|--------------|-----------|----|
| 固める | <u>しっかりと</u> | <u>守る</u> | こと |
| | 副詞表現 | 主要素の派生語 | |
- d. 故意に遅れる → わざと遅れる
- | | | | |
|----|------------|-----------|----|
| 故意 | <u>わざと</u> | <u>する</u> | こと |
| | 副詞表現 | 基本動詞 | |

(34a) は、「寄付」を「寄付する」に動詞化して，辞書定義文中の副詞表現「むりに」と付属語「せる」を付け加えることによって「むりに寄付させる」と言い換えることができる。(34b) の付属要素「浴びる」の辞書定義文は「受けること」である。「受ける」は，ヴォイス(受身)の働きをする基本動詞として，基本動詞リストに記述されている。そのため(34b)は「非難」を「非難する」と動詞化したのち，それを受身表現に変換することによって「非難される」と言い換えることができる。受身表現に変換する処理は，動詞の活用形を未然形に変換して，機能語「れる」を付与するというルールによって実現される。他のヴォイスやアスペクトの働きをする基本動詞を扱う場合にも，同様に人手で作成したルールを用いる。(34c) と(34d) も(34a) と同様に言い換え可能である。

6.2 重複表現の言い換え

重複表現は二通りの言い換えができる。

被包含要素の削除 認識された被包含要素を削除することによって、言い換えを行うことができる。

- (35) a. 賞を受賞する → 受賞する
 受賞 賞をもらうこと
- b. 貯金をためる → 貯金する
 貯金 お金をためること

提案手法によって、「賞を受賞する」の名詞要素「賞」や「貯金をためる」の動詞要素「ためる」が被包含要素であると分かる。その結果、それぞれの表現から被包含型を削除することによって、上記のような言い換えができる。(35b)のように名詞要素が主要素であれば、「貯金 → 貯金する」のように、主要素を動詞化する。

しかし、以下の四つの場合には、被包含要素を削除した言い換えはできない。

- 非完全包含型は被包含要素を削除すると元の意味が変化してしまうので、この言い換え方法を適用できない。

(36) ノーベル賞を受賞する

(37) 貯金をためこむ

- 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、被包含要素を削除すると、人間でも主要素の多義性を正しく解消できなくなる。

(38) 態度が軟化する → 軟化する

「軟化」の辞書定義文は「態度が穏やかになること」なので、これは対象包含型である。しかし「軟化」は多義語で、「物が軟らかくなること」というもう一つの辞書定義文を持っている。そのため、(38)から被包含要素「態度」を削除して言い換えを行うと、「軟化」の多義性を正しく解消できなくなる。

- 動作包含型を言い換えるためには、(35b)のように、主要素を動詞化する処理が必要となる。そのため、主要素が動詞化できない動作包含型は言い換えられない。

(39) 犯罪を犯す

- 機能包含型にはこの言い換え方法を適用できない。

(40) 先生が教える

本論文では「名詞 + 格助詞 + 動詞」形の限られた表現を扱っているが、より広い文脈の中では、(39) や (40) のような表現を言い換えられる場合が存在する。非包含要素の意味は包含要素の意味に含まれているため、(41) や (42) のように非包含要素を省略した表現に言い換えても、基本的には表現全体の意味は損なわれない。片岡らは、連体修飾節について同様の言い換えを行う方法を提案している(片岡, 増山, 山本 2000)。被包含要素は、片岡らのいう「語から連想される動詞」の一種であると考えることができる。

(41) 昔、犯罪を 犯したことを 後悔している

→ 昔の犯罪を後悔している

(42) その先生が 教えた おかげで、彼の成績が上がった

→ その先生のおかげで、彼の成績が上がった

主要素の変換 上記のような言い換えを行ったのち、残された主要素を、辞書定義文を使ってさらに言い換えることができる。

(43) 貯金をためる (→ 貯金する) → お金をためる

(44) 賞を受賞する (→ 受賞する) → 賞をもらう

すでに述べたように、被完全包含型は被包含要素を削除する言い換えはできないが、下に示すように、主要素の変換による言い換えは可能である。

(45) 貯金をためこむ → お金をためこむ

(46) ノーベル賞を受賞する → ノーベル賞をもらう

7 実験結果と考察

認識実験と言い換え実験の二つの実験で、提案手法を評価した。いずれの実験でも、辞書は例解小学国語辞典(田近 1997)を使用した。辞書定義文は、構文解析システム KNP(Kurohashi and Nagao 1994)を使って構文解析を行った。

認識実験 600 個の「名詞 + 格助詞 + 動詞」形の表現を用意して、適合率と再現率による評価を行った。テストセットは次のようにして作成した。まず、名詞要素と動詞要素が共に例解小学国語辞典に見出し語として登録されている「名詞 + 格助詞 + 動詞」形の表現を、解析済みの新聞記事コーパス(毎日新聞 1991 年 1 月分と 2 月分)からランダムに 600 個取り出した。そして、二人の被験者(被験者 A, 被験者 B とする)が個別に、600 表現のそれぞれに対して標準表

現，迂言表現，重複表現のいずれであるかを判断しタグ付けを行った．以下の表に実験の結果を示す．表 3は被験者 A，表 4は被験者 B の判断に基づく適合率と再現率である．それぞれの適合率は 84%と 71%，再現率は 53%と 51%であった．それらを合計して平均すると，適合率は $(41+35)/(49+49)=78\%$ ，再現率は $(41+35)/(78+68)=52\%$ であった(表 5)．

表 3 適合率と再現率 (被験者 A)

	適合率	再現率
迂言表現	0.90(38/42)	0.60(38/63)
重複表現	0.43(3/ 7)	0.20(3/15)
トータル	0.84(41/49)	0.53(41/78)

表 4 適合率と再現率 (被験者 B)

	適合率	再現率
迂言表現	0.79(33/42)	0.58(33/57)
重複表現	0.29(2/ 7)	0.18(2/11)
トータル	0.71(35/49)	0.51(35/68)

表 5 適合率と再現率 (平均)

適合率	再現率
0.78(76/98)	0.52(76/146)

同様のテストセットを用いて，多義性解消の処理を全く行わずに認識処理を行った結果を以下に示す．動詞要素が複数の辞書定義文を持っていた場合には，辞書定義文がどれか一つでも認識条件を満たしていれば，その表現を迂言表現/重複表現とした．被験者 A の判断にもとづく適合率と再現率は 71%と 59%であり(表 6)，被験者 B の判断にもとづく適合率と再現率は 58%と 56%であった(表 7)．それらを合計して平均すると，適合率は $(46+38)/(65+65)=65\%$ ，再現率は $(46+38)/(78+68)=58\%$ であった(表 8)．

表 6 適合率と再現率 (被験者 A)

	適合率	再現率
迂言表現	0.75(42/56)	0.67(42/63)
重複表現	0.44(4/ 9)	0.27(4/15)
トータル	0.71(46/65)	0.59(46/78)

表 7 適合率と再現率 (被験者 B)

	適合率	再現率
迂言表現	0.64(36/56)	0.63(36/57)
重複表現	0.22(2/ 9)	0.18(2/11)
トータル	0.58(38/65)	0.56(38/68)

表 8 適合率と再現率 (平均)

適合率	再現率
0.65(84/130)	0.58(84/146)

表 9 言い換え精度 (被験者 A)

	精度
認識	0.82(163/200)
言い換え	0.90(147/163)

表 10 言い換え精度 (被験者 B)

	精度
認識	0.85(169/200)
言い換え	0.92(155/169)

言い換え実験 次に、以下のような手順で、言い換え処理の評価を行った。まず、新聞記事コーパス (毎日新聞 1991 年 1 月分と 2 月分) から、提案手法が迂言表現または重複表現であると認識した「名詞 + 格助詞 + 動詞」形の表現をランダムに 200 個取り出した。そして、その 200 表現について、認識処理が適切であるかどうかを、認識実験のときと同一の二人の被験者が判断した。さらに被験者は、認識処理が適切であると判断された表現に対して、提案する言い換え処理が適切に働いたかどうかを判断した。重複表現には、非包含要素を削除する言い換えを行った。システムが重複表現を言い換え不能と判断した場合が一つだけあったが、被験者がその判断は妥当であると判断したため、それは言い換え処理が適切であったと数えた。下の表に実験の結果を示す。表 9 は被験者 A、表 10 は被験者 B の判断に基づく精度を表している。取り出された 200 表現のうち、それぞれの被験者によって認識処理が適切であったと判断されたものは、163 個と 169 個であった。そして、それらの中で、言い換え処理が適切に働いたと判断された表現は 147 個と 155 個であり、言い換え処理の精度はそれぞれ 90%、92%、平均で 91%であった。

考察 本手法によって言い換えられた表現の例を以下に示す。

- (47) 樹立に踏み切る → 思い切って樹立する
踏み切る 思いきってする
- (48) 承認を得る → 承認してもらおう
得る (1) 手に入れる
(2) してもらおう
- (49) お金を支払う → 支払う
支払う 相手にお金を払う

(47) と (48) は、それぞれ用言付属型として認識できた。(48) の用言要素「得る」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。(49) は対象包含型として認識することができた。

認識実験で、多義性解消を行った場合と行わなかった場合を比較すると、多義性解消を行った場合のほうが良好な結果となっている。例えば表 5 と表 8 を比べてみると、多義性解消を行った場合は、行わなかった場合より再現率が 6%下がっているが、適合率は 13%も上がっている。

この結果より多義性解消は有効に働いたといえる。

認識に失敗した大きな原因は、辞書定義文と要素のずれの吸収の失敗であった。本手法で扱える表現のずれは、同義語や上位語が原因となっているものだけなので、以下のような表現は扱えなかった。

(50) 構想をねる

ねる いろいろ考えてよいものにする

構想 考えをまとめる

(50) は動詞付属型 (特定付属型) であると考えられる。動詞付属型 (特定付属型) は、動詞要素の辞書定義文の主辞と名詞要素が一致すれば、提案手法によって正しく認識される。(50) の動詞要素の辞書定義文は「いろいろ考えてよいものにする」で、その主辞と名詞要素は一致していない。このような場合、提案手法が、表現のずれをうまく吸収して、動詞付属型 (特定付属型) を正しく認識するには、名詞要素の辞書定義文の主辞 (その名詞要素の同義語や上位語に相当する) と、動詞要素の辞書定義文の主辞が一致しなくてはならない。しかし、名詞要素の辞書定義文は「考えをまとめること」なので、結局、提案手法では (50) を正しく認識できない。しかし、動詞要素と名詞要素の辞書定義文は「考えて」と「考え」のところが部分的に一致しており、動詞要素と名詞要素はいずれも「考え」に関連する意味を持っていることが分かるので、今後はこのような情報も利用することが考えられる。

認識実験の結果をみると重複表現の認識結果が悪く、この部分には今後改善の余地があると思われるので、提案手法が誤って重複表現と認識した表現と、提案手法が重複表現と認識できなかった表現を調査した。まず提案手法が誤って重複表現と認識した例として、(51) ような慣用的な表現があった。(51) の動詞要素「ひきとる」は複数の辞書定義文をもつ多義語で、多義性解消処理により「息が絶える」という辞書定義文が選択された。この辞書定義文は入力の名詞要素と同じ格要素「息」をもっているため、システムは「息をひきとる」を対象包含型重複表現であると認識した。「息」と「ひきとる」は、(5c) の「馬がいなく」と同様、非常に強い選択制限で結び付いているので、重複表現であると言えなくもない。それにも関わらず、いずれの被験者にも重複表現であると判定されなかったのは、その結び付きが慣用的であったためと考えられる。今後は、重複表現と慣用表現の区別を明確にする必要があると考えている。

(51) 息をひきとる

ひきとる 息が絶える

次に、被験者が機能包含型であると判断したが、提案手法が認識できなかった表現として (52) があげられる。「笛」の辞書定義文は「合図のために、ふいて音を出す道具」で、提案手法は「(音を) 出す」だけしか機能として取り出せなかったため認識に失敗した。しかし、辞書定義文には「ふいて音を出す」という記述があるので、この部分を利用することができれば、(52) を

機能包含型として認識できる可能性がある。

(52) 笛を吹く

笛 合図のために、ふいて音を出す道具

当然,このような手がかりが副作用を及ぼす場合もあるため,すぐに(52)を機能包含型と認識することは難しい.例えば「鏡」の辞書定義文には「利用して,(顔や姿を)うつす」という記述があるが,「鏡」の機能は「(顔や姿を)うつす」だけで,「利用する」は機能でない考えるほうが適当である.そのため,「(顔や姿を)うつす」だけを機能として取り出す提案手法のままのほうが適切な結果を得られる.

(53) 鏡 光の反射を 利用して, 顔や姿をうつす 道具

不特定付属型の迂言表現を認識するには,基本的には不特定付属型の付属要素として働くことができる語をリストアップすればよい.そのため,不特定付属型の付属要素として働くことができる語が非常に限られているのならば,提案手法のように辞書定義文から自動学習する方法のほかに,人手でそのような語を整備するという方法も考えられる.不特定付属型の付属要素として働くことができる語の数を見積もるために辞書定義文を調べたところ,提案手法によって,1,356個の辞書定義文が不特定付属型の付属要素として働くことができると判定されていた.今回使用した例解小学国語辞典よりも大規模な辞書を使えば,この数字はさらに大きくなると思われる.この結果から,不特定付属要素として働くことができる単語の数は非常に多いのではないかと予想できるので,データを人手で整備するという手間のかかる手法よりも,提案するような自動的な手法のほうに利点があると考えられる.

8 関連研究

村木は,動詞付属型の付属要素となりうる動詞の分類リストを作成している(村木 1991).しかし,村木は付属要素が持つ副詞的な働きには注目していない.したがって,本論文で扱った「断行する」のように,副詞的な働きしか持っていない付属要素は分類の対象とはなっていない.さらに,「寄付を強制する → むりに寄付させる」のように,副詞表現(「むりに」)を伴った表現への言い換えは議論されていない.

言い換え処理に関する研究は,これまでに多数存在するが(Barzilay and McKeown 2001; Lin and Pantel 2001; Shinyama, Sekine, and Sudo 2002; Torisawa 2002; Barzilay and Lee 2003; Duclaye and Yvon 2003; Pang, Knight, and Marcu 2003),いずれの研究でも迂言表現や重複表現といったものは議論されていない.

Pustejovsky は,本論文で導入した「機能」と同様の概念を *telic role* と呼んでいる(Pustejovsky 1991).Pustejovsky は *telic role* を使って,形容詞の意味の解釈などの議論を行っている

る．例えば「タイピスト」の telic role は「タイプすること」なので、「速いタイピスト」は「速くタイプするタイピスト」と解釈される．Pustejovsky らはコーパスから telic role を学習する手法を提案しているので、今後はこのような研究を参考にしていきたい．

Boni らは、辞書定義文から手がかり表現を用いて telic role を学習する手法を提案している (Boni and Manadhar 2002)．しかし、すでに議論したように、手がかり表現を用いた手法だけでは不十分である．

動詞付属型の付属要素の中でも、「研究を行う」の「行う」や「世話をやく」の「やく」など、名詞要素を述語化する働きしか持っていないものは support verb と呼ばれている．support verb の認識に関しては、コーパスから自動学習する手法が提案されている (Grefenstette and Teufel 1995)．これに対して我々は、support verb の認識に対して、国語辞典というリソースが有効利用できることを示したと言える．

9 おわりに

本論文では「名詞 + 格助詞 + 動詞」形の迂言表現と重複表現を認識し、さらにそれらを言い換える手法を提案した．提案する認識処理と言い換え処理の精度を二人の被験者が判断したところ、認識処理の精度は、平均して適合率 78%、再現率 52%であった．また、言い換え処理の精度は平均して 91%であった．今後は、提案手法を応用して、書きことばを話しことばに変換するアプリケーションを開発することを考えている．また、情報検索や質問応答システムに組みこみ、システムの性能向上にどの程度寄与するかを観察することも、重要であると考えている．

参考文献

- Barzilay, R. and Lee, L. (2003). “Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment.” In *Proceedings of HLT/NAACL 2003*.
- Barzilay, R. and McKeown, K. R. (2001). “Extracting Paraphrases from a Parallel Corpus.” In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pp. 50–57.
- Boni, M. D. and Manadhar, S. (2002). “Automated Discovery of Telic Relations for WordNet.” In *Proceedings of the 1st International WordNet Conference*.
- Duclaye, F. and Yvon, F. (2003). “Learning Paraphrases to Improve a Question-Answering System.” In *Proceedings of the 10th Conference of EACL Workshop Natural Language Processing for Question-Answering*.

- Grefenstette, G. and Teufel, S. (1995). "Corpus-based Method for Automatic Identification of Support Verbs for Nominalizations." In *Proceedings of the 7th Conference of EACL*, pp. 98–103.
- Kaji, N., Kawahara, D., Kurohashi, S., and Sato, S. (2002). "Verb Paraphrase based on Case Frame Alignment." In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*.
- 片岡明, 増山繁, 山本和英 (2000). "動詞型連体修飾表現の“ N1 の N2 ”への言い換え." 自然言語処理, 7 (4).
- Kawahara, D. and Kurohashi, S. (2001). "Japanese Case Frame Construction by Coupling the Verb and its Closest Case Component." In *Proceedings of HLT 2001*, pp. 204–210.
- Kurohashi, S. and Nagao, M. (1994). "A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures." *Computational Linguistics*, 20 (4).
- Lin, D. and Pantel, P. (2001). "Discovery of inference Rules for Question Answering." *Journal of Natural Language Engineering*, 7 (4), pp. 343–360.
- 村木新次郎 (編) (1991). 日本語動詞の諸相. ひつじ書房.
- NTT コミュニケーション科学研究所 (編) (1997). 日本語語彙大系. 岩波書店.
- Pang, B., Knight, K., and Marcu, D. (2003). "Syntax-based Alignment of Multiple Translations: Extracting Paraphrases and Generating Sentences." In *Proceedings of HLT/NAACL 2003*.
- Pustejovsky, J. (1991). "The Generative Lexicon." *Computational Linguistics*, 17 (4), pp. 409–441.
- Shinyama, Y., Sekine, S., and Sudo, K. (2002). "Automatic Paraphrase Acquisition from News Articles." In *Proceedings of HLT 2002*.
- 田近洵一 (編) (1997). 例解小学国語辞典. 三省堂.
- Torisawa, K. (2002). "An Unsupervised Learning Method for Associative Relationships between Verb Phrases." In *Proceedings of the 19th International Conference on Computational Linguistics*, pp. 1009–1015.
- 山梨正明 (編) (1995). 認知文法論. ひつじ書房.

略歴

鍛冶 伸裕: 2000年京都大学工学部電気電子工学科卒業. 2002年京都大学大学院情報学研究科修了. 現在, 東京大学大学院情報理工学系研究科博士後期課程在学中. 自然言語処理の研究に従事.

黒橋 禎夫: 1989年京都大学工学部電気工学科第二学科卒業. 1994年同大学院

博士課程修了。京都大学工学部助手，京都大学大学院情報学研究科講師を経て，2001年東京大学大学院情報理工学系研究科助教授，現在に至る。自然言語処理，知識情報処理の研究に従事。

(2003年7月3日 受付)

(2003年9月25日 再受付)

(2003年10月23日 採録)