

言い換え技術に関する研究動向

乾 健太郎[†] 藤田 篤[†]

意味が近似的に等価な言語表現の異形を言い換えと言う。言い換え技術とは、所与の言語表現からその言い換えを生成する言い換え生成技術、および所与の言語表現対が言い換え関係にあるか否かを判定する言い換え認識技術の総称である。これらの技術は、機械翻訳の前編集や読解支援のための文章簡単化、質問応答や複数文書要約など、様々な応用に貢献する応用横断的なミドルウェア技術になると期待されており、近年研究者の関心を集めてきた。本論文では、こうした言い換え技術について、工学的研究を中心に近年の動向を紹介する。具体的には、言い換えの定義や言い換え技術の応用可能性について論じた後、構造変換による言い換え生成、質問応答・複数文書要約のための言い換え認識に関する研究を概観し、最後に言い換え知識の自動獲得に関する最新の研究動向を紹介する。

キーワード: 言い換え, 言い換え生成, 言い換え認識, 言い換え知識獲得, 意味の同一性

A Survey on Paraphrase Generation and Recognition

KENTARO INUI[†] and ATSUSHI FUJITA[†]

Paraphrases are alternative ways of conveying the same content. The language technology for processing paraphrases, namely, paraphrase generation and paraphrase recognition, has drawn the attention of an increasing number of researchers because of its potential contribution to a wide variety of natural language applications. This survey paper overviews recent research trends in paraphrase generation and recognition, and discusses future prospects, addressing the issues of the definition of paraphrases, transformation-based paraphrase generation, paraphrase recognition in question answering and multi-document summarization, and finally corpus-based knowledge acquisition.

KeyWords: *paraphrasing, paraphrase generation, paraphrase recognition, paraphrase acquisition, sameness of meanings*

1 はじめに

- (1) a. 旧友と飲む酒ほど楽しいものはない。
b. 昔の友達と飲む酒が一番楽しい。
- (2) a. 内戦状態に再突入する公算が大きい。
b. 再び内戦状態になる可能性が高い。

この例のように、言語には同じ情報を伝える表現がいくつも用意されている。意味が近似的に

[†] 奈良先端科学技術大学院大学情報科学研究科, Graduate School of Information Science, Nara Institute of Science and Technology

等価な言語表現の異形を言い換え (paraphrase) と言う。言い換えを指す用語には他に、言い替え、換言、書き換え、パラフレーズといった語も使われるが、統一のため本論文では一貫して「言い換え」という用語を使う。

これまでの言語処理研究の中心的課題は、曖昧性の問題、すなわち同じ言語表現が文脈によって異なる意味を持つ問題をどう解決するかにあった。これに対し、言い換えの問題、すなわち同じ意味内容を伝達する言語表現がいくつも存在するという問題も同様に重要である。

与えられた言語表現からさまざまな言い換えを自動生成することができれば、たとえば、所与の文章を読み手の読解能力に合わせて平易な表現に変換したり、音声合成の前編集として聴き取りやすい表現に変換したりすることができる。あるいは、機械翻訳の前編集として翻訳しやすい表現に変換するといったことも可能になるだろう。また、与えられた2つの言語表現が言い換えであるかどうかを自動判定することができれば、情報検索や質問応答、複数文書要約といったタスクにおける重要な問題の一つが解決する。

近年、こうした問題に関心を持つ研究者が増え、言い換えというキーワードが目立つようになってきた。本学会年次大会でも、2001年に言い換えのセッションが設置されて以来、4件 (2001年)、9件 (2002年)、10件 (2003年)、7件 (2004年) と投稿を集めた。また2001年、2003年には言い換えに関する国際ワークショップが開かれ、それぞれ8件、14件の発表、活発な議論が行なわれた (Sato and Nakagawa 2001; Inui and Hermjakob 2003)。

本論文では、言い換えに関する工学的研究を中心に、近年の動向を紹介する。以下、まず、2節で、言語学的研究および意味論研究における言い換えに関連の深い話題を取り上げ、言い換えの定義について考察する。次に、3節で言い換え技術の応用可能性について論じた後、4節で構造変換による言い換え生成、質問応答・複数文書要約のための言い換え認識に関する研究を概観する。最後に5節で言い換え知識の自動獲得に関する最新の研究動向を紹介する。

2 言い換えとは？

そもそも言い換えとはどのようなものか？ どのような種類があるのか？ 直感的なイメージをつかむには、末尾の付録Aに目を通されるのがよいかもしれない。日本語の言い換え現象を構文的特徴に基づいて整理してある。

では、こうした言い換えはどのように定義されるか？ 英語学習辞典 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 give its **meaning** using different words.

明らかなのは、言い換えを定義するには、“meaning” が何を指すか、つまり言葉の「意味」とは何かという問題に深く立ち入らなければならないということである。言語の意味論が多くの議論を要する、形式化が困難な問題であることは読者の良く知るところであろう。しかも、意

味論研究の成果と自然言語処理技術の現状には依然として深いギャップがある。ここでは言語学研究および意味論研究における言い換えに関連の深い話題をいくつか紹介し、言い換えとは何かを考える材料を提供する。

2.1 言語学から見た言い換え

2.1.1 変形文法

理論言語学における言い換への扱いは初期の変形文法まで遡る。変形文法における変形は、表層構造（統語構造）に対する統語的操作で、たとえば、能動文から受動文に変形する操作は次のように記述される。

能動文：	1 : NP	2 : [+V, +AUX]	3 : [+V, -AUX]	4 : NP
受動文：	4	2	BE EN	3 BY 1

こうした操作は、表す意味を保存したまま統語構造を変えるという意味で構造的な言い換えと見なすことができる。(Harris 1981) には、文から文への変形、名詞句から文へ変形といった具合に対象の粒度によって変形規則を約 20 種類に分類したものが掲載されている（ただし、これらの変換規則が構文的言い換えをどの程度カバーするかは明らかでない）。こうした研究の成果は言い換への工学的実現にも有益であろう。

変形規則のもう一つの問題は、変形の語彙依存性をうまく扱えていないことである。ある変形の適用が可能となる条件は単語に依存する 경우가少なくない。たとえば、上の受動化の例をとってみても、他動詞であれば常に受動化できるわけではなく、「resemble」のような動詞の場合は適用できない」といったきめの細かい条件指定が必要になる。これは個々の語の特性に立ち入る必要があることを意味しているが、変形文法のような伝統的文法理論ではそれらを明らかにする活発な研究は見られなかった。

2.1.2 Meaning-Text Theory

変形の語彙依存性の記述を試みた例として、Mel'čuk らが発展させた Meaning-Text Theory (以下, MTT) があげられる (Mel'čuk and Polguère 1987; Iordanskaja et al. 1991; Wanner 1994; Mel'čuk 1996; Iordanskaja et al. 1996)。MTT では、意味構造から深層統語構造、表層統語構造を経て音韻構造にいたるまで 7 層の表現レベルを用意し、次の 2 種類の規則で言い換えを説明する。

- **変形規則**：各レベル間の対応を変形規則として記述する。同じ意味構造に対して異なる変形規則を適用すると異なる文が得られる。これらの文は互いに言い換への関係にある。
- **言い換え規則**：同じレベルの表現どうしの間で起こる変形を言い換え規則として記述する。言い換え規則を適用すると言い換えが得られる。

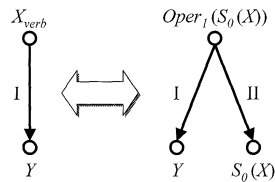


図 1 MTT における言い換え規則の例 (Iordanskaja et al. 1996)

変形規則と言い換え規則の語彙依存性の記述には語彙関数 (lexical function) を用いる. 語彙関数とは, 語の共起関係を記述するための道具で, たとえばある動詞 X に対してその名詞形を返す $S_0(X)$ と, ある動詞の名詞形 Y に対して元の動詞形の主語を同じく主語とするような機能動詞を返す $Oper_1(Y)$ という 2 つの語彙関数を組み合わせると, (3) のような深層統語構造の言い換え規則によって, (4) のような言い換えを記述することができる.

(3) $X_{verb} \Leftrightarrow S_0(X) + Oper_1(S_0(X))$ (図 1 のような構造を仮定)

$(S_0(decrease_{verb}) = decrease_{noun}, Oper_1(decrease_{noun}) = show)$

(4) s. Employment decreased sharply in October.

t. Employment showed a sharp decrease in October.

Mel'čuk (1996) によると, 言語の記述に必要な語の共起関係は 60 種類の言語独立な語彙関数でカバーできるとしており, それらの関数を組み合わせて表現する図 1 のような言い換え規則もまた言語に依存しない一般的な規則で記述する. ただし, 実際には個々の言語の語彙について語彙関数の大規模な辞書が必要となるが, 残念ながらそうした辞書はいまだ存在しない.

2.1.3 言い換えの構成性

変形文法や MTT における言い換えの扱いでもう一つ重要な原理は言い換えの構成性である. たとえば, 次の (5) の言い換えは, $purchase \Rightarrow buy$ という語彙的な言い換えと, $X \text{ be VERB-PP by } Y \Rightarrow Y \text{ VERB } X$ のような一般的な態交替の規則を組み合わせによって記述できる.

(5) s. This car was purchased by him.

t. He bought this car.

明らかに, 言い換えをよりプリミティブな変形に分解して記述するアプローチは, 分解せずに記述するアプローチに比べると理論的にも工学的にも理にかなっている. ただし, 前者を採用することにしても, それ以上分解できないプリミティブな言い換えにはどのような種類があるのか, それらは網羅的に数えあげられるのか, また, そもそも分解の可否を判断する基準を明確に規定できるか, という問題が残る. たとえば, MTT の研究者らはこれらを究明すべき中心的な問題の一部と考え, 前述のように言い換え規則や語彙関数としてその成果を形にしてい

た。しかし、語彙関数の定義が厳密性を欠くなど、不十分な点も多く、大規模な評価も試みられていないのが現状である。

2.1.4 言い換えの言語横断性

言い換えの言語横断的共通性の解明も注目すべき言語学的知見の一つである。

MTT では、フランス語や英語など、複数の言語の言い換えを対比させることによって、言語に依存しない言い換え規則を規定しようとしてきた。Mel'čuk (1996) によると、すべての言語の言い換えは 60 種類の語彙関数を組み合わせた言い換え規則で記述できるとされる。こうした成果が本当に正しいかどうかは十分な経験的評価による証明を待たねばならないが、少なくとも言い換えるモデル化に関して有益な知見を提供していることは間違いない。

経験的方法と組み合わせた大規模な調査も報告されている。たとえば、Kageura et al. (2004) は、日本語と英語の各々で同じ言い換え現象、とくに複合語を中心とした専門用語の言い換えを取り上げ、(6) のような日英各々の言い換え規則集合 (Jacquemin et al. 1997; Yoshikane et al. 2003) を用いて、両言語における言い換え可能性の共通性を調査している¹。

(6) j. 概念学習 \Rightarrow 概念を学習する ($X_1 N_2 \Rightarrow X_1$ “を” N_2 する)

e. word category \Rightarrow categorize words ($N_1 N_2 \Rightarrow V(N_2) N_1$)

彼らは、事例から演繹的に 50~60 個の変換規則を作成し、これを用いて、ある技術用語集の 6 割強の複合専門用語に対する両言語同時の言い換えを可能にしている。

ほかにも、(7) のような分裂文の言い換え (砂川 1995; Dras 1999) や (8) のような機能動詞結合の言い換え (村木 1991; Iordanskaja et al. 1991; Dras 1999; Fujita et al. 2004) など、さまざまな言語に共通の言い換え現象である。

(7) j. 収録されているのは約千人の人物だ。

\Rightarrow 約千人の人物が収録されている。

e. *It was his best suit that John wore to the dance last night.*

\Rightarrow John wore *his best suit* to the dance last night.

(8) j. 住民の強い要請を受け、廃棄物処理場の建設を中止した。

\Leftrightarrow 住民に (から) 強く要請され、廃棄物処理場の建設を中止した。

e. *Employment showed a sharp decrease in October.*

\Leftrightarrow *Employment decreased sharply in October.*

言い換えるの定式化・類型化にはこうした言語横断的視点からの分析が欠かせない。

2.1.5 言い換えるの機能論的説明

人間は言い換える。それはなぜだろうか？ この疑問に対する答えはいくつか考えられる。

¹ (6) の言い換え規則において、 N_i は名詞、 V_i は動詞、 X_i は任意の内容語、 $V(N_i)$ は名詞 N_i の動詞形を表す。

Walker (1993) は、言い換えによって意思の疎通をはかっていると指摘している。また、高塚 (1999) は、テクニカルコミュニケーションの立場から、第二言語学習者にとって有用な語彙の言い換えに着目し、言い換えの機能として次のような項目をあげている。

- 第二言語の学習者が自らの言語能力を補う。
- 相手の理解を促進させるために自分の先行発話を言い換える。
- 意味を確認するために相手の発話を言い換える。

上の例のようなコミュニケーションの促進という機能の他にも、言い換えは、社会的関係を保持する道具として用いられる。国広 (2000) は、会話の相手のメンツを傷つけることを避ける婉曲表現、仲間との連体性を保つ集団語として言い換えが用いられると述べている。自治体、企業のサービス窓口やアナウンサー向けのマニュアルにも社会的に不適切な表現の使用を避けるような項目が設けられており、代替表現（言い換え）が提示されていることもある。

2.2 意味が同じであるとは？

2.2.1 真理値意味論の意味の同一性

変形文法や MTT を含む多くの現代文法理論では、意味構造（深層構造）が同一であれば同義であり言い換えであると仮定しており、これが言い換えの定義になっている。ここで言う意味とは、語あるいは言語表現の内包的な意味（intension）を指す。本論文では、内包的意味の同一性に基づく言い換えを、以下で述べる参照的言い換えや語用論的言い換えと区別するために、語彙・構文的言い換え（lexical and structural paraphrase）と呼ぶ。

内包的意味が同一かどうかの判断は、真理値意味論（モデル論の意味論）を仮定すると考えやすい場合が少なくない。真理値意味論では、意味を真理値への写像と見なして定式化する。たとえば、「本」の意味は、個体 x が本であれば真、本でなければ偽を返す真理関数「本」(x) で与えられる。これによると、二つの表現があるとき、それぞれの真理関数において任意の個体の写像先の真理値がつねに同じであれば、またそのときに限り、それらの意味は同一である。「書籍」の関数「書籍」(x) を真にする個体の集合、すなわち書籍という概念の外延（extension）が「本」のそれと同じであれば、「書籍」と「本」は意味が同じといえる。よく知られるように、こうした見方は、「太郎が本を読む」のような命題を表す表現にも拡張できる。

ただし、こうした真理値意味論が言葉の意味を表現するのに十分でないことは明らかであり、真理値意味論の問題に関する議論は枚挙にいとまがない。以下、言い換えの定義に深く関連する問題について論じる。

2.2.2 言外の意味

変形文法では、前述のように能動文と受動文の対は同じ意味構造を持つと仮定した。誰が何をどうしたかという命題部分の意味は同じと考えて良いだろう。こうした仮定はたとえばより

最近の主辞駆動句構造文法 (Head-driven Phrase Structure Grammar; HPSG) にも受け継がれている。しかし、能動文と受動文は、話者の視点が違っていたり、どの情報を強調するか、あるいはどの情報が新情報かといった情報構造の違いがあるので、文脈によっては置換できない。現在の文法理論が仮定する意味表現は、こうした意味の違いを十分に扱えていない。

こうした問題は、最も基本的なものに見える単語間の言い換えにも見られる。たとえば、前述のように「書籍」と「本」は真理値意味論の意味、すなわち指示的意味 (denotation) はほぼ同じだが、厳密にはフォーマリティなどの暗示的意味 (connotation, いわゆるニュアンスあるいは言外の意味) の違いがあり、いつでも置き換えられるわけではない。言語は同義語を嫌う (Clark 1992)。同じ意味を持つ語があったとしても、語用論的な力が作用して、次第に違う意味、とくに違うニュアンスを帯びるようになる。したがって、実際には完全に意味が同じで常に置換可能な同義語はまれである。

言い換えと認められる表現対の多くになんらかの意味の違いがあるとすれば、ただちに、言い換え対の意味の差にはどのような種類があり、どのように記述すればよいかという問題が出てくる。これについては、すでに多くの言語学的蓄積 (Halliday 1994; 宮島, 仁田 1995a, 1995b; 影山 2001) があるものの、全貌はまだ遠く見えない。さらに、こうした言い換え対が所与の文脈で置換可能であるためには、意味の差がその文脈で無視できるものなくてはならない。したがって、言い換えの理論化には、言い換え対の意味の差分が所与の文脈に照らして無視できるかどうかを判別する機構の説明が必要である。

2.2.3 参照の同一性

語の意味に関する哲学的考察のなかで古くから論じられてきたように、言語表現の内包の意味が同じであることは、その表現の参照対象が同一であることと必ずしも一致しない。たとえば、Frege の「宵の明星」「明けの明星」の例は有名である。「宵の明星」と「明けの明星」は同じ参照対象を持つが、明らかに内包の意味は異なる。両者が言い換え可能 (置換可能) となる文脈を考えるのは容易でない。このように、参照対象が同一であることは言い換え可能であるための十分な条件にはならない。

ただし、内包の意味が違っていても、参照の同一性に基づいて言い換えることができる場合がある。典型的なのは、次の例のような参照表現の言い換えである (佐藤 1999)。

(9) s. 去年の出来事

t. 1998 年の出来事

(10) s. 筆者の考え

t. 佐藤の考え

こうした言い換えは特定の大域的文脈、談話の状況でのみ成り立つもので、内包の意味の同一性に基づく言い換えとは区別するべきである。本論文では、この種の言い換えを参照的言い換

え (referential paraphrase) と呼ぶ。

2.2.4 語用論的效果の同一性

先に意味の同一性を真理値意味論的に捉えた場合を議論したが、真理値意味論の欠陥を踏まえて状況論の意味論や言語行為理論が登場した経緯から容易に推測されるように、真理値意味論に基づいて言い換えを論じるのには限界がある。その一つが、語用論的效果の同一性に基づく言い換えである。言葉の語用論的效果とは、話者がそれを発することによって達成できると期待するコミュニケーションの目的である。次の例のように同じ語用論的效果を持つ発話は言い換え可能である (佐藤 1999; 川村 2000)。

(11) s. どなたか gcc のソースのありかをご存知ないでしょうか。

t. gcc のソースが置いてある ftp サイトを教えてください。

(12) s. Will you break this bill?

t. I want to use that vending machine.

これらの言い換えは、仮に内包的意味が真理値意味論的に与えられたとしても、同じではない。佐藤 (1999) はこうした言い換えを語用論的言い換えと呼んでいる。本論文でもこれに倣い、語用論的效果の同一性に基づく言い換えを語用論的言い換え (pragmatic paraphrase) と呼ぶ。

以上の議論をまとめると、言い換えには少なくとも、語彙・構文的言い換え、参照的言い換え、語用論的言い換の 3 種類がある。参照的言い換えは言葉が発せられた文脈や談話の状況を参照する必要がある。また、語用論的言い換えは明らかに、代表的な現代文法理論で仮定している意味論を超えるものである。このうち工学的実現が最も容易に見えるのは語彙・構文的言い換えである。実際、言い換えに関する工学的研究のほとんどが対象をこの語彙・構文的言い換えに限定している。以下、本論文でも語彙・構文的言い換えに話題をしぼる。

3 言い換え技術の使い方

言い換え技術の用途は広い。言い換えの実現方法に話をすすめる前に、さまざまな言い換えが自動化できるようになるとどのような使い方ができるかを整理してみよう。

3.1 人間のために言い換える

電子化文書データの爆発的な増加を背景に、そうした文書を利用者や利用形態に適した形に自動編集する技術の必要性が説かれるようになって久しい。冒頭の例のように、高齢者や子供、外国人、障害者など、利用者の言語能力にあわせて読みやすい平易な文面になおすタスク (Carroll et al. 1998; Canning and Tait 1999; 乾 2001; Higashinaka and Nagao 2002; Inui et al. 2003) はそのような編集の一例である。また、ニュース原稿から字幕を生成したり、Web の文

書を携帯端末に表示したり, ニュースを街頭や新幹線の電光掲示板に表示したい場合は, 1 行当たりの字数を考慮してコンパクトな表現に言い換える技術が必要になる (Robin and McKeown 1996; Kondo and Okumura 1997; 福島ら 1999; 三上ら 1999; 江原ら 2000; 片岡ら 2000; Masuda et al. 2001; 佐藤ら 2004; 池田ら 2004).

言い換え技術は, 人間が文書を書く現場でも有用である. 読みやすい文書を書く, スタイルを統一する, 規定の語彙と構文を使って (制限言語文書を) 書く, といった作業を支援する推敲支援でも, 読みにくい文や制限言語に合わない文を自動的に適切な文に言い換える技術が必要とされている (林, 菊井 1991; 高橋, 牛島 1991; 武石, 林 1992; Dras 1999; Mitamura and Nyberg 2001).

同様のことは, 機械翻訳や要約など, 機械が文章を生成する場合にもいえる. 機械が出力した文章をチェックし, 適格でない表現があれば自動的に修正するといった後編集 (Knight and Chander 1994; Mani et al. 1999; Nanba and Okumura 2000) が実現するとありがたい. また, 人間が要約する場合は, 原文にはない表現をうまく使って内容をまとめることができるが, これなども言い換えの一種といえ, そのような言い換えをいかに自動化するかが自動要約の重要な課題になっている (Kondo and Okumura 1997; 奥村, 難波 1999; 片岡ら 2000; 奥村, 難波 2002).

3.2 言語の機械処理のために言い換える

言い換えた結果を消費するのは人間ばかりとは限らない. 言い換えは入出力が同一言語であることから, さまざまな言語処理アプリケーションの中に部分タスクとして組み込むことができる.

機械翻訳では, 前編集段階で機械処理に適した言語表現にあらかじめ書き換えておくと訳質が上がる (白井ら 1995; 加藤ら 1997; 吉見ら 2000). この前編集を自動化する試みがすでに多数報告されている (金, 江原 1994; 白井ら 1995; Chandrasekar et al. 1996; Nyberg and Mitamura 2000; 吉見, 佐田 2000; 吉見ら 2000; 今村ら 2001; Yamamoto 2002a). 機械翻訳の他にも, 手話への翻訳のための前編集 (手話に変換しやすい表現に言い換える) (安達 1992; 徳田, 奥村 1998) や音声合成のための前編集 (耳で聴きとりやすい表現に言い換える) など, 前編集としての言い換え技術の潜在的応用範囲は広い.

言語には同じ内容を指す表現がいくつも用意されている. (佐藤 2001) の例を引こう.

(13) a. お名前をお願いしたいのですが.

⇒ Could you tell me your name, please?

b. お名前を頂戴することはできますか.

⇒ Could you tell me your name, please?

(13a) と (13b) はだいたい同じ意味の発話で, 同じ訳文を当てることができる. したがって, (a) を翻訳できる人なら, (b) も翻訳できるだろう. しかし, 翻訳システムが (a) と (b) の同義性を

理解できず、それぞれ別々に翻訳しようとする、(a)は翻訳できるが、(b)は翻訳できないといったことになりかねない。その場合でも、仮に前編集段階で(b)を(a)に言い換えることができれば、翻訳システムも対応できることになる。前編集が機械翻訳に効果的と考えられるのは、言語が持つこのような表現の多様性を前編集段階で吸収できることが期待されるためである。

言語表現の多様性が機械処理を難しくしている例は翻訳にとどまらない。文書集合に対する情報検索や質問応答では、検索要求や質問に使われる言語表現とそれに該当する記述の言語表現が異なれば、単純なキーワード照合ではうまく応答できない。これに対し、たとえば、

(14) a. 《著作名》の著者は《人名》だ。

b. 《人名》が《著作名》を発表する。

という2つの言い回しが広い意味での言い換えの関係になっていることを認識できれば、『坊ちゃん』の著者は誰ですか?』のような質問の答えを、情報源となる文書中の「夏目漱石が『坊ちゃん』を発表した明治39年は、…」のような記述から探し出すことができる。検索質問拡張(term expansion)は、その近似的な解決策の一つであるが、より洗練された同義性判定、すなわち言い換えの認識の仕組みが必要であることは明かであり、すでにさまざまな試みが報告されている(Jacquemin et al. 1997; 日笠ら 1999; Anick and Tipirneni 1999; 平田ら 2000; 白木, 黒橋 2000; Tomuro and Lytinen 2001; Ravichandran and Hovy 2002; Hermjakob et al. 2002; Sasaki et al. 2002; Duclaye et al. 2003; Moldovan et al. 2003; Takahashi et al. 2003; Yoshikane et al. 2003)。また、情報抽出において多様な言い換え表現から同じ情報を抽出する問題(関根 2001; Shinyama et al. 2002; Shinyama and Sekine 2003)や、複数文書要約において個々の文書から抽出したパッセージの中に同じ情報を冗長に伝える記述がないかどうかを判定する問題(McKeown et al. 1999; Barzilay et al. 1999; 上田, 小山 2000; 成松ら 2002; Barzilay 2003)なども、同様に言い換え認識の問題といえる。

3.3 言い換えを研究の道具として使う

言い換え技術は言語処理研究の道具としても使える可能性がある。直接的な使用例の一つに、機械翻訳システムの評価用正解翻訳例の自動生成がある。機械翻訳の研究では、評価用の各原文に対して複数の正解翻訳例を用意し、システムの出力を評価するのが一般的になってきた。たとえば、BLEU (Papineni et al. 2002) と呼ばれる自動評価手法では、豊富な正解例を用意することが評価結果の信頼性を確保するのに必要なことが分かっている。しかしながら、いくつもの正解翻訳例を手で作るのはコストが高い。こうした背景から、代表的な正解例からその他の翻訳例を言い換え生成によって自動的に入手する試みがいくつか報告されている(Pang et al. 2003; Kanayama 2003)。

4 言い換えの実現方法

3 節で述べたように、言い換えの工学的な処理は大きく生成と認識の 2 種類に分けられる。

言い換え生成は、与えられた言語表現からその言い換えを生成する作業であり、3 節で述べたように機械翻訳の前編集や後編集、文章読解の支援など、さまざまな応用がある。言い換え生成は、言語表現を入力とする生成という意味で、テキストからのテキスト生成 (text-to-text generation) の一種と行うことができる。これに対し、自然言語生成の文脈では、同一の意味構造からさまざまな言い換えを生成する作業を指して「言い換え生成」と呼ぶことがある。両者は相互に深く関係しあう問題であり、独立な問題として考えるべきではないが、混乱を避けるため、本論文ではもっぱらテキストからのテキスト生成の意味で言い換え生成という用語を使う。

言い換え生成は、意味を保存しながらある言語表現を別の意味表現に変換するという意味で、同一言語内の翻訳と見なすことができる。逆に言い換えの方を広く捉えて、翻訳は異なる言語間をまたぐ言い換えであると言ってもよい。本節では、まず 4.1 で言い換え生成の研究を歴史の長い機械翻訳研究と対比させながら紹介し、4.2 で機械翻訳研究では顕在化しなかった新しい問題を論じる。以下、「言い換え生成」と「(機械) 翻訳」を区別するため、それぞれを狭い意味で用いる。すなわち、「言い換え生成」は意味を保存したまま同一言語内の別の表現に変換する作業、「(機械) 翻訳」は意味を保存したまま他の言語の表現に変換する作業を指す。

一方、言い換え認識は、2 つの異なる言語表現が言い換えかどうかを判別する作業であり、情報検索や質問応答、複数文書要約などの応用がある。この問題は、一方から言い換えを生成して他方に到達できるかを調べる問題と見なせるので、言い換え生成と裏表の関係にあるといえる。4.3 ではこの点を論じる。

4.1 同一言語内翻訳としての言い換え生成

4.1.1 機械翻訳と言い換え生成

言い換え生成の研究が機械翻訳研究の長い歴史から学べることは多い。たとえば、統語構造変換と意味構造変換の長所短所、用例ベース翻訳や統計的翻訳などの経験的手法の有効性、対訳コーパスからの翻訳知識獲得の可能性といった議論は、言い換えにもほとんどそのまま当てはまる。実際、言い換え生成の実現方法に関するこれまでの提案や試みの多くは、こうした既存の機械翻訳技術からのアナロジーに基づいている。むしろ、処理方式などの方法論的議論に関するかぎり、言い換え生成の研究はこれまでのところ機械翻訳技術の後追いの域をほとんど出していないとさえいえる。

言い換えと翻訳の共通性をうまく切り取って形にした例として、Lavoie et al. (2000) の翻訳と言い換の統合モデル (図 2) は象徴的である。このモデルでは、MTT (2.1.2 を参照) に基づいて、まず入力文から深層の依存構造表現 (深層統語構造表現; DSyntS) を生成する。この

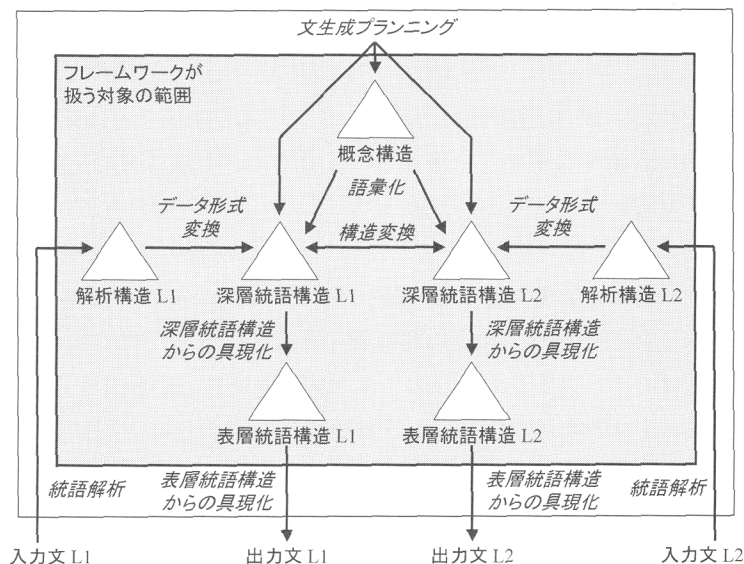


図 2 MTT に基づく翻訳と言い換えの統合モデル (Lavoie et al. 2000)

中間表現を他言語の中間表現に構造変換（トランスファ）してから生成すると翻訳になり，同言語内の別の中間表現に変換して生成すると言い換えになる．中間表現には言語に依存しない共通の表現形式を用いるので，言い換えにせよ翻訳にせよ同じエンジンを使って構造変換することができる．言い換えと翻訳の違いは，構造変換に用いる変換知識（変換パターン）が違うだけである．

4.1.2 構造変換（トランスファ）方式

既存の言い換えの実現方法の多くは機械翻訳のトランスファ方式に対応する．ただし，機械翻訳と異なって入力文の全体を変換する必要はないため，対象とする言い換えの種類ごとに，その現象を捉えるのに都合の良い変換レベルを設定して実現可能性を調査するというスタイルの研究が多い．たとえば，言い換えの対象が語や句のような比較的局所的な場合は，表層レベルでの局所的な置換によって言い換えを生成できる可能性があるが，埋め込み文を主節から切り離したり，主辞が交替するような言い換えを実現するためには，少なくとも依存構造や句構造などの統語レベルのトランスファが必要になる．

表層レベル 単語を同義語に言い換える場合や，慣用表現のような，要素が省略されにくく，語の間に別の語が割り込まない表現を言い換える場合，単純な文字列置換でも言い換えを生成できる．このレベルの言い換えには完全に語彙化された表現対が用いられる．実現例としては，単語から単語への置き換え (Edmonds 1999; 藤田, 乾 2001; Lapata

et al. 2001; Pearce 2001), 数語の単語列から同じく数語の単語列への言い換え (Barzilay and McKeown 2001; Barzilay and Lee 2002; Pang et al. 2003; 下畑ら 2003; Quirk et al. 2004), 慣用表現の言い換え (藤田, 乾 2003) などがあげられる。

統語レベル 言い換え生成モデルの中には依存構造を前提としているものも多い。統語レベルのトランスファでは、付録 A に示すようなさまざまな種類の言い換えが実現可能になる。たとえば, Kurohashi and Sakai (1999), 鍛治ら (2001) のモデルでは, 文節レベルの係り受け構造を用いている。そして, 「A の B」 \Leftrightarrow 連体節 (Kurohashi and Sakai 1999), 内容語の言い換え (鍛治ら 2003), 機能動詞結合の言い換え (鍛治, 黒橋 2004) などを実現している。一方, Takahashi et al. (2001) は, 言い換える対象が文節以上の単位であっても, その影響が文節よりも小さなレベルに及ぶことに着目し, 形態素レベルの依存構造を採用している。また, 機能語列や係り受けの順序などを表現するための記述言語を提供し, 依存構造レベルのトランスファ規則の作成を支援している。

意味レベル 2.1.2 で紹介した MTT の深層統語構造レベルでの言い換えは, 項の順序を区別して語彙関数を定義するなど, 依存構造や句構造から意味的なレベルに一步踏み込んだ構造変換を仮定している。その他, 動詞の格役割, 格要素間の関係を捉えるためにより意味に踏み込んだ例として, 語彙概念構造 (Lexical Conceptual Structure; LCS) を用いた複合動詞の言い換え (竹内ら 2002), 機能動詞結合の言い換え (Fujita et al. 2004) があげられる。

4.1.3 ピボット方式

機械翻訳のピボット方式相当のアプローチも見られる (Meteer and Shaked 1988; Huang and Fiedler 1996; Brun and Hagège 2003)。この方式では, 対象領域や捕らえるべき情報を限定して専用の意味構造を定義し, 対象テキストをその構造に当てはめることで言い換えを生成する。こうした研究では, 情報抽出の技術を用いて入力から意味構造の要素を抽出する技術や, 意味構造から表層表現を生成する技術が各々独立に論じられている。

たとえば, Brun and Hagège (2003) は, 薬品データベースの分析に基づいて 7 種類の述語-項構造を定義し, (15) の例のようにこの領域専用の意味表現として用いている。

- (15) a. Acetone is a manufactured chemical that is also found naturally in the environment.
 b. SUBSTANCE(acetone). PHYS_FORM(acetone,chemical).
 ORIGIN(acetone,natural,the environment,in).
 ORIGIN(acetone,man-made,NONE,NONE).

機械翻訳の場合と同様, ピボット方式では中間意味表現の設計と管理が問題になるため, 領域を十分に限定して, 過度に複雑な意味の問題をうまく避ける必要がある。

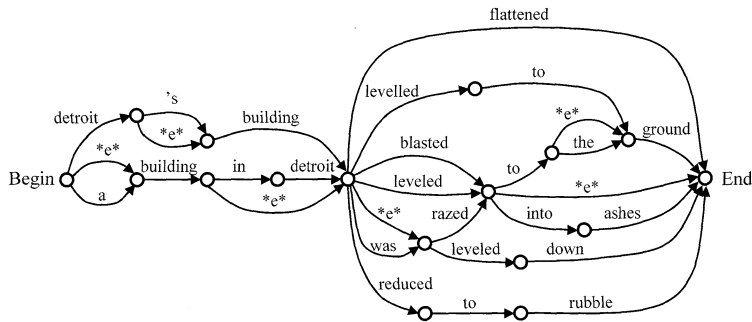


図 3 複数単語列アラインメントによって生成される単語ラティス (Pang et al. 2003)

4.1.4 言い換え知識の表現方法

言い換え知識の表現方法については、言い換え関係にある表現対を表層の単語列または構文木の対として表現する場合が多い。2.1.1 で紹介した変形文法の操作は統語レベルの言い換え知識とみなせるし、2.1.2 で紹介した MTT における言い換え規則も依存構造の対からなる。

文法を用いて言い換えを表現している例もある。Dras (1999) は、木接合による同期文法 (Synchronous Tree Adjoining Grammar; STAG) (Abeille et al. 1990; Shieber and Schabes 1990) を用いて言い換え知識を表現している。同期文法の枠組みでは、入力文の解析と同時に、解析に用いられる文法のそれぞれに対応 (同期) する文法が組み合わせさり、解析終了と同時に出力が得られる。

その他、言い換えの関係にある複数の文を、複数単語列アラインメント (Multi-Sequence Alignment; MSA) というアルゴリズムを用いて図 3 のような 1 つの単語ラティスで表現する試みもある (Barzilay and Lee 2002, 2003; Pang et al. 2003)。これは、英語のように語順の制約が比較的強い言語ならではのアプローチである。

4.2 機械翻訳と何が違うのか？

4.2.1 応用横断的技術としての言い換え

翻訳が異言語間の同義表現であるのに対し、言い換えは同一言語内の同義表現である。このことのおかげで、言い換え生成・認識技術は、単一言語を対象とするさまざまな言語処理アプリケーションへの利用が期待できる。別の言い方をすれば、言い換え技術は形態素・統語解析のような要素技術をさまざまな応用につながる応用横断的なミドルウェアであると考えてもよい。

3.2 の例 (13) をもう一度考えよう。機械翻訳の前編集で「お名前を頂戴することはできますか」を「お名前をお願いしたいのですが」に言い換えることが有益なのは、翻訳システムが前

者を正しく処理できず、後者を正しく処理できる場合である。しかし、この議論には少し誤魔化しがある。

この議論が成り立つためには、言い換えシステムが「お名前を頂戴することはできますか」という文を正しく解析し、正しく言い換える能力を持っていなければならない。しかし、上で述べたように翻訳と言い換え生成は本質的には同じ問題を扱う技術なので、もし「お名前を頂戴することはできますか」の言い換えが技術的に可能なのであれば、その技術を翻訳システムに組み込んで、両方の入力を正しく翻訳するシステムを作ることも原理的には可能なはずである。逆に、翻訳システムにとって解析が困難な文は言い換えもやはり困難なはずであり、翻訳の問題の一部を前編集に移したとしても、問題の難しさは変わらない。それでも翻訳の前編集に言い換え技術を使う試みが合理的に見えるのは、複雑な既存の翻訳システムの中身を触らずに済むといった短期的な利益のためばかりでなく、この技術が広く応用横断的だからである。翻訳という一つの応用技術から言い換えという応用横断技術を切り出す試みと言ってもよい。

このように、言い換え技術が応用横断的であることは、従来の機械翻訳研究では見過ごされてきた重要な特徴である。これまでの言い換えの研究では、特定の応用／言語／言い換えを想定して知識が構築されてきた。しかし、今後は、知識やシステムのポータビリティを考慮し、

- 言い換えのための知識をどのように整理し、分割し、記述しておけば応用横断的な再利用性が高くなるかを検討し、
- その成果にもとづいて実際に言い換える処理や知識を実現し、
- それらの部品を組み合わせる新しい用途に対応できる仕組みを作る

という努力を重ねる必要がある。

4.2.2 問題解決型タスクとしての言い換え

言い換え生成では、言い換えるべき対象を選択しなければならないという問題もある。

翻訳では、原文のすべての構成要素を目的言語に変換するという暗黙の前提があった (Wahlster 2000; Carl and Way 2003)。一方、言い換えは、「なんらかの目的を満たす表現への変換」(山本 2001) であるため、「原文のどの部分を言い換えるべきか」を目的に照らして判断し、その部分だけを選択的に変換する必要がでてくる。

たとえば、原文を平易な表現に言い換えて文章の読解を支援するといった用途の場合、原文のままでユーザが理解できる部分は言い換える必要がないし、むしろ言い換えない方がよい (Dras 1999; 乾 2001; Inui et al. 2003)。2 節で述べたように、言い換えは多くの場合原文の意味を厳密には保存できないため、不必要な言い換えは原文の情報を過度にねじまげてしまう恐れがあるからである。Dras は、原文を言い換えるたびに、話し手が伝えたい情報や微妙なニュアンスなど原文のなんらかの情報がかならず損われるため、人間の書いたテキストを言い換える際は目的を満たす範囲で言い換える程度を最小限に抑えるべきだと指摘している。そして、そ

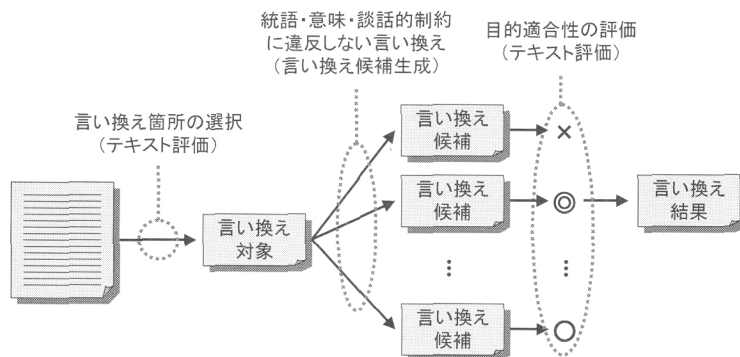


図 4 言い換え候補生成とテキスト評価

のような言い換えを「外部から与えられた制約を満たすために仕方なくやる言い換え」という意味で“reluctant paraphrasing”と呼んでいる。

これをもう少し一般化すると、言い換え生成は、「なんらかのテキストの評価基準が与えられたとき、原文から基準を満たさない言語表現を抽出し、満たす表現に言い換える」という問題解決型のタスクと見なせる。評価基準は言い換えの目的によって異なるだろう。読解支援の場合は「人間（外国人，子供，障害者，特定のユーザなど）にとってのテキストの読みやすさ」が基準になる (Carroll et al. 1998; Canning and Tait 1999; 乾 2001; Higashinaka and Nagao 2002; Inui et al. 2003) が、機械翻訳の前編集では「解析・翻訳の容易性」(白井ら 1995; 加藤ら 1997; 吉見ら 2000)、音声合成 (text-to-speech synthesis) の前編集では「聴覚理解の容易性」ということになる。また、特定の語彙と構文を基準として与えると制限言語への言い換え (Mitamura and Nyberg 2001) というタスクになり、書き言葉から話し言葉への変換 (Kaji et al. 2004) などの応用例がある。

以上から、言い換え生成技術は、次の 2 つの部分技術に集約できることがわかる。

- 言い換え候補生成 与えられた言語表現に対して、言語的に適格な種々の言い換えを網羅的に生成する技術
- テキスト評価 与えられた評価基準に基づいてテキストを評価する技術

言い換え候補生成が通常の意味での機械翻訳にほぼ相当するとすれば、テキスト評価は言い換えになって初めて顕在化される問題であるといえる。

テキスト評価の必要性は、すでに何人もの研究者が指摘するところである。山本 (2001) は、「対象特定」「仮説生成」「仮説選択」、Murata and Isahara (2001) は、「変換 (transformation)」「評価 (evaluation)」と呼んでいる。また、Mitamura and Nyberg (2001) も制限言語への言い換えを「checking」「rewriting」に分けている。これらはいずれもほぼ同じ分け方と見てよい。

両者の組み合わせ方にはいろいろ考えられる．最も単純には，次のような3段階のカスケード型のモデルが考えられる（図4）．

1. テキストを評価し，言い換えの対象を選択する
2. 選択された対象から可能な言い換えを網羅的に生成する
3. 生成された候補を評価し，最適解を出力する

もちろん，そのような単純な‘generate and test’方式に計算量的な問題がある場合は，評価基準に対する原文の「違反の仕方」に応じて言い換えの種類を絞り込むといった，いわば「言い換えプランニング」のような機構を検討してもよい．Dras (1999) が試みたように，制約下での最適化問題として定式化する方向も考えられる．言語生成のセンテンスプランニングを制約充足問題として定式化した Beale et al. (1998) のアプローチも参考になると思われる．また，読解支援などの場合，評価基準にユーザの読解能力の個人差（ユーザモデル）を反映することができれば，ユーザに適応的な支援も可能になるだろう．いずれにせよ，テキスト評価と言い換え候補生成を切り離した設計は，テキストの評価基準を取り替えることによってさまざまな用途の違いを吸収でき，より汎用的な枠組みを提供することができる点で有利である．

4.3 言い換えの認識

言い換えの認識は，質問応答や情報検索の中心的な部分問題の一つである（白木，黒橋 2000; 黒橋，酒井 2001; Lin and Pantel 2001; Ravichandran and Hovy 2002; Sasaki et al. 2002; Hermjakob et al. 2002; Duclaye et al. 2003; Moldovan et al. 2003; Takahashi et al. 2003; 高橋ら 2004）．高橋ら (2004) は，NTCIR QAC トラック (Fukumoto et al. 2002) で用いられた質問文と，人手で作成した質問の合計約 400 問とそれに対する解答文書の関係を分析している．それによると，質問と解答を結び付ける変換操作の約 85 % が含意・前提条件などの推論を含む広義の言い換えと見なせる．また，複数文書要約では，イベント間の関係を把握したり冗長な要約を避けたりするために，異なる文書中で同じ内容を指す部分（類似部分）を同定する必要がある．Barzilay et al. (1999) は，Topic Detection and Tracking コーパス (Allan et al. 1998) 中の同じ内容を示す文の対 200 組を分析し，語彙・構文的言い換えによって約 85 % の文対を結び付けることができると述べている．この文脈でも，言い換えの認識に関するいくつかの手法が提案されている (McKeown et al. 1999; Barzilay et al. 1999; 上田，小山 2000; 成松ら 2002; Barzilay 2003)．

言い換えの認識のアプローチは大きく 2 種類に分けられる．1 つ目は，語彙・構文的変換の到達可能性を調べるアプローチで，与えられた 2 つの言語表現のうち，一方を語彙・構文的に言い換えて，他方に到達できるか否かを判別する．2 つ目は，意味レベルの照合を明示的に扱うアプローチで，2 つの言語表現の各々をピボットの意味表現に変換し，それらが一致するか否かを判別する．以下，それぞれの代表的な研究を紹介する．

4.3.1 語彙・構文の変換に基づく言い換えの認識

語彙・構文の変換に基づく方法の例は上田, 小山 (2000) の複数文書要約アルゴリズムに見ることができる. 彼らの方法では, 複数の入力文書を依存構造の部分木の集合として表現し, その中から入力文書に共通に出現する部分木を取り出すことで複数文書要約を生成する. ただし, 文書によっては同じ情報が別の表現で言語化されている可能性がある. 同義な (すなわち言い換えの関係にある) 部分木どうしても共通の部分木として扱う必要がある. 彼らが扱った言い換えは, (16) のような類義語, 上位語への言い換え, および (17), (18) のような構文的な交替などである.

- (16) s1. ホウレンソウからダイオキシンが検出された.
s2. 白菜からダイオキシンが検出された.
t. 野菜からダイオキシンが検出された.
- (17) s. 軽量の携帯電話がフーバー社によって発売される.
t. フーバー社が軽量の携帯電話を発売する.
- (18) s. 全角スペースがシンタックスエラーを起こす.
t. 全角スペースでシンタックスエラーが起きる.

彼らのアプローチは, McKeown et al. (1999), Barzilay et al. (1999) の手法と次の点で共通する.

- 従来の bag-of-words 的な手法に替えて部分構文構造を導入し, 類似度をより正確に見積もる (McKeown et al. 1999),
- WordNet (Miller et al. 1990) の synset や動詞のクラス (Levin 1993) を用いて, 部分構造間の同義性を判定する (McKeown et al. 1999; Barzilay et al. 1999),
- 各種交替などの構文的な言い換えを規則として実装し, 表現の多様性を吸収する (Barzilay et al. 1999).

上田, 小山は, 語彙・構文の変換によって生成された部分木にペナルティを課すことで, 変換によって生じる元文との情報のずれを考慮している. そして, 変換によるペナルティと文書間での共通性から各部分木の重要度を計算し, 上位数個の部分木を要約生成に用いている. このアルゴリズムは, 入力となる n 文の各々について可能なすべての言い換えを生成することで, 複数の文書間で共通に出現し, かつ言い換え回数が少ない表現を効率良く選択している.

こうした語彙・構文の変換に基づく方式では, 入力表現の言い換え方が組み合わせ的な数に膨らむ可能性がある. 変換の種類に制限を加える, あるいは効率的な探索法を導入するといったなんらかの対策が必要になる. たとえば, Takahashi et al. (2003) は, 質問応答の文脈で, 山登り法探索を実現する枠組みを提案している. 彼らのアルゴリズムでは, 質問と解答候補文書の両方に言い換えを適用し, 最も類似度が高い (質問, 解答候補) の組を優先して繰り返し言い換える. また, 上田, 小山 のように, 語彙・構文の変換にペナルティを課す場合はそれによって探索空間が抑えられる.

4.3.2 意味表現に基づく言い換えの認識

ある表現の含意や前提条件などの推論は、統語構造上よりも意味表現上で扱う方が、知識記述や変形操作の実現という点で都合が良い。しかし、意味表現に基づいて言い換えを認識するには、やはり意味表現の設計が問題になる。ここでは、質問応答の文脈でのアプローチをいくつか紹介する。

Ravichandran and Hovy (2002) は、質問応答では多くの場合、質問に対して特有の表現パターンが解答になるということに着目している。たとえば、(19a) と (19b) は “Mozart” の誕生年を示す異なる表現だが、(20a), (20b) のように固有表現を抽象化すると、どちらも「ある人物の誕生年」に関する (21) のような意味表現に対する表現のパターンと見なすことができる。

- (19) a. Mozart was born in 1756.
- b. Mozart (1756-1791) ...
- (20) a. $\langle name \rangle$ was born in $\langle birthdate \rangle$
- b. $\langle name \rangle$ ($\langle birthdate \rangle$) -
- (21) BIRTHDATE (name, birthdate)

彼らは、TREC 2001 (Voorhees 2001) における質問の分析によって上の誕生年に関する質問を含む 6 つの質問タイプを選択し、それぞれを表すような典型的な表現パターンを Web から自動収集している。質問に対する解答としては、単純に上のような表現パターン中の、質問対象の固有表現のスロットに対応する表現が出力される。たとえば、“When was Mozart born?” という質問は、(21) の意味表現に対応付けられ、同時に $\langle birthdate \rangle$ が解答を示す固有表現スロットだと同定される。質問文中の固有表現 “Mozart” を用いて検索した文書中に、(19a) のような、(20) 中の表現パターンに対応する表現が見つければ、その文書と質問は (21) という意味表現において等価だと認識され、“1756” が解答として取り出される。

Sasaki et al. (2002), Moldovan et al. (2003) は、質問と解答候補文書を論理形式 (Logical Form) を意味表現とすることで、言い換えの関係にある文間の統語レベルの違いを捨象し、質問と解答候補文書の対応付けを可能にしている。(22LF) は、Sasaki et al. の質問応答システム、SAIQA-Is に (22q) という質問文を入力したときに得られる論理形式である。

- (22) q. Where is the capical city of Japan?
- LF. COUNTRY(Y1:'Japan'), R(Y1,Y2), 'city'(Y2:'capital city'),
- LOCATION(Y3:Z); ORGANIZATION(Y3:Z), R(Y2,Y3)

SAIQA-Is では、さらに、解答が得られなかったときのみ “World Cup” \Rightarrow “W-Cup” のような同義語レベルの言い換えを適用し、解答を再度探索する。一方、Moldovan et al. のシステム、Logic Prover は、WordNet (Miller et al. 1990) を用いて論理形式中の一部の語を同義語・上位語に置き換えるだけでなく、論理形式レベルで推論に関する書き換え処理を施して、質問と解答候補文書の照合を試みる。

5 言い換え知識の獲得

言い換えの生成や認識を実用規模で実現するには、言い換えに関する知識を既存の資源から効率的に獲得する手段の開発が必須である。本節では、言い換え知識の獲得に関するこれまでの試みを翻訳や情報抽出のための知識獲得技術に照らして紹介する。

5.1 既存の語彙資源を言い換えに利用する

5.1.1 シソーラスを使って同概念語に言い換える

言い換えに利用できる語彙資源と言うと、まず思い浮ぶのはシソーラスである。たとえば、WordNet (Miller et al. 1990)² や EDR 日本語単語辞書 (日本電子化辞書研究所 1995)³ には、非常に細かい意味分類に基づく単語間の同義関係が与えられているので、それを用いれば、入力中の単語を同義語に置換する語彙的言い換えを実現できるように思える。しかし、実際には同義語といえども意味や用法になんらかの差がある場合がほとんどで、無条件で置換できる語のペアは必ずしも多くない (Edmonds 1999; 藤田, 乾 2001; Lapata et al. 2001; Pearce 2001; Okamoto et al. 2003; Inkpen 2003)。たとえば、「随所」と「各地」は EDR 日本語単語辞書によると同概念に属する (同概念語) が厳密には意味が異なる。このため、互いに言い換え可能かどうかは、(23) のように、その差が周囲の文脈に照らして無視できるかどうか依存する。

- (23) a. 随所 (⇒ 各地) でがれきの山が生まれ、火災も発生し、死傷者も多数、確認されている。
b. 片仮名交じりの文語体、しかも難解な言葉が随所 (⇒*各地) にあり、法学専攻の学生をすら悩ます現行刑法の法文が現代用語に書き換えられる。

5.1.2 語釈文に言い換える

国語辞典の語釈文は見出し語の言い換え表現と見なせるので、国語辞典から〈見出し語、語釈文〉の対を取り出せば、そのまま大規模な語彙的言い換え知識として使える。たとえば、「廢材」は「いらなくなった木材 (岩波国語辞典)」という語釈文を持つので、(24) のような言い換えができる。

- (24) s. がれきや廢材の仮置き場
t. がれきやいらなくなった木材の仮置き場

しかし、いつも語釈文に置き換えるだけで正しい言い換えが作れるわけではない。たとえば、次の例 (25s) の「相乗り (する)」を語釈文「乗り物と一緒に乗る」にそのまま置き換えようとすると、(25t1) のような不適格な文になってしまう。正しくは、原文中の「タクシーに」と語釈

² <http://www.cogsci.princeton.edu/~wn/>

³ <http://www2.crl.go.jp/kk/e416/EDR/J.index.html>

文中の「乗り物に」の重複を検出して、(25t2)のように「乗り物に」を削除する必要がある。

(25) s. タクシーに相乗りする

t1. *タクシーに 乗り物に 一緒に乗る

t2. タクシーに一緒に乗る

語釈文への言い換えでおこる上の問題に注目した黒橋, 酒井 (2001), 鍛冶ら (2001) は, 言い換え対象語の周囲の文脈と語釈文の要素と重なり (上の例では「タクシー」と「乗り物」) を自動的に検出し, 重複をうまく取り除いた適格な言い換えを生成する手法を提案している. 彼らのアプローチは, (a) 国語辞典という既存の語彙資源を使うため, カバレッジの広い多様な語について語彙的言い換えを実現できる, (b) 自然言語で書かれた語釈文を知識源とするので, 知識の拡張・保守が容易であるなどの利点があり, 大きな可能性を秘めている.

5.1.3 語釈文から言い換えを見つける

さらに, 慣用表現など, 内容語の特別な用法について, 語釈文にヒントが隠されている場合がある. たとえば, 岩波国語辞典 (RWC 1998) の「はこぶ」の語釈文には, (26) のように, 慣用表現とそれに対応する表現の対が記述されている. ただし, こうした記述は網羅的なものではないため, この方法で十分なカバレッジを確保することは難しい. また, 運用の際には多義性も考慮する必要がある.

(26) はこ-ぶ【運ぶ】〈1〉((五他)) 何かのために, ものを他の所に進め移す.〈ア〉物を持ちたり車に積んだりして, 他の場所まで動かす.「机を別の部屋に―」「恋人の所へせつせと金を― (= みつぐ)」「筆を―」(文章を書き進める)「足を―」(行く. 通う)「ようこそおー・び (= おいで) くださいました」

5.1.4 対訳辞書から言い換えを見つける

対訳辞書を利用するという手も考えられる. たとえば, 日本語語彙大系 (池原ら 1997) の構文体系には (27) のような記述があり, そこから (28r) のような言い換え知識を獲得することができる.

(27) a. N1(名詞のクラス:主体) が N2(名詞のクラス:主体) の軍門に下る

⇔ N1 surrender to N2

b. N1(名詞のクラス:主体) が N2(名詞のクラス:主体) に降伏する

⇔ N1 surrender to N2

(28) r. N1(名詞のクラス:主体) が N2(名詞のクラス:主体) の軍門に下る

⇒ N1 が N2 に降伏する.

s. 英国を含む欧州がヒトラーの軍門に下るのを黙って見ているわけにはいかない.

t. 英国を含む欧州がヒトラーに降伏するのを黙って見ているわけにはいかない.

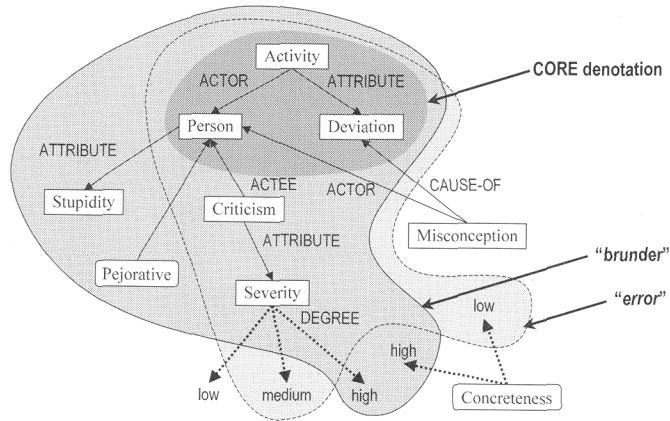


図 5 語の指示的意味と言外の意味を示すクラスモデル (Edmonds 1999, p.97)

5.1.5 意味の同一性を考えて言い換えを獲得する

より厳密に同義表現を獲得するためには、言い換え前後の表現対の共通の意味や意味の差を捉え、2.2 で述べたさまざまなレベルにおける同一性の問題に踏み込む必要がある。ここでは、最もプリミティブなレベル、すなわち類義語間の意味の差と捉えようとする試みをいくつか紹介する。

Edmonds (1999) は、類義語 (near-synonym) の意味を記述するオントロジを開発し、自然言語生成の語選択に用いている。図 5 では、類義語 “brunder” と “error” の意味が、それらを含む語のクラスの指示的意味、および一部の言外の意味のリンクによって示されている。この図からは、この 2 語が、(i) 非難の激しさの程度 (criticism → severity)、(ii) 誤りがばかげているか否か (stupidity)、(iii) 具体性 (concreteness)、(iv) 軽蔑的か否か (pejorative)、という 4 つの点で異なっていることが分かる。語の意味を記述する意味素の粒度について、Edmonds は、複数の言語を対象として上のような異なりを表現できているためある程度妥当であると評価している。

意味記述に替わる語彙知識のリソースとして国語辞典の語釈文を用いた研究がいくつかある。土屋、黒橋 (2000) は、国語辞典中の各語の語釈文を統語解析器を用いてグラフに変換し、MDS 原理に基づいて、辞書全体にわたる部分グラフの抽象化および辞書の圧縮を施している。結果として得られる辞書からは、任意の 2 語 w_1 と w_2 の共通の意味と個別の付加の意味を容易に取り出すことができる。藤田、乾 (2001) は、土屋、黒橋の手法を各々の w_1 と w_2 の対に対して適用し、語釈文間の重なり的大小さに基づいて言い換えの適格性を判定している。(23) の例では、「随所」の語釈が「限定されないどの場所にも。方々。」、「各所」の語釈が「ある範囲内のところどころ。」(いずれも角川類語新辞典 (大野、浜西 1981)) と完全に異なる。藤田、乾

の手法では、語釈文の差分と文脈における制約を独立にしか捉えていないため、文脈に関わらず常に（例 (23a) の言い換えも）適格でないと判断されるという問題がある。Okamoto et al. (2003) は、ある語が持つ指示的意味とその語が文脈中の他の語句の選択に与える制約（語彙的制約）を語釈文から取り出している。具体的には、語釈文中のすべての内容語を、その意味クラスとコーパス中の共起頻度を用いて指示的意味、語彙的制約に分類している。彼らの手法では、語釈文の比較などを要せずに単語そのものを表現できるため、Edmonds (1999) のモデルにおける知識獲得につながる可能性がある。しかし、藤田, 乾 (2001), Okamoto et al. (2003) の実験結果を見るかぎり、語釈文の情報を用いて同概念語間の可換性を判断するためには、かなり深い言語理解を必要とするように見える場合も多く、越えるべきハードルは高い。

言外の意味（フォーマリティや親密度など）をいかにして獲得するか、という課題もある。この課題に対する試みとして Inkpen (2003) の研究を紹介する。彼女はまず、Edmonds (1999) のオントロジに基づいて、類義語の意味を表現するための知識を、指示的意味 (denotation), 姿勢・態度 (attitude), スタイル (style) の 3 クラスの知識に分類・形式化している。それぞれの知識は次のような組で表現される。

- 指示的意味 〈語, 頻度 (sometimes, usually, always), 強さ (low, medium, high), 指示の
間接性 (SUGGESTION, IMPLICATION, DENOTATION), 周辺概念 (不定形)〉
姿勢・態度 〈語, 頻度, 強さ, 姿勢・態度の種類 (FAVORABLE, NEUTRAL, PEJORATIVE)〉
スタイル 〈語, 強さ, スタイルの種類 (FORMALITY, CONCRETENESS, FLORIDITY など)〉

図 5 で示されている言外の意味のうち、軽蔑的か否か (pejorative) は姿勢・態度の、具体性 (concreteness) はスタイルの下位クラスとして定義されており、それぞれ種類の項の値となる。Inkpen は次に、これらの知識を類義語の使い分け辞典から抽出する手法を提案している。この手法では、たとえば、(29a) の文章から (29b) に示す 3 つの語彙知識を獲得できる。(29b) の 1 つ目は動詞 “absorb” のスタイルに関する知識であり、残りは指示的意味に関する知識である。指示的意味における周辺概念だけは不定形であり、(29a) の文章中の句で表現される。

- (29) a. **Absorb** is slightly more informal than the others and has, perhaps, the widest range of uses. In its most restricted sense it suggests the taking in or soaking up specifically of liquids: the liquid *absorbed* by the sponge. In more general uses *absorb* may imply the thoroughness of the action: not merely to read the chapter, but to *absorb* its meaning.
- b. 〈absorb, low, FORMALITY〉
〈absorb, usually, medium, SUGGESTION, the taking in of liquids〉
〈absorb, sometimes, medium, IMPLICATION, the thoroughness of the action〉

5.2 パラレルコーパスから言い換え知識を獲得する

機械翻訳では、対訳コーパスから翻訳知識を自動獲得する試みが多数報告されており (Meyers et al. 1998; Watanabe et al. 2000; Melamed 2001; 山本, 松本 2001; 今村 2002), 大規模なパラレルコーパスまたはコンパラブルコーパスがあれば, そこから翻訳知識を獲得できることがわかっていて, 一方, 言い換えの場合, 大量の言い換え事例の入手は翻訳の場合ほど容易でない. 日本語の新聞記事や書籍, ホームページが英語や他の言語に翻訳されることはあっても, わざわざ「外国人日本語学習者にもわかる日本語」や「朗読して聴きとりやすい日本語」に言い換えられることはほとんどない. パラレル/コンパラブルコーパスを収集するためになんらかの工夫をするか, パラレルでないコーパスからの知識獲得を考える必要がでてくる.

5.2.1 同じ原文に対する複数の翻訳文を集める

同じ原文に対して複数の翻訳がある場合, それらは言い換えと見なすことができる. 機械翻訳では, システムの評価方法として, 1つの原文に対して例 (30) のような複数の正解翻訳例を用意するのが一般的になってきており, そうした複数の翻訳例を含む対訳コーパスもいくつか整備されつつある (Shirai et al. 2001; Shimohata et al. 2004).

(30) J0 競技場は大勢の観客で膨れ上がった.

J1 競技場は大勢の観客で身動きができなかった.

E0 The athletic field was swamped with spectators.

(Shirai et al. 2001) では, 既存の対訳コーパスに対して多様な別訳を作る作業をどうやってうまく制御し効率化するかといった問題も検討されており, 今後も研究者間で共有できる資源が増えるものと思われる.

対訳コーパスからの翻訳知識獲得では, 一対一の翻訳対を対象に句や節の対応を計算するという問題が一般的であった. 一方, 言い換えの場合は, 互いに言い換え関係にある複数の表現を含む集合を用意することができるので, 一般には3つ以上の要素間のアラインメントをとるという新しい問題が出てくる. もっとも単純なアプローチは, 集合内の各要素対ごとにアラインメントをとる方法である. たとえば, 今村ら (2001) は, 集合内の各言い換え対について構文木に基づく階層的アラインメントによって句や節レベルの言い換え対を獲得する方法を提案している. これに対し, Pang et al. (2003) の方法では, 構文木に基づく複数単語列アラインメントによって3つ以上の言い換え間の対応関係を1つの単語ラティスで表現する (4.1.4 を参照).

一方, 既存の翻訳を集めてくるという手もある. Barzilay and McKeown (2001) は, 『海底二万里⁴』のように同じ原著から何冊もの訳本がでている作品があることに着目し, そうした訳本から言い換え事例を大量に獲得しようと試みている. 彼女らによると, 複数の翻訳本から得られるパラレルコーパスはノイズが多く, また従来扱ってきた対訳コーパスに比べるときれい

4 Jules Verne (1869). *Vingt mille lieues sous les mers* / *Two Thousand Leagues Under the Sea*.

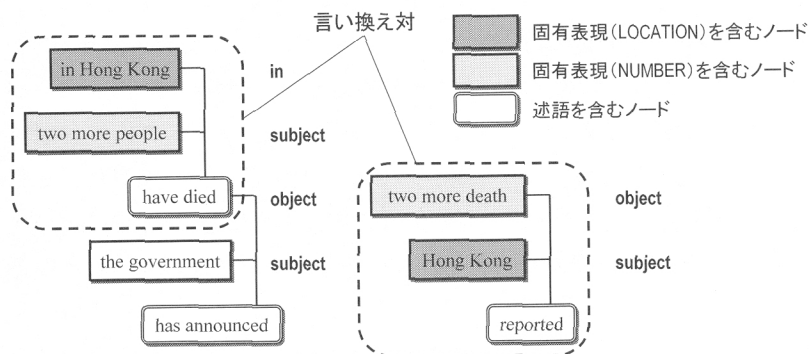


図 6 コンパラブルな文の対からの言い換え対の抽出 (Shinyama and Sekine 2003)

に言い換えの対応がとれる箇所は必ずしも多くない。さらに、獲得できた事例は *King's son* ⇔ *son of the king* や *countless* ⇔ *lots of* のような局所的な語句の言い換えが多く、5.1 で述べたような既存の資源から得られそうなものも少なくない。一方、Ohtake and Yamamoto (2003) が同様の実験を旅行対話に関する対訳コーパスで行ったところ、領域に特化した (31) のような言い換えが多数獲得できた。

- (31) s. それ以上は安くなりませんか。
t. それが最終的な値段ですか。

機械翻訳の場合と同じように、パラレルコーパス→アラインメント→言い換え知識の獲得、というシナリオが現実的に描けるか、難しいとすればどのような工夫が必要かなど、興味深い問題が課題として残されている。

5.2.2 同じ物事に対する複数の説明文を集める

同じ事件を報道している複数の違った新聞社の記事をコンパラブルコーパスと見なせば、そこから言い換え表現を発見できる可能性がある。厳密なパラレルコーパスと違って、記事は日々生産されるので大規模なコーパスを入手できるという利点がある。ただし、各記事がまったく同じ情報を同じ順序で過不足なく伝えている保証はないため、文単位、句単位の順番で厳密にアラインメントをとるという従来の翻訳知識獲得の方法を単純に適用するわけにはいかない。

この問題に対し、Shinyama et al. (2002), Shinyama and Sekine (2003) の方法では、まず記事対応をとった後、出現単語の類似度に基づいて文対応を同定する。次に、句単位のアラインメントをとる代わりに、次の条件をより良く満たす依存構造の部分木の対だけを言い換え対として獲得する (図 6)。

- 各部分依存構造木の根は用言である。
- 対となる部分依存構造木が共通の固有表現を含んでいる。

(c) 各用言が要求する格が部分依存構造木に過不足なく含まれている。

一方, Barzilay and Lee (2003) は, まずコーパスに含まれる各文を単語 n-gram に基づく類似度でクラスタリングし, 各クラスタ内に含まれる類似文から複数単語列アラインメントによって単語ラティス (図 3) を生成する. 単語ラティスを見れば, クラスタ内の各類似文のどの箇所が共通でどの箇所が文ごとに異なるかが分かるので, 共通部分を定型表現, それ以外を変数とする定型パターンが作れる. この方法を各新聞社の記事集合に適用し, Shinyama and Sekine (2003) と同様の方法で記事集合間の対応をとれば, 定型パターン間の言い換え関係を同定できる可能性がある.

句単位のアライメントがとれない場合に問題となるのは, 与えられた文の対のどの部分を言い換え対として抽出すれば良いかの判断が難しいことである. たとえば, 図 6 の例で言うと, 右側の木に対応する言い換えは “have died” を根とする左側の木の一部であって, “has announced” を根とする木全体でない. しかし, そうだと判断するに足る手がかりは, この文の対を見ているだけでは得られない. このとき, 有用な手がかりとなるのは, 言い換え関係に立つ表現の「表現らしさ」である. Shinyama and Sekine (2003) は, 上の条件 (c) を追加することによってこの「表現らしさ」を考慮しようとしている. また, Barzilay and Lee (2003) が複数単語列アラインメントによって定型パターンを事前に収集したのも, そのねらいは同じである.

5.3 パラレルでないコーパスを使う

5.3.1 文脈の類似性を測る

パラレルでないコーパス (ノンパラレルコーパス) から同義表現を獲得する場合に基本となるのは,

与えられた入力表現と (a) 似た文脈で出現する表現, あるいは (b) 内部構造が似ている表現がコーパス中に存在すれば, それは入力の言い換えである可能性が高い

という仮定である. とくに, (a) の出現文脈の類似性に基づいて推定される言語表現の類似度は分布類似度 (distributional similarity) と呼ばれ, 単語間の類似度の推定に効果的であることが知られている (Pereira et al. 1993; Lin 1998). 言い換えの獲得は同義語の獲得を一般化した問題と見なせるので, 単語間の分布類似度の推定方法をうまく拡張すれば, より多様な構造の言い換えを獲得できる可能性がある.

代表的なのは Lin and Pantel (2001) の手法である. DIRT と呼ばれる彼らのアルゴリズムは, (Lin 1998) で提案した単語間分布類似度の推定方法を一般化したもので, 図 7 のような依存構造の部分木間の類似度を推定する. ここで対象とする部分木は, 両端を名詞の変数スロットとする枝分かれなしのパスである. 図の例のように, 2つのパスの両端のスロット X , Y に現れる単語の分布が互いに十分に似ていれば, それらのパスは言い換えと同定される.

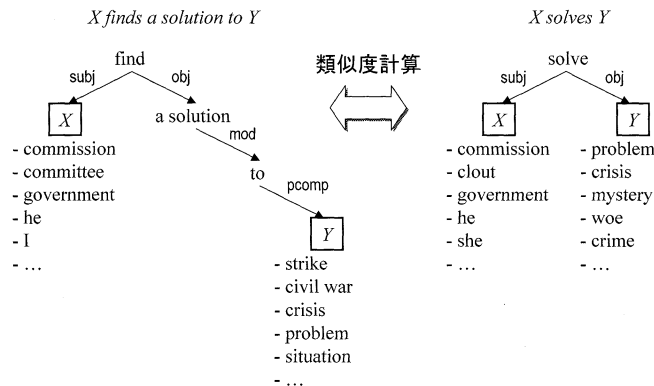


図 7 分布類似度に基づく部分依存構造木間の同義性の判定 (Lin and Pantel 2001)

また、鳥澤 (2002) の手法は、「アメリカの車」のような入力に対して「アメリカで生産する車」のような言い換えをコーパスから獲得する。二つの名詞（「アメリカ」と「車」）を出現文脈とし、与えられた文脈と確率的に良く共起する表現を選択する点は同じで、違うのは、共起の強さを測る方法と獲得の対象を動詞格構造（「で生産する」）に限定している点である。また、(Torisawa 2002) では、同様の方法が (近藤ら 1999) と同様の動詞格構造間の言い換えにも適用できることが報告されている。

コンパラブルコーパスに比べると、ノンパラレルコーパスははるかに容易に入手できるので、これと分布類似度の組み合わせは良い解決策であるように見える。ただし、分布類似度にも問題がある。まず、分布類似度を推定するには参照する文脈のスコープを固定する必要があるため、予め固定したパターンの言い換えしか獲得できないという制限がつきまとう。たとえば、Lin and Pantel (2001) の手法では、両端を名詞の変数スロットとする枝分かれなしのパスに対象が限定されており、3つ以上の変数スロットを持つ部分構文木を同時に扱うことはできない。分布類似度に基づく方法にはもう一つ、文脈の分布の偏りが大きい表現の言い換えしか獲得できないという欠点もある。たとえば、「X が Y を告訴する」の言い換えを同様の方法で獲得しようとしても、スロット X, Y に出現する人間や組織の間でなされる行為は「告訴」だけではないので、これだけの情報で正しい言い換えを選別するのは困難である。

5.3.2 内部構造の類似性を測る

内部構造の類似に基づく方法には次のような例がある。木村ら (2001), Tokunaga et al. (2003) の手法では、同義表現を探し出す手段として、漢字インデックスによる情報検索を利用する。同じ漢字をより多く共有する2つ名詞句は意味が似ている可能性が高い。漢字をインデックスとすることによって多様な表現の間の類似性が計算できるので、次のように単語の置換だ

けでは抽出できない言い換えも生成できる。

(32) s. 収益の減少

t. 減収減益

(33) s. 倍額の増資

t. 出資額倍増

また, Terada and Tokunaga (2001) は, 略語をもとの単語に復元する (34) のような言い換えをとりあげ, 言い換えの候補を文脈に応じて候補選択するモデルを略語の多いコーパスと略語の少ないコーパスから獲得する手法を提案している。

(34) s. TWR

t. tower / toward

彼らの手法では, 文字ベースの類似性と単語が出現する文脈の類似性の両方を考慮して, 略語と同義な単語をコーパス中から探す。入力と「似ている」表現をコーパスから探し出すという点で, やはり上述のアプローチと同様の方向性を持っている。

Jacquemin et al. (1997), Yoshikane et al. (2003) は, 文書検索の文脈で, 索引語 (ここではとくに複合専門用語, multi-word term; MWT) の表現の多様性を吸収する言い換え規則を単言語コーパスから発見する手法を提案している。

(35) a. 構文的変形: technique for performing volumetric measurements

⇒ measurement technique

b. 形態的変形: electrophoresed on a neutral polyacrylamide gel ⇒ gel electrophoresis

彼らの手法では, まず, 専門用語辞書中の複合専門用語 (以下, 単に複合語) を種にして, その言い換えをコーパスから抽出する。ここでは, 所与の複合語を構成する内容語がコーパス中で有意に共起しているパターンを, この複合語の言い換えパターンとして取り出す。次に, さまざまな複合語について得られた言い換えパターンを人手で類型化し, 大きく 6 種類 (Yoshikane et al. は 7 種類) の言い換え規則集合を作成している。両研究とも, 作成した言い換え規則集合によって生成される複合語の言い換えがどれだけ正しいか, 情報検索の精度向上にどれだけ寄与するかの 2 段階で評価している。

5.4 言い換えの適格性を判定するための知識

言い換えた後の表現が言語的に適格か否かを判定する必要がある。言い換え生成はテキストの一部に関する操作であるため, 適格性の判定も, 文や文章全体の良さではなく, その操作を受けた部分と文脈がうまくあうかどうかだけを評価すればよさそうに思える。ただし, 言い換えの言語的適格性に関わる要因には, 形態素・構文レベルから意味レベル, 談話レベルまで性質が異なるさまざまなものがあり, 言い換えの種類によって共通性は見られるものの, その傾向は異なっている (藤田, 乾 2003)。これらをまとめて捉えるようなモデルは現状では存在しな

いので、それぞれの適格性に関わる要因を個別に捉えて整理・モデル化し、うまく融合させる必要がある。

自然言語生成の分野では、近年、出力テキストの候補を複数生成し、最後にランキングして候補を1つに絞る方式が有力になってきた (Knight and Hatzivassiloglou 1995; Langkilde and Knight 1998; Bangalore and Ranbow 2000). これにならったランキング方式のモデルの一例として、単語の共起の是非を判定する研究 (Pearce 2001; Lapata et al. 2001; 藤田ら 2004) を取り上げる. 例 (23b) では、「言葉が（法文の）各地にある」という表現について、「法文の」と「各地」が共起しない（修飾関係にならない）ことがわかれば、適格ではないと判定できる. しかし、「随所」とは共起するが、同概念語の「各地」とは共起しないといった粒度の細かい共起制約を必要とするということは、「意味クラスに基づく共起データの抽象化」という常套手段が通用しないことを意味するので、問題は見た目ほど単純ではない. この共起の是非を判定するためには、厳密には個々の語に関する詳細な知識や共起に関する知識が必要になるが、個々の単語を区別する統計モデルを洗練するだけでも比較的良好な成果をあげている. Pearce (2001), Lapata et al. (2001) は、WordNet (Miller et al. 1990) を用いて名詞を修飾する形容詞を言い換えたときの単語の共起の是非を判定し、単語間の分布類似度が人間の判断と相関を持っていることを示した. 一方、藤田ら (2004) は、態の交替や内容語の言い換えなど、さまざまな言い換えの際に頻繁に不適格になる、動詞とその名詞格要素の共起を対象としている. 藤田らは、大規模な生コーパスから得られる共起用例の分布クラスタリング (Pereira et al. 1993) によって単語間の潜在的な類似性を考慮するとともに、人手で収集した不適格な共起用例（負例）を組み合わせ、共起の適格性を判定するモデルを構築している.

その他、言い換えの適格性判定の研究としては、談話構造や結束性を対象とした研究 (Inui and Nogami 2001; 野上, 乾 2002; Siddharthan 2003) がある. さらに、言語的適格性を超えたレベルでも評価が必要になる場合がある. 言語生成の分野では、ユーザの知識に合わせて表現を変える (Cawsey 1992). 語用論的效果を考慮する (Hovy 1988). 丁寧さや性別などを考慮する (金子ら 1996; 内元ら 1996) など、さまざまな試みが報告されている (乾 1999).

6 おわりに

本論文では、近年研究者間で関心が高まってきた言い換え技術について、最近の研究動向を紹介した.

言い換え技術は言い換え生成と言い換え認識に大きく分けて考えることができる. 言い換え生成は、ある言語表現を意味を保存しながら別の言語表現に変換する作業であり、機械翻訳の前編集や読解支援のための文章簡単化など、さまざまな応用に利用できる. 変換の方式、曖昧性解消、言語生成、知識の表現方法と自動獲得など、個別の部分問題に関する限り、言い換え生成に必要な技術は機械翻訳技術と重なるところが大きい. 一方、言い換え認識は、2つの異

なる言語表現が言い換えかどうかを判別する作業であり、情報検索や質問応答、複数文書要約などの応用がある。この問題は、一方から言い換えを生成して他方に到達できるかを調べる問題と見なせるので、言い換え生成と裏表の関係にある。こうした技術は、これまで、それぞれの応用で別々に必要性が論じられ、個別に研究されてきた。しかし、それらの間には必要な技術や蓄積すべき言語知識に共通する部分も多い。今後はこれら個別の試みを統合し、応用横断的なミドルウェア技術に発展させていくことが重要である。

最大の技術的関心は知識獲得である。言い換えのパラレルコーパスは、翻訳の場合よりさらに入手が難しいので、大規模なコンパラブルコーパスを収集するためになんらかの新しい工夫をするか、ノンパラレルコーパスから言い換え知識を獲得する方法を考える必要がある。コンパラブルコーパスの収集については、同じ事件を報道した複数の新聞記事を集める方法や、同じ原著に対する複数の翻訳を集める方法を紹介したが、ノイズの多いコンパラブルコーパス上で高精度な知識獲得を実現するには克服すべき課題も多い。一方、ノンパラレルコーパスからの知識獲得も、分布類似度に代表される従来の単語間類似度の推定手法を拡張・一般化する方向で発展してきたが、実際の応用に耐える実用規模の知識獲得に至った例はほとんどない。

目指すべき方向性の一つは、コーパスから獲得した言い換え知識と既知の言い換え知識を組み合わせる知識の洗練をはかるアプローチであろう。これまでに報告された知識獲得の研究は、同義語の知識以外は既知の言い換え知識を仮定せず、スクラッチから言い換え知識を獲得しようとする試みがほとんどである。しかし、2.1 で見た言語学的研究が示すように、我々はかなり多くの言い換えをすでに知っている。また、麻岡ら (2004), Fujita et al. (2004) の試みに見られるように、規則化がある程度可能なタイプの言い換えも少なくない (付録 A)。安易にすべてを知識獲得技術に頼るのではなく、まずはこうした人手による管理が可能な言い換え知識を収集・蓄積し、コーパスからの知識獲得に積極的に活用する姿勢が必要のように思われる。そのためにはまず、言い換えにはどのような種類があるのか、どのタイプの言い換えの知識を人手で書くのが合理的なのか、自動獲得に頼るべき言い換え知識はどのようなものかを明らかにしていく必要がある。

言い換え技術をさらに発展させるためには、多様な言い換え現象を包括的に調査・分類し、それに基づいて言い換えコーパスや言い換え知識などの言語資源を共有可能な形に設計・蓄積する努力が必要である。しかし、言い換えが言語表現の同義性に立脚する概念である以上、これに厳密な定義や分類を与えるには、意味とは何か、意味が同じとはどういうことかといった深い意味論の問題に立ち入らなければならない。本論文では、語彙・構文的言い換え、参照的言い換え、語用論的言い換えを混同すべきでないことに言及したが、これは議論の出発点の一つに過ぎない。こうした意味に深く根ざす問題にどうやって工学的にアプローチするか、統計的言語処理技術がある程度成熟を見た今、再度じっくり議論してよいテーマである。言い換え技術が現在の言語処理技術をより深い意味処理に一步近づけるための良い例題になると期待したい。

謝辞

本論文に対する, 佐藤理史氏 (京都大学), 山本和英氏 (長岡技術科学大学), 難波英嗣氏 (広島市立大学), 高橋哲朗氏 (奈良先端科学技術大学院大学) の有益なコメントに感謝する.

参考文献

- Abeille, A., Schabes, Y., and Joshi, A. K. (1990). "Using lexicalized TAGs for machine translation." In *Proceedings of the 13th International Conference on Computational Linguistics (COLING)*, Vol. 3, pp. 1–6.
- 安達久博 (1992). "手話通訳のためのニュース文の話しコトバへの変換処理." 電子情報通信学会技術研究報告, NLC92-47.
- Allan, J., Carbonell, J., Doddington, G., Yamron, J., and Yang, Y. (1998). "Topic detection and tracking pilot study: final report." In *Proceedings of the Broadcast News Understanding and Transcription Workshop*, pp. 194–218.
- Anick, P. G. and Tipirneni, S. (1999). "The paraphrase search assistant: terminological feedback for iterative information seeking." In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) Workshop on Customised Information Delivery*, pp. 153–159.
- 麻岡正洋, 佐藤理史, 宇津呂武仁 (2004). "語構成を利用した言い換え表現の自動生成." 言語処理学会第 10 回年次大会発表論文集, pp. 488–491.
- Bangalore, S. and Ranbow, O. (2000). "Corpus-based lexical choice in natural language generation." In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 464–471.
- Barzilay, R., McKeown, K. R., and Elhadad, M. (1999). "Information fusion in the context of multi-document summarization." In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 550–557.
- 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 (ACL)*, pp. 50–57.
- Barzilay, R. and Lee, L. (2002). "Bootstrapping lexical choice via multiple-sequence alignment." In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 164–171.
- Barzilay, R. and Lee, L. (2003). "Learning to paraphrase: an unsupervised approach using multiple-sequence alignment." In *Proceedings of the 2003 Human Language Technol-*

- ogy Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL), pp. 16–23.
- Barzilay, R. (2003). *Information fusion for multidocument summarization: paraphrasing and generation*. Ph.D. thesis, Columbia University.
- Beale, S., Nirenburg, S., Viegas, E., and Wanner, L. (1998). “De-constraining text generation.” In *Proceedings of the 9th International Workshop on Natural Language Generation (INLG)*, pp. 48–57.
- Brun, C. and Hagège, C. (2003). “Normalization and paraphrasing using symbolic methods.” In *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)*, pp. 41–48.
- Canning, Y. and Tait, J. (1999). “Syntactic simplification of newspaper text for aphasic readers.” In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) Workshop on Customised Information Delivery*, pp. 6–11.
- Carl, M. and Way, A. (Eds.) (2003). *Recent advances in example-based machine translation*. Kluwer Academic Publishers.
- Carroll, J., Minnen, G., Canning, Y., Devlin, S., and Tait, J. (1998). “Practical simplification of English newspaper text to assist aphasic readers.” In *Proceedings of the 15th National Conference on Artificial Intelligence and 10th Conference on Innovative Applications of Artificial Intelligence (AAAI-IAAI) Workshop on Integrating Artificial Intelligence and Assistive Technology*.
- Cawsey, A. (1992). *Explanation and interaction*. The MIT Press.
- Chandrasekar, R., Doran, C., and Srinivas, B. (1996). “Motivations and methods for text simplification.” In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, pp. 1041–1044.
- Clark, E. V. (1992). “Conventionality and contrast: pragmatic principles with lexical consequences.” In *Kittay and Lehrer (Eds.), Frames, fields, and contrasts: New essays in semantic and lexical organization*, pp. 171–188. Lawrence Erlbaum Associates.
- Dras, M. (1999). *Tree adjoining grammar and the reluctant paraphrasing of text*. Ph.D. thesis, Department of Computing, Macquarie University.
- Duclaye, F., Yvon, F., and Collin, O. (2003). “Learning paraphrases to improve a question-answering system.” In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL) Workshop on Natural Language Processing for Question-Answering*, pp. 35–41.

- Edmonds, P. (1999). *Semantic representations of near-synonyms for automatic lexical choice*. Ph.D. thesis, CSRI-399, Department of Computer Science, University of Toronto.
- 江原暉将, 福島孝博, 和田裕二, 白井克彦 (2000). “聴覚障害者向け字幕放送のためのニュース文自動短文分割.” 情報処理学会研究報告, NL-138-3, pp. 17–22.
- 藤田篤, 乾健太郎 (2001). “語釈文を利用した普通名詞の同概念語への言い換え.” 言語処理学会第7回年次大会発表論文集, pp. 331–334.
- 藤田篤, 乾健太郎 (2003). “語彙・構文的言い換えにおける変換誤りの分析.” 情報処理学会論文誌, **44** (11), pp. 2826–2838.
- 藤田篤, 乾健太郎, 松本裕治 (2004). “自動生成された言い換え文における不適格な動詞格構造の検出.” 情報処理学会論文誌, **45** (4), pp. 1176–1187.
- Fujita, A., Furihata, K., Inui, K., Matsumoto, Y., and Takeuchi, K. (2004). “Paraphrasing of Japanese light-verb constructions based on lexical conceptual structure.” In *Proceedings of the 42th Annual Meeting of the Association for Computational Linguistics (ACL) Workshop on Multiword Expressions: Integrating Processing*, pp. 9–16.
- Fukumoto, J., Kato, T., and Masui, F. (2002). “Question answering challenge (QAC): question answering evaluation at NTCIR workshop 3.” In *Working Notes of the 3rd NTCIR Workshop Meeting: QAC1*, pp. 1–10.
- 福島孝博, 江原暉将, 白井克彦 (1999). “短文分割の自動要約への効果.” 自然言語処理, **6** (6), pp. 131–147.
- Halliday, M. A. K. (1994). *An introduction to functional grammar (second edition)*. Edward Arnold.
- Harris, Z. (1981). “Co-occurrence and transformation in linguistic structure.” In *Hiz (Ed.) Papers on Syntax*, pp. 143–210. D. Reidel Publishing Company.
- 林良彦, 菊井玄一郎 (1991). “日本文推敲支援システムにおける書換え支援機能の実現方式.” 情報処理学会論文誌, **32** (8), pp. 962–970.
- Hermjakob, U., Echibahi, A., and Marcu, D. (2002). “Natural language based reformulation resource and web exploitation for question answering.” In *Proceedings of the 11th Text Retrieval Conference (TREC 2002)*.
- 日笠亘, 藤井綱貴, 黒橋禎夫 (1999). “入力質問と知識表現の柔軟なマッチングによる対話的ヘルプシステムの構築.” 情報処理学会研究報告, NL-134-14, pp. 101–108.
- Higashinaka, R. and Nagao, K. (2002). “Interactive paraphrasing based on linguistic annotation.” In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pp. 1218–1222.
- 平田大志, 日笠亘, 藤井綱貴, 黒橋禎夫 (2000). “図書館の自動リファレンス・サービス・シス

- テムの構築.” 言語処理学会第6回年次大会発表論文集, pp. 463–466.
- Hovy, E. H. (1988). *Generating natural language under pragmatic constraints*. Lawrence Erlbaum Associates.
- Huang, X. and Fiedler, A. (1996). “Paraphrasing and aggregating argumentative text using text structure.” In *Proceedings of the 8th International Workshop on Natural Language Generation (INLG)*, pp. 21–30.
- 飯田龍, 徳永泰浩, 乾健太郎, 衛藤純司 (2001). “言い換えエンジンKURAを用いた節内構造および機能語相当表現レベルの言い換え.” 第63回情報処理学会全国大会予稿集第二分冊, pp. 5–6.
- 池田諭史, 大橋一輝, 山本和英 (2004). “「新幹線要約」のための文末の整形.” 情報処理学会研究報告, NL-163-22.
- 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦 (編) (1997). 日本語語彙大系: CD-ROM 版. 岩波書店.
- 今村賢治, 秋葉泰弘, 隅田英一郎 (2001). “階層的句アライメントを用いた日本語翻訳文の換言.” 言語処理学会第7回年次大会ワークショップ論文集, pp. 15–20.
- 今村賢治 (2002). “構文解析と融合した階層的句アライメント.” 自然言語処理, 9 (5), pp. 23–42.
- Inkpen, D. Z. (2003). *Building a lexical knowledge-base of near-synonym differences*. Ph.D. thesis, Department of Computer Science, University of Toronto.
- 乾健太郎, 山本聡美, 野上優, 藤田篤, 乾裕子 (1999). “聾者向け文章読解支援における構文的言い換えの効果について.” 電子情報通信学会技術研究報告, WIT1999-2, pp. 9–14.
- 乾健太郎 (1999). “文章生成.” 自然言語処理—基礎と応用—, pp. 116–158. 電子情報通信学会.
- 乾健太郎 (2001). “コミュニケーション支援のための言い換え.” 言語処理学会第7回年次大会ワークショップ論文集, pp. 71–76.
- Inui, K. and Nogami, M. (2001). “A paraphrase-based exploration of cohesiveness criteria.” In *Proceedings of the 8th European Workshop on Natural Language Generation (EWNLG)*, pp. 101–110.
- Inui, K. and Hermjakob, U. (Eds.) (2003). *The 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)*. ACL-2003 Workshop.
- Inui, K., Fujita, A., Takahashi, T., Iida, R., and Iwakura, T. (2003). “Text simplification for reading assistance: a project note.” In *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)*, pp. 9–16.
- Iordanskaja, L., Kittredge, R., and Polguère, A. (1991). “Lexical selection and paraphrase in a meaning-text generation model.” In *Paris et al. (Eds.) Natural Language Generation in Artificial Intelligence and Computational Linguistics*, pp. 293–312. Kluwer Academic

Publishers.

- Iordanskaja, L., Kim, M., and Polguère, A. (1996). "Some procedural problems in the implementation of lexical functions for text generation." In *Wanner (Ed.) Lexical Functions in Lexicography and Natural Language Processing*, pp. 279–297. John Benjamin Publishing Company.
- Jacquemin, C., Klavans, J. L., and Tzoukermann, E. (1997). "Expansion of multi-word terms for indexing and retrieval using morphology and syntax." In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL-EACL)*, pp. 24–31.
- Kageura, K., Yoshikane, F., and Nozawa, T. (2004). "Parallel bilingual paraphrase rule for noun compounds: concepts and rules for exploring web language resources." In *Proceedings of the 4th Workshop on Asian Language Resources (ALR)*, pp. 54–61.
- 影山太郎 (編) (2001). 日英対照 動詞の意味と構文. 大修館書店.
- 鍛冶伸裕, 黒橋禎夫, 佐藤理史 (2001). "国語辞典に基づく平易文へのパラフレーズ." 情報処理学会研究報告, NL-144-23, pp. 167–174.
- 鍛冶伸裕, 河原大輔, 黒橋禎夫, 佐藤理史 (2003). "格フレームの対応付けに基づく用言の言い換え." 自然言語処理, **10** (4), pp. 65–81.
- 鍛冶伸裕, 黒橋禎夫 (2004). "迂言表現と重複表現の認識と言い換え." 自然言語処理, **11** (1), pp. 81–106.
- Kaji, N., Okamoto, M., and Kurohashi, S. (2004). "Paraphrasing predicates from written language to spoken language using the Web." In *Proceedings of the 2004 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*.
- Kanayama, H. (2003). "Paraphrasing rules for automatic evaluation of translation into Japanese." In *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)*, pp. 88–93.
- 金子和恵, 八木沢津義, 藤田稔 (1996). "話し手の性別・年齢を反映する文生成システム." 情報処理学会研究報告, NL-116-19, pp. 129–136.
- 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), pp. 79–98.
- 加藤輝政, 小川清, 佐良木昌 (1997). "英語構文の構文解析と編集、その論理と方法." 情報処理学会研究報告, NL-120-10, pp. 65–70.
- 川村珠巨 (2000). "言語生成プロセスの一視点—命題を言い換えるというコミュニケーション方

- 略について—.” 大阪府立工業高等専門学校研究紀要, Vol.34, pp. 77–82.
- 金淵培, 江原暉将 (1994). “日英機械翻訳のための日本語長文自動短文分割と主語の補完.” 情報処理学会論文誌, **35** (6), pp. 1018–1028.
- 木村健司, 徳永健伸, 田中穂積 (2001). “漢字インデックスを利用したパラフレーズの抽出.” 情報処理学会研究報告, NL-146-7, pp. 39–45.
- 木村健司, 徳永健伸, 田中穂積 (2002). “日本語名詞句に対するパラフレーズ事例の自動抽出に関する研究.” 言語処理学会第8回年次大会発表論文集, pp. 327–330.
- 金水敏, 工藤真由美, 沼田善子 (2000). 時・否定と取り立て. 岩波書店.
- Knight, K. and Chander, I. (1994). “Automated postediting of documents.” In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI)*, pp. 779–784.
- Knight, K. and Hatzivassiloglou, V. (1995). “Two-level, many-paths generation.” In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 252–260.
- Kondo, K. and Okumura, M. (1997). “Summarization with dictionary-based paraphrasing.” In *Proceedings of the 4th Natural Language Processing Pacific Rim Symposium (NLPRS)*, pp. 649–652.
- 近藤恵子, 佐藤理史, 奥村学 (1999). “「サ変名詞+する」から動詞相当句への言い換え.” 情報処理学会論文誌, **40** (11), pp. 4064–4074.
- 近藤恵子, 佐藤理史, 奥村学 (2001). “格変換による単文の言い換え.” 情報処理学会論文誌, **42** (3), pp. 465–477.
- 神田慎哉, 藤田篤, 乾健太郎 (2001). “連用節主節化に関する規則の追試と洗練.” 第15回人工知能学会全国大会, 1A1-06.
- 国広哲弥 (2000). “人はなぜ言葉を言い換えるか.” 言語, **29** (10), pp. 20–25.
- Kurohashi, S. and Sakai, Y. (1999). “Semantic analysis of Japanese noun phrases: a new approach to dictionary-based understanding.” In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 481–488.
- 黒橋禎夫, 酒井康行 (2001). “日本語の柔軟な照合.” 言語処理学会第7回年次大会発表論文集, pp. 343–346.
- 黒川和也 (2003). “日本語機能表現の正用・誤用の判別および日本語学習支援における利用に関する研究.” 豊橋技術科学大学大学院工学研究科修士論文.
- Langkilde, I. and Knight, K. (1998). “Generation that exploits corpus-based statistical knowledge.” In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL)*, pp. 704–710.

- Lapata, M., Keller, F., and McDonald, S. (2001). "Evaluating smoothing algorithms against plausibility judgements." In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 346–353.
- Lavoie, B., Kittredge, R., Korelsky, T., and Rambow, O. (2000). "A framework for MT and multilingual NLG systems based on uniform lexico-structural processing." In *Proceedings of the 6th Applied Natural Language Processing Conference and the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL)*, pp. 60–67.
- Levin, B. (1993). *English verb classes and alternations: a preliminary investigation*. Chicago Press.
- Lin, D. (1998). "Extracting collocations from text corpora." In *Proceedings of the 1st International Workshop on Computational Terminology (CompuTerm)*, pp. 57–63.
- Lin, D. and Pantel, P. (2001). "Discovery of inference rules for question answering." *Natural Language Engineering*, **7** (4), pp. 343–360.
- Mani, I., Gates, B., and Bloedorn, E. (1999). "Improving summaries by revising them." In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 558–565.
- Masuda, H., Yasutomi, D., and Nakagawa, H. (2001). "How to transform tables in HTML for displaying on mobile terminals." In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 29–36.
- 益岡隆志, 田窪行則 (1994). 基礎日本語文法 (改訂版). くろしお出版.
- 松吉俊, 佐藤理史, 宇津呂武仁 (2004). "機能表現「なら」の機械翻訳のための言い換え." 情報処理学会研究報告, NL-159-28, pp. 201–208.
- McKeown, K. R., Klavans, J. L., Hatzivassiloglou, V., Barzilay, R., and Eskin, E. (1999). "Towards multidocument summarization by reformulation: progress and prospects." In *Proceedings of the 16th National Conference on Artificial Intelligence and 11th Conference on Innovative Applications of Artificial Intelligence (AAAI-IAAI)*, pp. 453–460.
- Melamed, I. D. (2001). *Empirical methods for exploiting parallel texts*. The MIT Press.
- Mel'čuk, I. and Polguère, A. (1987). "A formal lexicon in meaning-text theory (or how to do lexica with words)." *Computational Linguistics*, **13** (3-4), pp. 261–275.
- Mel'čuk, I. (1996). "Lexical functions: a tool for the description of lexical relations in a lexicon." In Wanner (Ed.) *Lexical Functions in Lexicography and Natural Language Processing*, pp. 37–102. John Benjamin Publishing Company.

- Meteer, M. and Shaked, V. (1988). "Strategies for effective paraphrasing." In *Proceedings of the 12th International Conference on Computational Linguistics (COLING)*, pp. 431–436.
- Meyers, A., Yangarber, R., and Grishman, R. (1998). "Deriving transfer rules from dominance-preserving alignments." In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL)*, pp. 843–847.
- 三上真, 増山繁, 中川聖一 (1999). "ニュース番組における字幕生成のための文内短縮による要約." 自然言語処理, 6 (6), pp. 65–81.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J. (1990). "Introduction to WordNet: an on-line lexical database." *International Journal of Lexicography*, 3 (4), pp. 235–244.
- Mitamura, T. and Nyberg, E. (2001). "Automatic rewriting for controlled language translation." In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 1–12.
- 宮島達夫, 仁田義雄 (編) (1995a). 日本語類義表現の文法 (上) 単文編. くろしお出版.
- 宮島達夫, 仁田義雄 (編) (1995b). 日本語類義表現の文法 (下) 複文・連文編. くろしお出版.
- Moldovan, D., Clark, C., Harabagiu, S., and Maiorano, S. (2003). "COGEX: a logic prover for question answering." In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pp. 87–93.
- 森田良行, 松木正恵 (1989). 日本語表現文型—用例中心・複合辞の意味と用法. アルク.
- 森田良行 (1994). 動詞の意味論的文法研究. 明治書院.
- 村木新次郎 (1991). 日本語動詞の諸相. ひつじ書房.
- Murata, M. and Isahara, H. (2001). "Universal model for paraphrasing — using transformation based on a defined criteria —." In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 47–54.
- 村田真樹, 井佐原均 (2002). "受け身/使役文の能動文への変換における機械学習を用いた格助詞の変換." 情報処理学会研究報告, NL-149-6, pp. 39–44.
- Nanba, H. and Okumura, M. (2000). "Producing more readable extracts by revising them." In *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, pp. 1071–1075.
- 成松深, 河原大輔, 黒橋禎夫, 西田豊明 (2002). "格関係の比較を用いた複数テキスト間の重複・差分の検出." 言語処理学会第8回年次大会発表論文集, pp. 535–538.

- 日本電子化辞書研究所 (1995). EDR 電子化辞書仕様説明書.
- 野上優, 乾健太郎 (2001). “結束性を考慮した連体修飾節の言い換え.” 言語処理学会第 7 回年次大会発表論文集, pp. 339–342.
- 野上優, 乾健太郎 (2002). “言い換えを用いた結束性評価基準の構築.” 言語処理学会第 8 回年次大会発表論文集, pp. 335–338.
- Nyberg, E. and Mitamura, T. (2000). “The KANTOO machine translation environment.” In *Proceedings of AMTA 2000 Conference*, pp. 192–195.
- 小川修太, 石崎俊 (2004). “概念辞書における深層格の相互作用について—壁塗り構文を例として—.” 言語処理学会第 10 回年次大会発表論文集, pp. 572–575.
- 大橋一輝, 山本和英 (2004). “「サ変動詞＋名詞」の複合名詞への換言.” 言語処理学会第 10 回年次大会発表論文集, pp. 693–696.
- 大野満, 横山晶一, 西原典孝 (2003). “日本語敬語表現の変換・解析システム.” 言語処理学会第 9 回年次大会発表論文集, pp. 218–222.
- Ohtake, K. and Yamamoto, K. (2001). “Paraphrasing honorifics.” In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 13–20.
- Ohtake, K. and Yamamoto, K. (2003). “Applicability analysis of corpus-derived paraphrases toward example-based paraphrasing.” In *Proceedings of the 17th Pacific Asia Conference on Language, Information and Computation (PACLIC)*, pp. 380–391.
- 大泉敏貴, 鍛冶伸裕, 河原大輔, 岡本雅史, 黒橋禎夫, 西田豊明 (2003). “書きことばから話しことばへの変換.” 言語処理学会第 9 回年次大会発表論文集, pp. 93–96.
- Okamoto, H., Sato, K., and Saito, H. (2003). “Preferential presentation of Japanese near-synonyms using definition statements.” In *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)*, pp. 17–24.
- 奥雅博 (1990). “日本文解析における述語相当の慣用的表現の扱い.” 情報処理学会論文誌, **31** (12), pp. 1727–1734.
- 奥村学, 難波英嗣 (1999). “テキスト自動要約に関する研究動向.” 自然言語処理, **6** (6), pp. 1–26.
- 奥村学, 難波英嗣 (2002). “テキスト自動要約に関する最近の話題.” 自然言語処理, **9** (4), pp. 97–116.
- 大野晋, 浜西正人 (1981). 角川類語新辞典. 角川書店.
- Pang, B., Knight, K., and Marcu, D. (2003). “Syntax-based alignment of multiple translations: extracting paraphrases and generating new sentences.” In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pp. 102–109.

- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). "BLEU: a method for automatic evaluation of machine translation." In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 311–318.
- Pearce, D. (2001). "Synonymy in collocation extraction." In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL) Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, pp. 41–46.
- Pereira, F., Tishby, N., and Lee, L. (1993). "Distributional clustering of English words." In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 183–190.
- Quirk, C., Brockett, C., and Dolan, W. (2004). "Monolingual machine translation for paraphrase generation." In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 142–149.
- Ravichandran, D. and Hovy, E. (2002). "Learning surface text patterns for a question answering system." In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 215–222.
- Robin, J. and McKeown, K. (1996). "Empirically designing and evaluating a new revision-based model for summary generation." *Artificial Intelligence*, **85** (1-2), pp. 135–179.
- RWC (1998). RWC テキストデータベース第2版, 岩波国語辞典タグ付き/形態素解析データ第5版.
- 斉藤健太郎, 池原悟, 村上仁一 (2002). "日英文型パターンの意味的対応方式." 言語処理学会第9回年次大会発表論文集, pp. 295–298.
- 酒井浩之, 増山繁 (2003). "コーパスからの名詞と略語の対応関係の自動獲得." 言語処理学会第9回年次大会発表論文集, pp. 226–229.
- Sasaki, Y., Isozaki, H., Kokuryou, K., Hirao, T., , and Maeda, E. (2002). "NTT's QA Systems for NTCIR QAC-1." In *NTCIR Workshop 3 Question Answering Challenge (QAC)*.
- 佐藤大, 岩越守孝, 増田英孝, 中川裕志 (2004). "Web と携帯端末向けの新聞記事の対応コーパスからの言い換え抽出." 情報処理学会研究報告, NL-159-27, pp. 193–200.
- 佐藤理史 (1999). "論文表題を言い換える." 情報処理学会論文誌, **40** (7), pp. 2937–2945.
- 佐藤理史 (2001). "なぜ言い換え/パラフレーズを研究するのか." 言語処理学会第7回年次大会ワークショップ論文集, pp. 1–2.
- Sato, S. and Nakagawa, H. (Eds.) (2001). *Workshop on Automatic Paraphrasing: Theories and Applications*. NLP RS-2001 Workshop.
- 関根聡 (2001). "複数の新聞を使用した言い替え表現の自動抽出." 言語処理学会第7回年次大

- 会ワークショップ論文集, pp. 9–14.
- Shieber, S. M. and Schabes, Y. (1990). “Synchronous tree-adjoining grammars.” In *Proceedings of the 13th International Conference on Computational Linguistics (COLING)*, Vol. 3, pp. 253–258.
- 下畑光夫, 渡辺太郎, 隅田英一郎, 松本裕治 (2003). “パラレルコーパスからの機械翻訳向け同義表現抽出.” 情報処理学会論文誌, **44** (11), pp. 2854–2863.
- Shimohata, M., Sumita, E., and Matsumoto, Y. (2004). “Building a paraphrase corpus for speech translation.” In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, pp. 1407–1410.
- Shinyama, Y., Sekine, S., Sudo, K., and Grishman, R. (2002). “Automatic paraphrase acquisition from news articles.” In *Proceedings of the Human Language Technology Conference (HLT)*.
- Shinyama, Y. and Sekine, S. (2003). “Paraphrase acquisition for information extraction.” In *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)*, pp. 65–71.
- 白井諭, 池原悟, 河岡司, 中村行宏 (1995). “日英機械翻訳における原文自動書き換え型翻訳方式とその効果.” 情報処理学会論文誌, **36** (1), pp. 12–21.
- Shirai, S., Yamamoto, K., and Bond, F. (2001). “Japanese-English paraphrase corpus.” In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Language Resources in Asia*, pp. 23–30.
- 白木伸征, 黒橋禎夫 (2000). “自然言語入力と目次との柔軟な照合による図書検索システム.” 情報処理学会論文誌, **41** (4), pp. 1162–1170.
- 首藤公昭, 田辺利文, 吉村賢治 (2001). “日本語モダリティ表現とその言い換え.” 言語処理学会第7回年次大会ワークショップ論文集, pp. 47–50.
- Siddharthan, A. (2003). “Preserving discourse structure when simplifying text.” In *Proceedings of the 9th European Workshop on Natural Language Generation (EWNLG)*, pp. 103–110.
- 砂川有里子 (1995). “日本語における分裂文の機能と語順の原理.” 複文の研究 (下), 仁田義雄 (編), pp. 353–388. くろしお出版.
- Takahashi, T., Iwakura, T., Iida, R., Fujita, A., and Inui, K. (2001). “KURA: a transfer-based lexico-structural paraphrasing engine.” In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 37–46.
- Takahashi, T., Nawata, K., Inui, K., and Matsumoto, Y. (2003). “Effects of structural match-

- ing and paraphrasing in question answering.” *IEICE Transactions on Information and Systems*, **E86-D** (9), pp. 1677–1685.
- 高橋哲朗, 乾健太郎, 関根聡, 松本裕治 (2004). “質問応答に必要な言い換えの分析.” 言語処理学会第 10 回年次大会発表論文集, pp. 309–312.
- 高橋善文, 牛島和夫 (1991). “計算機マニュアルのわかりやすさの定量的評価方法.” 情報処理学会論文誌, **32** (4), pp. 460–469.
- 高塚成信 (1999). “コミュニケーション方略としての「言い換え」—その指導内容と方法に関する基礎的研究—.” 岡山大学教育学部研究集録, 第 10 号, pp. 1–12.
- 武石英二, 林良彦 (1992). “接続構造解析に基づく日本語複文の分割.” 情報処理学会論文誌, **33** (5), pp. 652–663.
- 竹内孔一, 内山清子, 吉岡真治, 影浦峯, 小山照夫 (2002). “語彙概念構造を利用した複合名詞内の係り関係の解析.” 情報処理学会論文誌, **43** (5), pp. 1446–1456.
- Terada, A. and Tokunaga, T. (2001). “Automatic disabbreviation by using context information.” In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 21–28.
- 徳田昌晃, 奥村学 (1998). “日本語から手話への機械翻訳における手話単語辞書の補完方法について.” 情報処理学会論文誌, **39** (3), pp. 542–550.
- Tokunaga, T., Tanaka, H., and Kimura, K. (2003). “Paraphrasing Japanese noun phrases using character-based indexing.” In *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)*, pp. 80–87.
- 徳永泰浩 (2002). “取り立て詞に着目した否定表現の言い換えと意味解析.” 九州工業大学情報工学部知能情報工学科卒業論文.
- Tomuro, N. and Lytinen, S. L. (2001). “Selecting features for paraphrasing question sentences.” In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 55–62.
- 鳥澤健太郎 (2002). “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 回年次大会発表論文集, pp. 323–326.
- Torisawa, K. (2002). “An unsupervised learning method for associative relationships between verb phrases.” In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pp. 1009–1015.
- 土屋雅稔, 黒橋禎夫 (2000). “MDL 原理に基づく辞書定義文の圧縮と共通性の発見.” 情報処理学会研究報告, NL-140-7, pp. 47–54.
- 土屋雅稔, 佐藤理史, 宇津呂武仁 (2004). “機能表現言い換えデータからの言い換え規則の自動生成.” 言語処理学会第 10 回年次大会発表論文集, pp. 492–495.

- 内元清貴, 黒橋禎夫, 長尾眞 (1996). “日本語文生成における語彙選択に必要な要因とその性質.” 情報処理学会研究報告, NL-116-21, pp. 143–150.
- 内山清子, 石崎俊 (2003). “複合動詞の多義性解消のための意味解析法.” 言語処理学会第9回年次大会発表論文集, pp. 163–166.
- 内山清子, Baldwin, T. (2004). “機械学習を用いた複合動詞の多義性解消.” 言語処理学会第10回年次大会発表論文集, pp. 741–744.
- 上田良寛, 小山剛弘 (2000). “共通意味断片の抽出による複数文書要約.” 言語処理学会第6回年次大会発表論文集, pp. 360–363.
- Voorhees, E. M. (2001). “Overview of the TREC 2001 question answering track.” In *Proceedings of the 10th Text Retrieval Conference (TREC 2001)*, pp. 42–51.
- Wahlster, W. (Ed.) (2000). *VerbMobil: foundations of speech-to-speech translation*. Springer.
- 若尾孝博, 江原暉将, 白井克彦 (1997). “テレビニュース番組の字幕に見られる要約の手法.” 情報処理学会研究報告, NL-122-13, pp. 83–89.
- Walker, M. A. (1993). “When given information is accented: repetition, paraphrase and inference in dialogue.” In *Proceedings of Linguistics Society of America Annual Meeting*.
- Wanner, L. (1994). *Current issues in Meaning-Text Theory*. Pinter Publishers.
- Watanabe, H., Kurohashi, S., and Aramaki, E. (2000). “Finding structural correspondences from bilingual parsed corpus for corpus-based translation.” In *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, pp. 933–939.
- 山口昌也, 乾伸雄, 小谷善行, 西村恕彦 (1998). “前編集結果を利用した前編集自動化規則の獲得.” 情報処理学会論文誌, **39** (1), pp. 17–28.
- 山本薫, 松本裕治 (2001). “統計的係り受け結果を用いた対訳表現抽出.” 情報処理学会論文誌, **42** (9), pp. 2239–2247.
- 山本和英 (2001). “換言処理の現状と課題.” 言語処理学会第7回年次大会ワークショップ論文集, pp. 93–96.
- Yamamoto, K. (2002a). “Machine translation by interaction between paraphraser and transfer.” In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pp. 1107–1113.
- Yamamoto, K. (2002b). “Acquisition of lexical paraphrases from texts.” In *Proceedings of the 2nd International Workshop on Computational Terminology (CompuTerm)*, pp. 22–28.
- Yoshikane, F., Tsuji, K., Kageura, K., and Jacquemin, C. (2003). “Morpho-syntactic rules for detecting Japanese term variation: establishment and evaluation.” 自然言語処理, **10** (4), pp. 3–32.
- 吉見毅彦, 佐田いち子 (2000). “英字新聞記事見出し翻訳の自動前編集による改良.” 自然言語

処理, 7 (2), pp. 27-43.

吉見毅彦, 佐田いち子, 福持陽士 (2000). “頑健な英日機械翻訳システム実現のための原文自動前編集.” 自然言語処理, 7 (4), pp. 99-117.

付録 A 語彙・構文的言い換えの分類

これまでの事例研究の中で扱われてきたさまざまな種類の言い換えや, 言語学の分野で示されてきた交替現象, 表現の使い分けなどの分析結果を集めた. さらに, それぞれの言い換えを実現するための課題を考察し, (a) 言い換えのスコープ, (b) 内容表現か機能表現か, (c) 必要な語彙知識の種類, という観点から分類した.

A.1 節間の言い換え

2つ以上の節にまたがる言い換えである. (36), (37) のような言い換えでは, 主題が変更するため, それにともなって対応する名詞述語表現が必要になる. 一方, (38), (39) のような言い換えでは, 節間の修辭的關係を表す接続詞を改めて選択しなければならない. このように, 節間の言い換えでは, 節間の順序や關係が変化するため, 結束性の評価が必要になる.

(36) 連体節主節化 (Chandrasekar et al. 1996; Dras 1999; 野上, 乾 2001)

- s. 昨年, 区制施行 70 周年という大きな節目を迎えた本区は, 新たな 10 年に向けて順調な区政運営をスタートいたしました.
- t. 昨年, 本区は区制施行 70 周年という大きな節目を迎えました. そして, 新たな 10 年に向けて順調な区政運営をスタートいたしました.

(37) 分裂文から非分裂文への言い換え (砂川 1995; Dras 1999)

- s. 今週当選したのは, 奈良県の男性でした.
- t. 今週は, 奈良県の男性が当選しました.

(38) 連用節・並列節の分割 (武石, 林 1992; 神田ら 2001; Mitamura and Nyberg 2001)

- s. 情報化に向けての前向きな意見が多くを占めていますが, 情報格差などの不安もみられます.
- t. 情報化に向けての前向きな意見が多くを占めています. しかし, 情報格差などの不安もみられます.

(39) 接続表現の言い換え (宮島, 仁田 1995b)

- s. 用紙は各事務所に置いてありますから, どうし意見をお寄せください.
- t. 用紙は各事務所に置いてありますので, どうし意見をお寄せください.

A.2 節内の言い換え

(41) の主題交替や, (42) の格交替など, 操作の対象が節内で閉じている言い換えである. 変換パターンのバリエーションはそれほど多くなく, 人手で書き尽くせる程度のように見えるが, 変換パターンによっては適用の可否が語に依存するため, その判断に必要な語彙知識をいかにして発見・構築するかが課題となる. また, 視点のような対人関係の意味や主題/陳述構造のような文脈レベルの意味が変化するため, これを捉えるモデルを形式化する必要もある.

- (40) 否定表現の言い換え (林, 菊井 1991; 近藤ら 2001; 飯田ら 2001; 徳永 2002)
- s. 返信しないと, 申込みは取り消されます.
 - t. 返信すると, 申込みは取り消されません.
- (41) 比較表現の言い換え (近藤ら 2001; 斉藤ら 2002)
- s. 隣町は我が町より山林資源が乏しい.
 - t. 我が町は隣町より山林資源が豊かだ.
- (42) 態・使役の交替 (山口ら 1998; 近藤ら 2001; 村田, 井佐原 2002)
- s. 今年は湾岸などの都市基盤の整備が行われました.
 - t. 今年は湾岸などの都市基盤の整備を行いました.
- (43) 動詞交替 (自他) (Levin 1993; 影山 2001; 近藤ら 2001)
- s. 無制限な個人情報の収集に一定の制限を加える.
 - t. 無制限な個人情報の収集に一定の制限が加わる.
- (44) 動詞交替 (壁塗り/場所格) (Levin 1993; 影山 2001; 小川, 石崎 2004)
- s. 課長は大きな杯に日本酒を満たした.
 - t. 課長は大きな杯を日本酒で満たした.
- (45) 機能動詞結合の言い換え (奥 1990; 村木 1991; Iordanskaja et al. 1991; 森田 1994; Dras 1999; 鍛冶, 黒橋 2004; Fujita et al. 2004)
- s. 住民の熱心な要請を受け, 工事を中止した.
 - t. 住民に熱心に要請され, 工事を中止した.
- (46) 授受の構文の言い換え (益岡, 田窪 1994; 乾ら 1999)
- s. 区民の健康保持の立場から, 清掃活動を頑張ってくれている.
 - t. 区民の健康保持の立場から, 清掃活動を頑張っている.
- (47) 可能動詞の言い換え (宮島, 仁田 1995a; 乾ら 1999)
- s. 電車は込んでいたけど吊革に掴まれた.
 - t. 電車は込んでいたけど吊革に掴まることができた.
- (48) 修飾要素の交替 (宮島, 仁田 1995b)
- s. ブロック, レンガなど大きくて重いものは, ひも掛けをしてそのまま出してください.
 - t. ブロック, レンガなど大きく重いものは, ひも掛けをしてそのまま出してください.

(49) 数量詞の遊離 (村木 1991)

- s. 一件の開示請求がありました。
- t. 開示請求が一件ありました。

A.3 内容語の複合表現の言い換え

複数の内容語が複合表現を形成する場合、接続詞や格助詞、共通の動詞などの関係は明示的には現れない。この隠れている関係を明示的に示すように言い換えれば（下記 $s \Rightarrow t$ ），読解支援という目的に対しては有効だと考えられる。一方、要約のような応用には、複合表現への言い換え（下記 $t \Rightarrow s$ ）の方が有効である。これらの言い換える可否は、構成素となる内容語からそれらの関係（複合表現における結び付き方）が容易に連想できるかどうかにも依存する。

(50) 複合名詞の分解・構成 (佐藤 1999; 木村ら 2002; 竹内ら 2002; 大橋, 山本 2004)

- s. 区政施行 70 周年という大きな節目を迎えました。
- t. 区の行政が施行されてから 70 周年という大きな節目を迎えました。

(51) 「A の B」 \Leftrightarrow 連体節 (Kurohashi and Sakai 1999; 片岡ら 2000; 鳥澤 2002)

- s. 奈良県知事選への出馬の挨拶を行った。
- t. 奈良県知事選への出馬を表明する挨拶を行った。

(52) 複合動詞の分解・構成 (内山, 石崎 2003; 内山, Baldwin 2004)

- s1. 人に頷きかける。
- t1. 人に向かって頷く。
- s2. 夕飯を食べ過ぎた。
- t2. 夕飯を必要以上に食べた。

A.4 機能語／モダリティの言い換え

機能語相当表現（助詞・助動詞）やモダリティのレベルの言い換えは、上で示した言い換えに比べて語彙的な性格が強く、局所的な情報を参照するだけで言い換えられるものも多い。このレベルの言い換えを実現するには、同義の機能語／モダリティ表現をグループ化して辞書を整備するとともに、個々の言い換えで生じる意味差分をどのように計算するかが課題となる。

(53) 機能語相当表現の言い換え (森田, 松木 1989; 飯田ら 2001; 黒川 2003; 松吉ら 2004; 土屋ら 2004)

- s. 市民はもとより全国に誇れるものにしていきたい。
- t. 市民だけでなく全国に誇れるものにしていきたい。

(54) 取り立て助詞の移動 (金水ら 2000; 徳永 2002)

- s. ご飯は食べずに、辛いおかずを食べてばかりいた。
- t. ご飯は食べずに、辛いおかずばかりを食べていた。

- (55) 助詞による特徴づけの削除 (森田 1994)
- s. 人口は一時 10 万人を超えこそしたが, 今は 7 万人まで減少している.
 - t. 人口は一時 10 万人を超えたが, 今は 7 万人まで減少している.
- (56) 伝達のもダリティ (宮島, 仁田 1995a; 乾ら 1999; 首藤ら 2001)
- s. 秋には紅葉を見に多くの人が集まるという.
 - t. 秋には紅葉を見に多くの人が集まるそうだ.
- (57) 敬語表現の言い換え (Ohtake and Yamamoto 2001; 大野ら 2003)
- s. お支払いの方はいかがなさいますか.
 - t. お支払いの方はどうなさいますか.
- (58) 文体の変換 (大泉ら 2003; Kaji et al. 2004)
- s. 不本意だが仕方がない.
 - t. 不本意ですが仕方ありません.

A.5 内容語句の言い換え

内容語の言い換え表現は個々の単語ごとに記述する必要があるため, パラレルコーパスのアラインメント, シソーラス中の同概念語, 国語辞典の語釈文などの既存の資源から機械的に収集する手段が検討されている (5 節).

- (59) 名詞の言い換え (藤田, 乾 2001; Pearce 2001; Yamamoto 2002b; Okamoto et al. 2003)
- s. 太平洋を一望する桂浜公園の丘陵に完成.
 - t. 太平洋を一望する桂浜公園の高台に完成.
- (60) 動詞の言い換え (Kondo and Okumura 1997; 近藤ら 1999; 鍛冶ら 2001; Torisawa 2002; 鍛冶ら 2003)
- s. 警官が犯人を逮捕する.
 - t. 警官が犯人を捕まえる.

A.6 慣用表現の言い換え

構成語の変形では生成できない特有の言い回しの言い換えは, 言い換え表現対を辞書に蓄える必要がある.

- (61) 慣用句 (森田 1994; Mitamura and Nyberg 2001)
- s. 「ひかり都市」として脚光を浴びることとなりました.
 - t. 「ひかり都市」として注目されることとなりました.
- (62) 表記のゆれ/略語 (若尾ら 1997; Terada and Tokunaga 2001; 酒井, 増山 2003)
- s. 多くの市民が原発の建設に反対している.
 - t. 多くの市民が原子力発電所の建設に反対している.

(63) 換喩

- s. シェイクスピアを読む.
- t. シェイクスピアが書いた本を読む

略歴

乾 健太郎: 1967 年生. 1995 年東京工業大学大学院情報理工学研究科博士課程修了. 同年より同研究科助手. 1998 年より九州工業大学情報工学部助教授. 1998 年~2001 年科学技術振興事業団さきがけ研究 21 研究員を兼任. 2001 年より奈良先端科学技術大学院大学情報科学研究科助教授. 現在にいたる. 博士(工学). 自然言語処理の研究に従事. 情報処理学会, 人工知能学会, ACL 各会員. inui@is.naist.jp.

藤田 篤: 1977 年生. 2000 年九州工業大学情報工学部卒業. 2002 年同大学大学院情報理工学研究科博士前期課程修了. 同年, 奈良先端科学技術大学院大学情報科学研究科博士課程入学. 現在にいたる. 自然言語処理, 特にテキストの自動言い換えの研究に従事. 情報処理学会, ACL 各学生会員. atsush-f@is.naist.jp.

(2004 年 5 月 20 日 受付)

(2004 年 7 月 5 日 採録)