

## 統合的言い換え表現の抽出

呉 浩東<sup>†</sup> 下川 浩<sup>†</sup> 二宮 哲<sup>†</sup>

<sup>†</sup> 獨協大学外国語学部 〒340-0042 埼玉県草加市学園町 1-1

E-mail: {kgo, yshimo, sninomiya}@dokkyo.ac.jp

あらまし 言い換えに関する研究は自動要約、質疑応答、平易文生成と多岐の分野において重要な道具になりつつある。本稿では、語彙的換言処理、構文的換言処理と意味的換言処理を実現するために、辞書、シソーラス、コーパス、WWWなどの言語資源から、また言語学の視点から複数の手掛かりを得る方法を提案し、その有効性と適用性を論じる。

キーワード 言い換え、抽出、類似性、シソーラス、コーパス、WWW

## A Hybrid Model for Paraphrase Acquisition

Kotoh GO<sup>†</sup>, Yutaka SHIMOKAWA<sup>†</sup>, Satoshi NINOMIYA<sup>†</sup>

<sup>†</sup> Faculty of Foreign Languages, Dokkyo University 1-1 Gakuencho, Soka-shi, Saitama, 340-0042 Japan

E-mail: {kgo, taro, jiro}@dokkyo.ac.jp

**Abstract** Studies on paraphrasing are very important instruments in various research topics as automatic summarization, sentence generation, and question-answering. We in this paper propose a new approach to automatically extract paraphrase in lexical, syntactic and semantic levels using multiple cues extracted from linguistic resources as dictionary, thesaurus, parallel corpus and WWW, and cues from linguistic observations. This approach is applicable to many natural language tasks.

**Keyword** Paraphrase, Extraction, Similarity, Thesaurus, Corpus, WWW

### 1. はじめに

言い換えとは、ある言語表現をできるだけもとの意味を保持したまま同一言語の別の表現に変換する処理である。つまり、特定の言語表現が持つ各種のパラメータ（長さ、読みやすさ、スタイル等）を変更することである。英語学習辞書 COBUILD によると、「言い換え」(paraphrase) は次のように説明されている。

If you *paraphrase* something written or spoken, or the person who said it, you gives it meaning using different words.

近年、言い換えに関する研究は多くの研究者から注目され、言い換え表現の自動獲得の研究も重視されつつある。その研究は多岐の分野に展開されている。

- 機械翻訳の前編集と後編集  
構文構造の複雑さと語彙の曖昧さに対処することは従来の機械翻訳システムの難点である。ここでは、換言処理によって、曖昧な文型表現を簡単なものに分解することと、訳語選択の複雑さを軽減させるために単語を入れ替えるなどの対処を行う。
- 平易な文への置き換えによる読解支援  
第二言語の学習者や子供が自らの言語能力を補う。
- 要約の生成  
字幕生成や短信の生成に重要な道具である。
- 質問応答  
より柔軟な会話を生成するためには語彙・構文の言い換えは不可欠である。
- テキスト検索

言い換えによって検索的的中率が大幅に上げられる。

本稿では、語彙的・構文的また語用論的言い換え表現の自動抽出の新しいアプローチを提案し、その仕組みについて述べる。複数の言語資源から言い換え表現の抽出のための手掛かりについて論じ、その有効性を評価する。

## 2. 言い換え処理のモデル

言い換えの抽出のために、本稿では以下のものから必要な手掛かりを抽出する。

- 文脈的な類似性
- シソーラス
- コーパス
- WEB

そして、それぞれの手掛かりを用いて、語彙的言い換え、構文的言い換え、意味的言い換えの抽出の仕組みを提案する。

三つの換言処理：

- (I) 語彙的換言処理では、表記の違い(妻 ⇔ 家内 ⇔ ワイフ)や動詞の変化(入手する ⇔ 手に入れる)などを扱う。
- (II) 構文的換言処理は、語順の入れ換え(65キロの体重 ⇔ 体重65キロ)や述部の換言(コーヒーだけでいいです ⇔ コーヒーだけお願いします。)などを対象とする。
- (III) 意味的換言処理は、(I)と(II)併用しても処理できない換言を扱う。ここでは文脈や世界知識を使う場合が多い。たとえば、食堂で、塩がありますか?を相手に言う場合、塩をください?に言い換える。

以下、上記の三つの換言処理対策を論究する。以下で使う記号の意味は：

- s. は source の略。
- t. は target の略。
- A ⇔ B A 表現と B 表現が対等的なもので、お互いに置き換えられる。
- A ⇒ B A 表現を B 表現に置き換えることだけが可能である。

## 2.1 語彙的換言処理

(あ) 名詞の言い換え

- (1) s. タイン川流域のローマ時代の城塞  
t. タイン川流域のローマ時代の城

(い) 動詞の言い換え

- (2) s. 警察が殺人犯を逮捕する。  
t. 警察が殺人犯を逮まえる。

(1) シソーラスの同義語による換言処理

語彙的換言処理にシソーラス(類語辞書、たとえば、日本語語彙大系)は単語の同義語の抽出、単語間の意味的類似性の推定に非常に役立つ。図1は首相の同義語を示す。それによると

- (3) 内閣総理大臣 ⇔ 首相

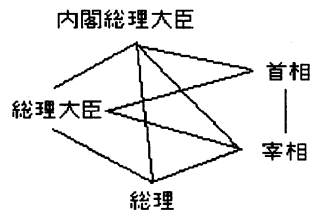


図1 シソーラスに同義語の例

以上のような語彙的類似性はシソーラス(類義関係を中心に語を分類し、それらの上位・下位関係を木構造で表した辞書)に基づく類似性のことである。

(II) 略語抽出による換言処理

- (4) 理化学研究所(理研)の使命は、我が国唯一の自然科学における総合研究機関として、第一級の自然科学の研究をし、さらにその成果を社会に還元することです。

から(理化学研究所 ⇔ 理研)という言い換えが得ることができる。同様に、

- (5) 昭和14年(1939年)9月1日に第二次世界大戦が勃発した。

から(昭和14年 ⇔ 1939年)を抽出することが

可能である。

また、辞書から略語とその本来の表現を見つける場合もよくある。

(6) ASAP ⇔ as soon as possible

## 2.2 構文的換言処理

構文的換言処理は言語知識によって換言規則を作り、適用可能なものを処理する。また、変形文法法を利用し語順を調節する。

(I) 否定表現の言い換え

ここでは、しかとだけを例とする。

変換規則 NしかVない ⇒ NだけV

(7) s. 高橋君しか英検一級に合格しなかった。

t. 高橋君だけ英検一級に合格した。

(II) 構文変形による言い換え

日本語の構文変換規則

s. [N1]+は+[N2]+を+[V]

t. [N2]+が+[N1]+に+[Vの受動態]

から

(8) s. 太郎は花子愛している。

t. 花子が太郎に愛されている。

そのほかに、「サ変動詞」を普通動詞に言い換えることや、格助詞の補完や置き換えなども上記のような変換規則によって実現できる。

## 2.3 意味的換言処理

(I) 辞書の定義文

国語辞書(広辞苑)またはEDR単語・概念辞書の定義文(語釈文)より言い換え表現を獲得することができる。たとえば、

「説得力」は「よく話して納得させること」から

(9) よく話して納得させる ⇔ 説得する

という言い換えを抽出することができる。

同じように、

(10) 物事を理解して承認する ⇔ 納得する

(11) 相手の言い分を聞き入れる ⇔ 承認する

が辞書から抽出できる。

(II) パラレルコーパスからの同(類)義表現の抽出  
パラレルコーパスで同じ原文から複数の翻訳がある場合、それらは言い換えと考えられる。

(12) s. いいですか?

t. よろしいですか?

t. いいのですか?

t. いいでしょうか?

同様に

(13) 収益の増加 ⇔ 増収増益

(14) 日本一の山 ⇔ 日本最高の山

## 3. 用例に基づく換言処理

言い換え表現の信憑性を確認するため、文脈にける類似性をコーパスあるいはWWWから下記のように計算する。F(x)がWWW上でxが観察されたページ数である。

ソース構造Psに複数の言い換え候補からPtを選択し、PsとPtの類似性を測る尺度スコアを以下の通り定義する:

$$SIM(P_s, P_t) = \log_2 \sum_{i=1}^n f(w_i) \frac{f(P_s, P_t)}{f(P_s)f(P_t)}$$

(1)

ここでは、 $w_i (1 \leq i \leq n)$ はPs, Ptに含まれる内容語(動詞、名詞、形容詞、副詞)である。

たとえば、雪が降る日は雪の日に換言できるかどうかを判断する際、Psは雪が降る日で、Ptは雪の日で、内容語はそれぞれ雪、降る、日(n=3)である。F(x)はxという言葉構造(単語、句、文節)が検索エンジンGOOGLEを使ってWEBから観測されたページ数である。

f(雪が降る日)	f(雪の日)	f(雪が降る日, 雪の日)	内容語頻度合計	SIM
916	144000	252	25360000	9.08

今までの実験では、SIMの閾値は6.8以上の場合正解率が80%を超えている。

そして、

(15) 雪が降る日 ⇒ 雪の日

以上の方法を用いて、

(16) 雨が降る日 ⇒ 雨の日

の換言処理もできる。

また、(15)の結果から、シソーラスによる方法も利用可能である。雨が降る日がウェブから観測されるので、雨と雪が強い類似性を持つことがシソーラスから観察されれば、(15)の例から(16)が生成できる。このような言い換えのプロセスを用例に基づいた方法と呼ぶことができる。

#### 4. 考察

本稿で提案した手法で、カタカナ(外来語)を他の日本語に言い換えられる。たとえば、

(17) プライムミニスター ⇔ 総理

語彙的な言い換え処理にシソーラスから同義語による置き換え処理を用いると、使用頻度が低い単語を使用頻度高い単語に置き換えられる。たとえば

(18) ワイフ ⇒ 妻

(17, 18)のような換言処理によって言葉がより日常的・平易のものになる。あるいは、求める文体に即したものとすることができる。

シソーラスの階層構造の上位一下位関係によって、類義的言い換え表現が得られる。たとえば、

(18) りんご、バナナとイチゴを買った。⇒ 果物を買った。

しかし、下位概念は上位概念に置き換えるとき過剰一般化の問題が生じる可能性が高い。たとえば、

(19) りんご、バナナとイチゴが嫌い。≠ 果物が嫌い。

というようなケースの対応が難しい。

#### 5. まとめ

本稿では、言語学また計算学的な視点から、辞書、コーパス、WWW、変形文法などを活用し、語彙的・構文的・意味的レベルにおける言い換への抽出にそれぞれ有効な方法を提案した。これから、言い換への部品化を目指し、機械翻訳・自動要約・第二言語学習支援など分野に統合的な言い換えモデルを構築する。

#### 文 献

- [1] 乾健太郎. “言語表現を言い換える技術.” 言語処理学会第8回年次大会チュートリアル資料, pp.1-21, 2002.
- [2] 鍛冶伸裕, 河原大輔, 黒橋禎夫, 佐藤理史. “ゲームの対応付けに基づく用言の言い換え.” 自然言語処理, Vol.10, No.4, pp.65-81, 2003.
- [3] 鍛冶伸裕, 岡本伸裕, 黒橋禎夫. “WWWを用いた書き言葉特有語彙から話し言葉語彙への用言の言い換え.” 自然言語処理, Vol.10, No.5, pp.19-37, 2003.
- [4] 乾健太郎, 藤田篤. “言い換え技術に関する研究動向.” 自然言語処理, Vol.11, No.5, pp.151-198, 2004.
- [5] NTTコミュニケーション科学研究所(1997). 日本語語彙大系. 岩波書店.
- [6] Wu, Haodong. “Contextual Word Similarity and Its Applications.” Dokkyo Studies in Data Processing and Computer Science, No.20, pp.21-28, 2002.
- [7] 山本和英. “換言処理の現状と課題.” 第7回言語処理学会年次大会ワークショップ論文集, pp.93-96, 2001.
- [8] 下川浩. 現代日本語構文法 大久保文法の継承発展, 三省堂, 1993.
- [9] 鳥澤健太郎. “教師無し学習による名詞句の言い換え.” 言語処理学会第8回年次大会年次大会発表論文集, pp.323-326, 2002.