

言い換え計算モデルの工学的実現への 語彙意味論的アプローチ

乾 健太郎，藤田 篤

キーワード：自然言語処理，言い換え，言い換え生成，語彙概念構造，統語構造変換，機能動詞構文

1 言い換え技術と語彙意味論

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。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。自然言語処理の立場からこの言語横断的な言語現象を見るとき、言い換えの問題は言い換え生成と言い換え認識という2種類の問題に集約することができる。

言い換え生成は、所与の言語表現からその言い換えを生成する問題である。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 Web 上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 (Canning and Tait 1999, Inui et al. 2003, etc.)。あるいは、機械翻訳の前編集として入力文を翻訳しやすい言い回しに自動変換するといったこ

とも可能になるだろう (吉見・佐田 2000, 吉見ら 2000, Mitamura and Nyberg 2001, etc.)。

一方, 言い換え認識は, 異なる 2 つの言語表現が同義か否かを判定する問題である。例として, 与えられた文書集合から質問の答えを探し出す質問応答の問題を考えよう (Lin and Pantel 2001, Ravichandran and Hovy 2002, etc.)。もしも計算機が「《著作名》の著者は《人名》だ」という言い回しと「《人名》が《著作名》を発表する」という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?』のような質問に対して「夏目漱石が『坊ちゃん』を発表した明治 39 年は, …」のような文書集合中の記述からその答えを見つけることができるだろう。こうした技術は文書からの情報抽出や複数文書要約 (Shinyama and Sekine 2003, Barzilay 2003, etc.) など幅広いアプリケーションで有用である。

これら言い換えに関する 2 つの問題は一見別々の問題のように見えるが, 実はそうでない。言い換え認識は, 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。逆に, 言い換え生成は, 生成物の候補群から言い換えと判定できるものを選択する問題と見なせる。

このように, 言い換えの問題は, 広範な自然言語処理アプリケーションに貢献する重要な研究課題と言える。近年, これに関心を持つ言語処理研究者が増え, 言い換えというキーワードが目立つようになってきた。たとえば, 言語処理学会の年次大会でも, 2001 年に言い換えのセッションが設置されて以来, 毎年数件の投稿を着実に集め, 現在に至っている。また, 2001 年から言い換えに関する国際ワークショップが隔年で開催され, 2005 年には第 3 回を迎えた (Sato and Nakagawa 2001, Inui and Hermjakob 2003, Dras and Yamamoto 2005)。

さて, 言い換えは意味の等価性に基づいて行う統語的操作と見なせるので, 語の統語的振る舞いをその意味的特性で説明しようとする語彙意味論は, 言い換え研究に極めて有用な基盤を提供すると期待できる。にもかかわらず, これまでの工学的言い換え研究はその成果をほとんど活用してこなかった。本論文では, 言い換えの問題, とくに言い換え生成に焦点を当て, 言い換えの計算モデルを構築する方法について工学的立場から論じ, 我々の試みを紹

介する。その中で、語彙意味論の研究成果をどのように利用すべきか、また今後の語彙意味論研究に何を期待するかを論じる。

以下、本論文は、まず2章で言い換えと呼ばれる言語現象の広がり概観し、我々が当面の研究対象と考える言い換えの種類を位置付ける。3章では言い換え生成の技術的背景となる機械翻訳と述語項構造解析の研究動向を紹介する。次に、2章、3章の議論を受けて、4章で語彙意味論に基づく言い換え生成モデルの構築を目指す語彙意味論的アプローチを論じる。最後に、5章、6章でこのアプローチを実現するための我々の取り組みを紹介する。

2 どのような言い換えを対象とするか

言い換えと呼べる現象は多岐にわたる。その中には談話の状況に関する高度な推論を要するものもあり、現在の技術ですべてをカバーするのは難しい。本題に入る前に、当面の研究対象と考える言い換えの種類について概観する。

2.1 意味の等価性

言い換えは意味が近似的に等価な言語表現の異形である。とすれば、言い換えの定義と分類は「意味が等価である」ことの意味によって与えられると考えられる。意味の等価性には少なくとも次の3つのレベルがある。

内包的意味の等価性 変形文法やLFG, HPSGなどを含む現代文法理論の多くは、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、多くの場合、語あるいは言語表現の内包的な意味(intension)、より具体的には真理値意味論(モデル意味論)の意味を指す。

参照的意味の等価性 語の意味に関する多くの哲学的考察が明らかにしたように、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。たとえば、Frege の「宵の明星」「明けの明星」の例は有名である。ただし、内包的意味が違っていても、次の例のように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。

- (1) a. 去年の出来事 ⇔ 2005 年の出来事 (佐藤 1999)
 b. 筆者の考え ⇔ 乾の考え

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、内包的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。

語用論的效果の等価性 言語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、内包的意味が異なっても、語用論的效果が同じであれば、言い換えと見なせる場合がある。

- (2) a. どなたか gcc のソースのありかをご存知ないでしょうか
 ⇔ gcc のソースが置いてある ftp サイトを教えてください (佐藤 1999)
 b. Will you break this bill?
 ⇔ I want to use that vending machine. (川村 2000)

こうした言い換えは世界知識による高度な推論が必要か、あるいは社会の中で慣習化された固定的なものである場合が多く、統語論・意味論だけでは説明が難しい。

参照的意味の等価性による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、語用論的效果の等価性による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。したがって、工学的に最も扱いやすい対象は内包的意味の等価性による言い換えということになる。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タイプの言い換えに限定している。以下、内包的意味の等価性に基づく言い換えを便宜的に語彙・構文的言い換えと呼び、これに対象をしぼって議論する。

2.2 語彙・構文的言い換え

語彙・構文的言い換えに限っても、純粋に統語論で扱えそうな言い換えから語の詳細な意味に立ち入る必要のある言い換えまで多岐にわたる。こうした言い換えは、実現に必要な知識の種類の観点から便宜的に次の4種類に分けられる。

統語的言い換え 個別の語の意味に立ち入らなくても統語論の記述レベルで概ね説明できる言い換え

- (3) 最初に合格したのは高橋さんだ ⇔ 高橋さんが最初に合格した

語彙的言い換え 語の同義性だけで概ね説明できる、統語操作を伴わない局所的言い換え

- (4) 一層の苦境に陥る恐れがある ⇔ 一層の窮地に陥る可能性がある

語彙構成的言い換え 語の統語的特性と意味的特性に基づいて構成的に説明できると考えられる規則性の高い言い換え

- (5) 2位が先頭との距離を縮めた ⇔ 2位と先頭の距離が縮まった

推論的言い換え 内包的意味が近似的に等価な言い換えのうち、上のどの種類にも当てはまらないもの

- (6) 財政再建が急務の課題だ ⇔ 緊急に財政再建する必要がある

これらのうち統語的言い換えは、人手で書き尽くせる程度の数の一般性の高い書き換えパターンで実現できると考えられる。この種の言い換えでは、上の例(2)のように文の構造が大きく変わる場合が多いので、主題、焦点、新情報・旧情報といった情報構造の変化の適否が主要な問題になる。一方語彙的言い換えは、語の置換の可否に関する大規模な語彙知識を必要とする問題であり、語の同義性や語と語の共起に関する知識を例えばコーパスのような既存の資源からどのように獲得するかが課題となる。これらに対し語彙構成

的言い換えは、統語的言い換えと語彙的言い換えの中間に位置するもので、統語的変換操作を伴う点では統語的言い換えに近いが、変換操作が語の意味に依存して決まるという点で語彙的でもある。個々の語の意味と統語的性質の両方の知識が必要になるため、この種の言い換えの実現には語彙意味論研究の成果が有用であると考えられる。4章以降では、この語彙構成的言い換えを対象に語彙意味論的アプローチを論じる。

3 言い換え技術に関する技術的背景

3.1 機械翻訳

言い換えの問題、とくに言い換え生成は、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、機械翻訳と本質的には同じ問題である。機械翻訳は自然言語処理の中心的な研究テーマとして同領域を黎明期から牽引してきた。したがって、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少なくない。

現在の機械翻訳研究は大きく2つの方向に進んでいる。1つ目は、統計的手法や機械学習を用いることにより頑健な翻訳システムを安価に構築することを目指す方向である。大規模な対訳コーパスから翻訳知識を自動獲得したり、翻訳の確率モデルを学習する技術が発展を見せ、まったく新しい言語対についてもある程度の性能を示す翻訳システムが短期間で開発できるようになった¹。

2つ目は、精密文法と高度な語彙知識を用いた知識集約的な意味構造変換を目指す方向である。機械翻訳のプロセスは典型的には、(i) 原言語の入力文から中間構造（統語構造あるいは意味構造）を得る解析処理、(ii) 得られた中間構造を目的言語の中間構造に変換する構造変換処理、(iii) 変換後の中間構造から目的言語の表層文を得る生成処理の3段階からなる。中間構造として統語構造を仮定するか、意味構造を仮定するかによって、さらに統語構造変換方式あるいは意味構造変換方式に分類される。

意味構造変換は、統語構造変換に比べて中間構造の抽象度が高いため、言

語表現の多様性や言語間の差異をより高度に捨象し、変換規則の複雑性を抑えることができる (Dorna et al. 1998)。ただし、その反面、初期の技術では頑健で高精度な意味構造解析を実現することは難しく、また言語の意味に匹敵する表現能力を持った意味構造を設計すること自体が困難であるといった問題もあった。近年になると、主辞句構造文法 (HPSG) や語彙機能文法 (LFG) のような精密文法 (precision grammar) による統語・意味解析の技術が飛躍的に進歩し、これに基づく意味構造変換方式の翻訳システムを構築する計画も報告されている (Lønning et al. 2004)。

3.2 述語項構造解析

入力文を解析して意味構造 (主として述語項構造) を得るためには、形態素、統語、意味の各レベルのさまざまな曖昧性を解消する必要がある。このうち形態素・統語レベルの曖昧性解消については、大規模な正解タグ付きコーパスを用意し、種々の統計的手法を用いて曖昧性解消規則を機械的に獲得する方法が成功を収め、実用に耐える解析精度を得るに至っている (Manning and Schütze 1999)。

形態素・統語レベルの曖昧性が解消され、入力文の統語構造が決まったとしても、そこから述語項構造を得るには、さらに次のような曖昧性を解消する必要がある。後述するように、本論文で論じる語彙意味論的アプローチでは入力文の語彙概念構造 (LCS) を解析する処理が必要になるが、この問題は以下に述べる述語項構造解析と本質的には同じ問題である。

係助詞が兼務する表層格の曖昧性 「は」や「も」などの係助詞はガ格やヲ格を兼務するので、どの表層格を兼務しているかを推定する必要がある。

助詞落ち名詞句の表層格の曖昧性 「宿題やった？」のような助詞落ち名詞句がどの表層格 (あるいは項) のフィラーなのかを推定する必要がある。

連体修飾節と被修飾名詞の意味関係に関する曖昧性 連体修飾節の主辞の述語と被修飾名詞の間には、格関係がある場合（内の関係。例えば、「魚を焼く男」の「焼く」と「男」）と格関係がない場合（外の関係。例えば、「魚を焼くにおい」の「焼く」と「におい」）がある。さらに内の関係の場合は、被修飾名詞が連体修飾節の述語のどの表層格を埋めるかに関する曖昧性がある。

述語が取る項構造の曖昧性 「女性が悲鳴をあげる」と「彼が彼女にプレゼントをあげる」の動詞「あげる」のように、述語が複数の項構造を取りうる場合に、その曖昧性を解消する必要がある。

格の意味役割の曖昧性 「その本なら3日前に花子に返した」の「に」のように、項への対応が一意に決まらない表層格標示があるため、これを解消する必要がある。

ヴォイスによる格交替の曖昧性 「れる、られる、せる、させる」などの助動詞によって態が変化している場合に、受動態か可能態かと言った多義性や、格交替の曖昧性が生じる場合がある。

近年は、述語項構造レベルの曖昧性解消にも多くの研究者が関心を寄せるようになり、少なくとも英語については、項構造情報を含む詳細な意味情報を記述した大規模な辞書が開発され (Dorr 1997, Baker et al. 1998, etc.) , そうした情報を付与したタグ付きコーパスも入手可能になっている (Palmer et al. 2005)。こうした資源の整備は、述語項構造解析にも統計的手法の適用を可能にし、頑健で高精度な解析技術の開発への道を拓いた。

統計的手法には、人手で正解情報を与えた訓練用事例を用意し、そこから曖昧性解消のための確率モデルなどを学習する教師あり手法と、そうした訓練用事例を使わない教師なし手法がある。教師なし手法は訓練用事例を作成する人的コストを必要としない反面、一般に精度が低い (Lapata and Brew 1999)。したがって、高い精度を得るにはやはり大量の訓練用事例が必要で

あり、どうすれば十分な量の訓練用事例を安価に収集することができるかが重要な課題になっている。

4 語彙意味論に基づく言い換え計算モデルの構築に向けて

4.1 語彙構成的言い換えへの語彙意味論的アプローチ

統語構造変換を選択するにせよ、意味構造変換を目指すにせよ、与えられた構造(の一部)を同義な構造に変換するには、部分構造間の同義性に関する膨大な知識が必要である。こうした知識をここでは言い換え知識と呼ぶ。言い換えの計算モデルが実用規模で機能するためには、大規模な言い換え知識が必要となるので、その開発および保守を効率化するための方法論が重要な研究課題になる。これには主として、人手で作成された既存の語彙資源を利用するアプローチとコーパスから言い換え知識を自動獲得するアプローチがある。詳細は別稿(乾・藤田 2004)に譲るが、既存の語彙資源から抽出できるのは限定的な種類の言い換え知識だけであり、またコーパスから力任せに自動獲得する方法もこれまでのところ実用に耐える成果を上げられていない。

一方、2章で見たように、我々が当面の目標とする語彙・構文的言い換えの中には、我々が語彙構成的言い換えと呼んだような構成的に計算できる可能性が高い言い換えも少なくない。こうした種類の言い換えについては、個別の語の統語的あるいは意味的特性に関する知識と一般性の高い原理的な変換規則によって実現することが望ましい。語彙構成的言い換えが構成要素の語彙的知識から組み合わせ的に計算できるとすれば、少なくともその種類の言い換えについては、人手で開発・保守できる規模の語彙資源でカバーすることができるだろう。

さらに、構成的言い換えの計算モデルが確立すれば、それを使って非構成的言い換えの知識獲得を効率化することが可能になる。たとえば、コーパスから何らかの方法で(7)のような言い換え事例が獲得できたとしよう。

(7) 財政再建が急務の課題である

⇔ 緊急に財政の立て直しをはかる必要がある

現在の技術では、「財政」という名詞以外に構成要素間の対応をとるのが困難なため、この事例から一般的で有用な言い換え知識を獲得するのは難しい。しかし、もし仮に(8a)~(8c)のような言い換え(これらはいずれも語彙構成的言い換えと考えられる)の対応を別の方法で計算することができれば、同じ言い換え事例(7)から、(8d)のように非構成的言い換えに関する、より一般的で有用な知識が獲得できる可能性が出てくる。

(8) a. 財政再建 ⇔ 財政を再建する

b. 財政を再建する ⇔ 財政を立て直す

c. 財政を立て直す ⇔ 財政の立て直しをはかる

d. 〈行為 X 〉_n が急務の課題である ⇔ 緊急に〈行為 X 〉_v 必要がある

以上の考察から、次のような3段階の研究シナリオを考えることができる。

1. まず、語彙構成的言い換への計算を工学的に実現するための方法論を研究し、必要となる語彙資源を開発する。
2. つぎに、1の成果に基づいて、所与の文のペアから言い換え関係(同義関係)にある構成素の対応を特定する言い換え認識機構を構築する。
3. 最後に、コーパスからの言い換え知識獲得にこの言い換え認識機構を導入し、少数の訓練事例からより多くの有用な言い換え知識を抽出する新しい知識獲得方式を開発する。

こうしたシナリオを現実のものとするためにも、語彙構成的言い換への計算モデルの構築は優先されるべき課題と言える。

さて、語彙構成的言い換えが構成要素の統語的・意味的特性の組み合わせによって計算できるとすれば、語の統語的振る舞いを意味と対応づける理論的枠組みとして近年目覚ましい発展を遂げた語彙意味論を言い換え計算の

道具に使うのは自然な選択であろう。以下に述べる語彙意味論的アプローチでは、例えば「...に動揺を与える ⇔ ...を動揺させる」という言い換えを、(a)「動揺(する)」、「与える」など個別の語の統語・意味特性と(b)一般性の高い計算手順に分解して実現する。(a)の語彙知識は、個別の言い換えに依存しない知識であり、同じ知識が格・態の交替、統語範疇間の交替、複合語の言い換えなど、様々な種類の有用性の高い言い換えに共通に利用できる。我々の取り組みは、この命題の現実性を大規模な工学実験によって証明する試みであると言える。

語の意味を記述する枠組みにはいろいろな提案があるが、我々は手始めに語彙概念構造(LCS)を採用することにした。LCSは要素還元型の意味記述モデルであり、語の意味をプリミティブな意味特性(意味述語)に分解して記述する。個々の意味特性は特定の統語的振る舞いと対応するように設計される。語の振る舞いは語の意味と強い相関があるので、語の意味特性が分かれば、その語の振る舞いがかなり予測できる。

4.2 語彙概念構造にガイドされた統語構造変換

語の意味記述にLCSを採用することは、必ずしも意味構造変換方式や中間言語方式の採用を意味するわけではない。LCSで表現できる意味は語の意味の一部に過ぎないし、3.1節で述べたように文の意味を完全に表現できる意味表現を設計すること自体が極めて困難な課題である。そこで、意味構造変換に一足飛びに移行するのではなく、まずは統語構造レベルでの構造変換の枠組みを仮定し、語彙概念構造から得られる意味情報を利用して変換処理を制御することを考える。

動詞の自他交替を例として考えよう。よく知られるように、自他交替に伴って生じる格構造の交替は動詞の種類によって異なる。たとえば、(9a)の他動詞文のガ格は自動詞文でデ格あるいは二格に交替するが、(9b)のそれは自動詞文のガ格の修飾語に替わる。

- (9) a. 風が洗濯物を揺らす ⇒ 風で/に洗濯物が揺れる
 b. 風が勢いを弱める ⇒ 風の勢いが弱まる

こうした振る舞いの違いは交替に参加する動詞の LCS から予測することができる(影山 1996)ので、自他交替のための知識を個別に記述する必要はない。そこで、自他交替の知識としては、(10a)のように、他動詞をそれに対応する自動詞に置き換える手続きだけを記述しておき、それに伴う(10b)の格構造の交替のような処理は LCS 上での一般的な手続きによって計算する、という枠組みを考える。

- (10) a. 他動詞 V_t をそれに対応する自動詞 V_i に置き換える。
 b. V_i の補語の表層格を再計算する。

本論文では、(10a)のような統語構造上の変換規則を便宜的に統語構造変換スキームと呼ぶ。

機能動詞構文を言い換える場合にも言い換え方の曖昧性が観察される。(11)の3例はともに動詞「与える」のヲ格の動作性名詞を主動詞に据えるように言い換えているが、(11a)、(11b)が能動態になるのに対して(11c)では受動態になる。また、例(11a)、(11c)のように、動詞のヴォイスに関わらず、格助詞の修正が必要になる場合もある。

- (11) a. 友人の活躍が息子に刺激を与える
 ⇒ 友人の活躍が息子を刺激する
 b. 株価の変動が為替に影響を与える
 ⇒ 株価の変動が為替に影響する
 c. その映画が彼に感動を与える
 ⇒ その映画が彼を感動させる

そこで、この種の言い換えについても、(12)のような統語構造変換スキームだけを記述しておき、機能動詞構文の主辞と動作性名詞の動詞形の間で意味が変化しないように補語の表層格を別途計算する枠組みを考える。

- (12) a. 動詞 V のある補語 X_n をそれに対応する動詞 X_v に置き換える。
- b. X_v 以外の V の補語の係り先を X_v に変更し、 V を削除する。
- c. X_v の態および X_v の補語の表層格を再計算する。

図1に統語構造変換スキームと語彙概念構造による言い換え生成の模式図を示す。まずは入力文を解析し、統語構造を得る。さらにLCS解析も行い、入力文中の各述語についてLCSに基づく項構造を得る。次に、統語構造変換スキームによって統語構造の一部を書き換える。統語構造変換スキームは必要な構造変換を完全には指定しない。たとえば、例(11)にあげた機能動詞構文に統語構造変換スキーム(12a-b)を適用した時点では、次の例の下線部に示すように言い換え先の表現の態および格助詞は未指定である(図中“?”, 例中“_”で示してある箇所)。

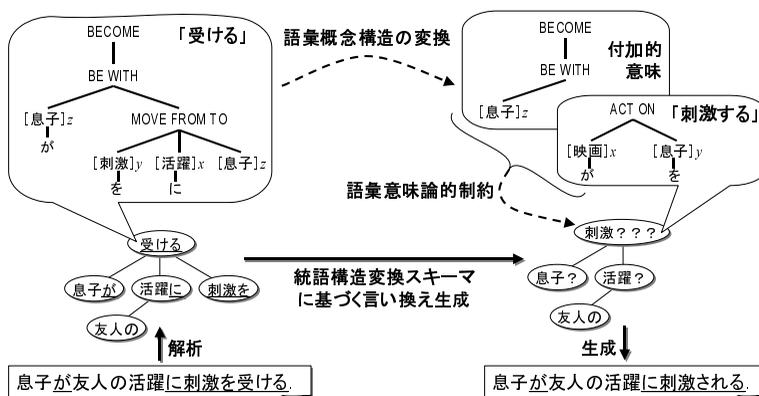


図1: 統語構造変換スキームと語彙意味論的制約に基づく言い換え生成モデル

- (13) a. 友人の活躍が息子に刺激を与える
 ⇒ 友人の活躍 _ 息子 _ 刺激 _
- b. 株価の変動が為替に影響を与える
 ⇒ 株価の変動 _ 為替 _ 影響 _
- c. その映画が彼に感動を与える
 ⇒ その映画 _ 彼 _ 感動 _

これらを LCS の情報に基づいて決定し、最後に統語構造から表層表現を生成する。

この枠組みで語彙構成的言い換えを実現するために、統語構造変換スキーマおよび LCS 以外の語彙的な知識も用いる。たとえば、上にあげた動詞の自他交替では自動詞と他動詞の対応に関する知識、機能動詞構文の言い換えでは動作性名詞と動詞形の対応に関する知識を仮定している。他にも、5.2 節で例示する (14d) のような動詞交替のためには「売る」と「買う」のような授受の関係にある動詞の組、(14e) のような統語範疇間の交替のためには「暖まる」と「暖かい」のような派生語の組に関する知識が必要である。WordNet² や CatVar³ など、英語についてはこのような知識を含む大規模なレキシコンがすでにあり、様々な自然言語処理タスクに用いられている。日本語に関してもこのような知識を整備すれば言い換えや語彙論研究の基盤として役立つと期待できる。

上述のような知識集約的な枠組みがうまく機能するかどうかは、実用規模の知識ベースを現実的なコストで構築できるかに大きくかかっている。この点について、我々は次のような LCS の利点をそのまま継承することができる (竹内ら 2005, 2006)。

客観性 LCS では、語の統語的振る舞いをその語が持つ意味的な特性から予測する。ある語がある意味特性を持つかどうかは、その意味特性に対応する統語的振る舞いをテストすることによって客観的に判断できる。すなわち、LCS の仕様をうまく設計すれば、LCS 辞書を構築する過程に作業者の恣意性が入りにくくなる。

分類軸の独立性 ある意味特性を固定したまま別の特性を細分化したり、定義修正したりすることができる。このことは辞書の漸進的な洗練が可能になることを意味し、辞書の開発に有利である。

予測能力 意味特性の決定には、それに対応する統語的振る舞いの一部をテストするだけでよい。残りの未教示の統語的振る舞いは意味特性から予測できる。

研究の蓄積 言い換え生成にとって重要な統語的振る舞い(アスペクトや態に関する特性、動詞交替、複合語など)と意味特性の関係がLCSの研究者らによってすでに活発に調査されており、その成果を利用することができる。

先に言い換え生成は本質的には機械翻訳と同じ問題であると述べた(3.1節)が、機械翻訳の領域では意味構造変換を指向する研究でさえ未だに語彙意味論研究の成果を採り入れる試みがほとんど見られない。これは、異なる言語の間で意味特性の対応をとるところに高いハードルがあったためと推測できる。これに対し、言い換え生成の場合は同一言語内の意味構造だけを扱えばよいという点でこの問題を当面は避けて通ることができる。したがって、語彙意味論的な概念を採り入れることが機械翻訳の場合よりも容易であると期待できる。

語彙意味論的アプローチの最大の問題は、LCS辞書のような語彙資源の開発・保守のコストが高いことであろう。しかし、LCSのような語彙知識は、言い換えのような特定のタスクに限らず、意味処理を必要とする様々な自然言語処理技術において有用な資源となるであろうし、言い換えに限っても、その応用横断的な有用性を考えれば、ある程度のコストを支払っても意味構造変換の実現を目指すことの意義は機械翻訳の場合よりもさらに大きい。

5 3つの部分問題

前節で述べた枠組みを工学的に実現するためには、統語構造の解析および同構造からの生成の他に、少なくとも次の3つの部分問題に取り組む必要がある：

1. 頑健で高精度な述語項構造（LCS）解析の実現
2. LCS による構造変換の制御
3. 変換後の文の適格性の判定

以下、それぞれについて我々の試みを紹介する。

5.1 LCS 解析

3.2 節で述べたように、LCS のタグを付与した大量の正解事例を用意できれば、高い精度で LCS 解析を実現することができる。そこで、現実的には、正解事例の収集コストをいかに削減できるかが重要な課題である。この問題に対する有効なアプローチの一つは事例間の類似性を考慮することである。

たとえば「女性が悲鳴をあげる」と「こどもが大声をあげる」という2つ用例があったとしよう。これらは、格の出現パターンやそれぞれの要素が似ているため、同じタイプの LCS を持つ可能性が高い。同様に、動詞が異なる場合でも、それらが意味的に近いものであれば、「大統領が98年に計画を公表する」と「首相が10月にプランを発表する」のようにやはり同じタイプの LCS を持つ可能性が高い。

類似した事例がいくつもある場合、それらの一つ一つに個別に正解の LCS を付与するのはあまり賢いやり方ではない。むしろ、同じ手間をかけるなら、類似した事例は一括りにしてその代表例だけに正解タグを付けることにして、一つでも多く新しい種類の事例を調査する方が解析精度に効果的に貢献するだろう。

そこで、次の手順で正解タグを付与することにより事例収集コストを削減することを考える。まず、大規模平文コーパスから抽出した大量の事例集合を上述のような類似性に基づいてクラスタリングする。つぎに、各クラスタについて、そこに所属する事例から代表事例を選び、正解を手で教示する。最後に、同じクラスタ内の他の事例に対し、代表事例のタグに準じたタグを自動的に付与する。また、クラスタ数が多い場合は、解析精度向上に貢献しそうなクラスタから優先的に教示するといった選択的サンプリングも有

効であろう。たとえば、事例数の多いクラスタを優先したり、既教示の事例にあまり似ていないクラスタを優先するといった戦略が考えられる。

我々が行った予備調査(平野ら 2005)では、上述のような工夫によって、それほど多くのノイズ(異なる LCS に対応する事例を同じクラスタに含めてしまうこと)を入り込ませることなく事例を比較的少数のクラスタ数にまとめあげることができた。また、クラスタリングに多少のノイズが入るとしても、用いる事例の数を増やすほど解析精度を改善できることも明らかになった。こうした調査から我々は頑健な LCS 解析に近い将来可能になると考えている。

5.2 LCS に基づく言い換え候補生成

語彙意味論的アプローチが対象とする語彙構成的言い換えは、語の統語的特性と意味的特性に基づいて構成的に説明できると考えられる規則性の高い言い換えである。これには、次のような言い換えが含まれる。

- (14) a. 子供がプレゼントに喜ぶ ⇔ プレゼントが子供を喜ばせる (態交替)
 b. 通りが群衆であふれた ⇔ 群衆が通りにあふれた (場所格交替)
 c. 2 位が先頭との距離を縮めた
 ⇔ 2 位と先頭の距離が縮まった (自他交替)
 d. うちの営業が S 社にコピー機を 5 台売った
 ⇔ S 社がうちの営業からコピー機を 5 台買った (動詞交替)
 e. 部屋は十分暖まっている
 ⇔ 部屋は十分暖かい (統語範疇間の交替)
 f. 父が息子に土地を譲る
 ⇔ 息子が父から土地を譲り受ける (複合動詞)
 g. 財政再建 ⇔ 財政を再建する (複合名詞)
 h. 太郎が犯人であると認める ⇔ 太郎を犯人と認める (補文構文)
 i. 息子が友人の活躍に刺激を受ける
 ⇔ 息子が友人の活躍に刺激される (機能動詞構文)

- j. 新しい機材の必要性を議論する
 - ⇨ 新しい機材が必要かどうかを議論する（名詞接尾辞の着脱）
- k. 厳密に審査基準を定める
 - ⇨ 厳密な審査基準を定める（修飾語の係り先の交替）
- l. リサイクルの効率化が求められる
 - ⇨ 効率的なリサイクルが求められる（主辞交替）

(14a-g)の言い換えは、これまでのLCSの研究(Levin and Rappaport Hovav 1995, 影山 1996, 2001, etc.)が対象としてきた中心的な言語現象であり、その研究成果を直接的に利用することができる。また、(14h)のように補文や引用文をとる構文についても、現在のLCSの枠組みの自然な拡張で扱えるようになると期待できる。(14i)のような機能動詞構文では、動作性名詞(「刺激」)のLCSと主動詞(「受ける」)のLCSが相互に作用して全体の意味を構成していると考えられる。これについては、次節で詳しく述べる。

一方、(14j-l)に見られるような言い換えは、これまでの語彙意味論研究ではほとんど対象とされてこなかった言語現象である。しかし、これらのタイプの言い換えも、かなり広範な語について同様の交替が起こること、ただしそこには語の統語的・意味的特性が制約として働いているように見えることなど、(14a-i)と同様の語彙構成的性質を持っていると考えられる。こうした言い換えは、現在のLCSの枠組みで記述できていない語彙知識は何か、現在のLCSをどのように拡張すればよいか、他にどのような語彙資源が必要か、といった極めて興味深い問題を提供するものと期待できる。

5.3 統計的言語モデルによる適格性判定

語彙意味論的アプローチは、語の意味特性から予測される統語的振る舞いに基づいて語彙構成的言い換えの計算を制御する。個別の語の統語的振る舞いを網羅的に記述する方法では規則の記述量が組み合わせ的に増えてしまうが、語彙意味論的アプローチではこの問題を回避することができる。ただし、意味特性からの予測に従わない例外的な語の振る舞いに対する頑健性が

低いという欠点もある。対処法はいくつか考えられるが、現実的で有効な解決策の一つは、生成された言い換え文の適格性を判定する統計的言語モデルを導入することである。

我々は、名詞・動詞などの内容語の言い換えや態の交替などの様々な種類の言い換え生成実験を通じて、言い換え生成時に生じる誤りの傾向を調査し(藤田・乾 2003)、とくに頻繁に不適格になる動詞とその名詞格要素(補語および付加詞)の共起を対象として統計的言語モデルを構築した(藤田ら 2004)。このモデルは、大規模なコーパスから得られる共起用例(正例)に基づいて構築した確率モデルと、人手で収集した不適格な共起用例(負例)に基づく判別モデルを組み合わせたものであり、(15)の2つの例のように、動詞とその格要素の共起に対して適格性のスコア(カッコ内の数値)を与える。このスコアに適切な閾値を設けることで、(15a)を適格、(15b)を不適格と判定できる。

- (15) a. 失敗を理由に処分を受ける
 ⇒ 失敗を理由に処分される (「理由に処分される」: 0.9607)
 b. 今月中に試験を実施する
 ⇒ * 今月中を試験する (「中を試験する」: 0.5073)

6 例題：機能動詞構文の言い換え

例(14)に示したように比較的多様な言い換えが語彙意味論に基づいて説明できそうである。ただし、言い換えの種類ごとに語のどのような統語的・意味的特性がどのように関わっているかは異なると考えられる。たとえば、動詞の自他交替や場所格交替については、LCSに基づく現象分析の試みがいくらか見られる(影山 1996, 岸本 2001)ものの、説明に用いられている統語的・意味的特性は異なる。本節では、機能動詞構文の言い換えを取り上げ、言い換え生成モデルおよび工学的評価実験の結果について述べる。機能動詞構文において動詞はいくつかの文法的機能を果たすが、ここでは、例(11)のようなヴォイスに相当する機能を持つ機能動詞構文の言い換えを取り上げ

る。この現象は、4.2 節で述べたように、LCS の言い換えへの応用可能性を調査するための適当な例題である。

6.1 言い換え生成モデル

例 (16) を用いて、意味構造変換と表層生成の処理について述べる。

- (16) s. 健が映画に刺激を受ける
LCS_{V1} [BECOME [[健]_z BE WITH
 [[刺激]_y MOVE FROM [映画]_x TO [健]_z]]]
LCS_{N0} [[]_{x'} ACT ON []_{y'}]
LCS_{N1} [[映画]_{x'} ACT ON [健]_{y'}]
LCS_S [BECOME [[健]_z BE WITH []]
 t0. 映画が健を刺激する
 t1. 健が映画に刺激される

6.1.1 意味構造変換

村木 (1991) は、機能動詞構文を言い換える際のヴォイス表現の選択には、動作主体から見た行為の遠心性（動作主体から外部に動作が向かうこと）と求心性（主体へ動作が向かうこと）という動詞に関する 2 種類の特性をあげている。一方 LCS では、上位事象と下位事象の階層関係や、述語項構造における項の順序によって、動詞が表す出来事における要素（補語）間の作用関係、出来事の視点が明示的に表される。そこで、言い換え前後でこれらの情報が変化しないように LCS 上の操作を施す。具体的には、次の 3 ステップで言い換え前後の 2 つの LCS、すなわち機能動詞構文の LCS 解析結果 *LCS_{V1}* と動作性名詞の動詞形に対応する LCS のテンプレート *LCS_{N0}* を対応付ける。

- (i) 意味述語の対応付け まず LCS 中の意味述語を対応付ける。機能動詞構文の言い換えに関わる Agent の行為の方向性をとらえるために、Agent

が関与するか否か，アスペクトを表すかで意味述語を次の3種に分類し，同じ分類の意味述語ならば対応付ける。ただし，アスペクトを表す意味述語は機能動詞構文の言い換えに関与しないため無視する。

- Agent が関与する意味述語：“CONTROL”，“ACT”，“ACT ON”，“ACT TO”，“MOVE FROM TO”⁴
- 状態を表す意味述語：“MOVE TO”，“BE AT”，“BE WITH”
- アスペクトを表す意味述語：“BECOME”

例(16)では LCS_{V_1} の“MOVE FROM TO”と LCS_{N_0} の“ACT ON”を対応付ける。

(ii) 項の対応付け 意味述語を対応付けた後は，その対応に基づいて LCS_{V_1} と LCS_{N_0} の項を作用の受け手あるいは状態変化の結果状態から順に対応付け， LCS_{N_0} の値を埋める (LCS_{N_1})。例(16)では，意味述語の対応に基づいて， z と y' ， x と x' が順に対応付けられ「健」と「映画」が y' ， x' を埋める。ただし，動作性名詞の意味は LCS_{N_0} によって表されているため，対応付けの対象とはしない。

(iii) 余った構造の付加 上述の手続き (i)，(ii) を終えた時点で LCS_{V_1} の部分構造が LCS_{N_1} に対応付けられずに残っている場合がある。この部分構造は， LCS_{N_1} が示す一つの動詞の意味としては表すことができないが，可能ならば何らかの表現を用いて表されるべきである。とくに，LCS の左端の部分構造は，出来事の視点を特定する手がかりであるため，機能動詞構文の言い換えでは無視できない。ゆえに，この余った部分構造 LCS_S を LCS_{N_1} とともに出力し，次の表層生成のステップで用いる。例(16)では， LCS_{V_1} 中の “[BECOME [[健]_z BE WITH []]]” が LCS_S となる。

6.1.2 表層生成

このステップでは、 LCS_{N1} と LCS_S を言語表現生成の制約として用い、統語構造変換スキーマでは未指定のまま残っていたヴォイス表現と補語の表層格を特定する。

まず、 LCS_{N1} に基づいて動作性名詞の動詞形の補語に表層格を割り当てる。この時点では、能動文 (16t0) が生成される。次に、ヴォイス表現を特定し、格交替を施す。ここでは、 LCS_{N1} と LCS_S がどのような意味述語を持っているか、項に共通の要素を持っているか、などを手がかりとする。たとえば、例 (16) では、 LCS_{N1} の作用の受け手にあたる「健」が、 LCS_S によって出来事の視点に据えられていると解釈する。そこで、ヴォイス表現を「される」とし、 LCS_S の当該項の要素（例では「健」）がガ格になるように受動化を施す。この解釈は長谷部 (2005) が述べた統語的な態交替に等しい。他にもたとえば、 LCS_S が Agent の項を持ち、その項が LCS_{N1} のいずれの項とも異なる要素を持つ場合、その要素を他動使役の使役者とみなしてヴォイス表現を「させる」とし、格構造に使役化を施す、などの規則を 5 種類作成した。詳細は別稿 (Fujita et al. 2005) を参照されたい。

6.2 言い換えの生成実験

6.2.1 語彙および評価用データ

提案する言い換え生成モデルの性能を評価する実験を行った。実験に用いる語の選定、および評価用データの作成手順の概要を以下に示す。詳細な手順は別稿 (Fujita et al. 2005) を参照されたい。

動作性名詞 竹内 (2004) の T-LCS 辞書⁵ (バージョン 0.95) 中の 1,165 語を動作性名詞とした。これらは、「売り」のような動詞の連用形、「刺激」のような漢語由来のサ変名詞、および「ドライブ」のような英語からの借用語で構成されている。なお、複数の語義を持つ語には、語義ごとに LCS が付与されている。

機能動詞構文パターン（動詞と格助詞の組）村木（1991）は機能動詞構文を形成しうる数十語の動詞を示しているが、これらは必ずしも優先して分析すべき頻出語とは限らない。また、同じ動詞でも、動作性名詞の表層格に依存して、機能動詞構文にならない場合や文法的機能が異なる場合もある。そこで、「を与える」、「を受ける」、「に取り組む」のように機能動詞構文を形成しうる動詞と格助詞の組を一つの機能動詞構文パターンとする。新聞記事コーパス中の頻度上位 50 組について機能動詞構文を形成しうるか否かを判定した結果、40 組が可能とみなされたので、これらに LCS を付与した。

評価用言い換え事例集 機能動詞構文とそれに対応する言い換え文の組（言い換え事例）も上と同じコーパス中の頻度を参照して収集した。まず、上述の動作性名詞と機能動詞構文パターンの組み合わせの頻度上位 245 組を取り出し、次に各組について、それを含むような動詞格構造の頻度上位 3 例（単文）、合計 735 例を取り出した。このうち、人手で言い換え文を付与できた 711 文を用いて言い換え生成モデルの性能を評価する。

6.2.2 言い換え生成と評価

提案モデル（表 1 の LCS）は、入力中の機能動詞あるいは動作性名詞に対して複数の LCS が付与されている場合、機能動詞の LCS と動作性名詞の LCS のあらゆる組み合わせについて 6.1 節で述べた手続きを試行し、可能な言い換え候補を全て出力する。さらに、5.3 節で述べたように、LCS を参照して生成された言い換えを適格と不適格に分類するために言語モデルを適用する（表 1 の LCS+LM）。他方、ヴォイス表現「する」、「される」、「られる」の中のいずれか、および格助詞「が」、「を」、「に」が重複しないようなあらゆる組み合わせの中から確率的に最適な組合せを選択する言語モデル（表 1 の LM）を構築する。ここでは、5.3 節で述べた確率モデルを単独で用いる。このモデルは、学習用のコーパス中に現れない表現が入力された場合は確率を計算できないため、いかなる言い換えも出力しない。

評価用データ 711 文を言い換え生成モデルに入力したときに得られる言い換え文を，人手で作成した言い換え文と比較して適格な言い換え文と不適格な言い換え文に分類する。ここでは，正しいヴォイス表現を選択し，それぞれの補語に表層格が正しく付与されていれば，格要素の語順が異なっても適格と判定する。評価には次の 2 つの指標を用いた。再現率とは求める言い換え文をどれだけ自動的に生成できたかを図る指標，精度とは自動的に生成した言い換え文のどれだけが適格かを図る指標である。

$$\begin{aligned} \text{再現率} &= \frac{\text{適格な言い換え文の数}}{\text{人手で作成した言い換え文の数}} \\ \text{精度} &= \frac{\text{適格な言い換え文の数}}{\text{自動生成された言い換え文の数}} \end{aligned}$$

各モデルの評価結果を表 1 に示す。提案モデルは言語モデルによって最尤の表現を選択するのに比べて顕著に高い性能を示した。このことから，LCS が表す動詞の語彙特性はヴォイス表現と補語の表層格を定めるのに有効であると言える。また，提案モデルが生成する誤りは，言語モデルと組み合わせることで精度良く解消できることが確認できた。

表 1: ベースラインおよび提案モデルの性能比較

モデル	LCS	LM	LCS+LM
自動生成された言い換え文の数	798	547	717
適格な言い換え文の数	624	322	609
不適格な言い換え文の数	174	225	108
再現率 (%)	87.8	45.3	85.7
精度 (%)	78.2	58.9	84.9

7 まとめ

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，あるいは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

は、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、近年言語処理研究者の関心を集めてきた。

言い換えは意味の等価性に基づいて行う統語的操作と見なせるので、語の統語的振る舞いをその意味的特性で説明しようとする語彙意味論は、言い換え研究に極めて有用な基盤を提供すると期待できる。本論文では、語彙意味論に基づく言い換えモデル構築の意義と方法論を論じ、LCS(述語項構造)解析、LCSに基づく構造変換、言い換え文の適格性判定の各部分問題について我々の具体的な取り組みを紹介した。このアプローチを現実のものとするためには、もう一つ、大規模な語彙概念構造辞書の構築という大きな仕事を成し遂げる必要があるが、これについても竹内論文(本号)で報告するように、具体的な作業を進めている。今後は、6章で述べたような方法論を5.2節で例示したようなより広範な種類の語彙構成的言い換えに展開し、既存の語彙概念構造理論で扱える範囲はどこまでか、残りの言い換えをカバーするにはこれをどのように拡張する必要があるかを工学的に検証していきたい。

言うまでもなく語彙意味論は現在も発展中の理論群である。影山(2005)や小野(2005)の議論にもあるように、語彙概念構造論と生成語彙論の融合の可能性も極めて興味深い。我々はこうした最新の研究成果を工学的な計算モデルに採り入れる努力を惜しむべきでない。それとともに、工学的実験を通して得られる知見が語彙意味論研究への何らかのフィードバックになるとすれば幸いである。

[謝辞] 本論文執筆の機会を与えてくださった影山太郎氏(関西学院大学)と伊藤たかね氏(東京大学)に深く感謝いたします。伊藤氏には執筆中にも有益なコメントをいただきました。また、竹内孔一氏(岡山大学)にはT-LCS辞書の利用を快諾いただき、LCSに関する様々なご教示をいただきました。記して深く感謝いたします。

注

(1) 統計的機械翻訳の有効性は機械翻訳に関する評価型ワークショップ(例えば、<http://www.nist.gov/speech/tests/mt/>)などでも示されている。

(2) WordNet : <http://wordnet.princeton.edu/>

- (3) The Categorical Variation Database : <http://clipdemos.umiacs.umd.edu/catvar/>
- (4) 意味述語“MOVE FROM TO”は文献(影山 2001)で「教える」と「教わる」のような動詞対の授受の関係を説明するために導入された。竹内らの LCS 辞書(竹内ら 2002, 竹内 2004)でもこれを踏襲しているが, これは“MOVE TO”とはアスペクト特性を異にする概念である。我々は, 機能動詞構文の言い換えへの応用にあたって“FROM”の項から“TO”の項への作用を動作主性を持つ作用とみなし, 試験的に, “FROM”の項を Agent, “MOVE FROM TO”を Agent が関与する意味述語とした。
- (5) 語彙概念構造辞書 : <http://cl.it.okayama-u.ac.jp/rsc/lcs/>

参考文献

- Baker, Collin F., Fillmore, Charles J., and Lowe, John B. (1998) The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL)* 86–90.
- Barzilay, Regina (2003) *Information fusion for multidocument summarization: paraphrasing and generation*. Ph.D. thesis Graduate School of Arts and Sciences, Columbia University.
- Canning, Yvonne, and Tait, John (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 Customized Information Delivery* 6–11.
- Dorna, Michael, Frank, Anette, van Genabith, Josef, and Emele, Martin C. (1998) Syntactic and semantic transfer with F-structures. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL)* 341–347.
- Dorr, Bonnie Jean (1997) Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation* 12 (4) 271–322.
- Dras, Mark, and Yamamoto, Kazuhide (Eds.) (2005) *The 3rd International Workshop on Paraphrasing (IWP)*. IJCNLP-2005 Workshop.
- 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) 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) 1176–1187.
- Fujita, Atsushi, Inui, Kentaro, and Matsumoto, Yuji (2005) Exploiting Lexical Conceptual Structure for paraphrase generation. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP)* 908–919.

- 長谷部郁子 (2005) 「日英語の非対格構造の統語的使役化」, 影山太郎 (編) 『レキシコンフォーラム No.1』 103–131. ひつじ書房.
- 平野徹, 飯田龍, 藤田篤, 乾健太郎, 松本裕治 (2005) 「動詞項構造辞書への大規模用例付与」, 『言語処理学会第 11 回年次大会発表論文集』 396–399.
- Inui, Kentaro, and Hermjakob, Ulf (Eds.) (2003) *The 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)*. ACL-2003 Workshop.
- Inui, Kentaro, Fujita, Atsushi, Takahashi, Tetsuro, Iida, Ryu, and Iwakura, Tomoya (2003) Text simplification for reading assistance: a project note. In *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)* 9–16.
- 乾健太郎, 藤田篤 (2004) 「言い換え技術に関する研究動向」, 『自然言語処理』 11 (5) 151–198.
- 影山太郎 (1996) 『動詞意味論—言語と認知の接点』 くろしお出版.
- 影山太郎 (編) (2001) 『日英対照 動詞の意味と構文』 大修館書店.
- 影山太郎 (2005) 「辞書的知識と語用論的知識—語彙概念構造とクオリア構造の融合にむけて」, 影山太郎 (編) 『レキシコンフォーラム No.1』 65–101. ひつじ書房.
- 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』 34 巻 77–82.
- 岸本秀樹 (2001) 「壁塗り構文」, 影山太郎 (編) 『日英対照 動詞の意味と構文』 100–126. 大修館書店.
- Lapata, Maria, and Brew, Chris (1999) Using subcategorization to resolve verb class ambiguity. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP-VLC)* 266–274.
- Levin, Beth, and Rappaport Hovav, Malka (1995) *Unaccusativity: at the syntax-lexical semantics interface* Linguistic Inquiry Monograph 26 The MIT Press.
- Lin, Dekang, and Pantel, Patrick (2001) Discovery of inference rules for question answering. *Natural Language Engineering* 7 (4) 343–360.
- Lønning, Jan Tore, Oepen, Stephan, Beermann, Dorothee, Hellan, Lars, Carroll, John, Dyvik, Helge, Flickinger, Dan, Johannessen, Janne Bondi, Meurer, Paul, Nordgård, Torbjørn, Rosén, Victoria, and Velldal, Erik (2004) LOGON. A Norwegian MT Effort. In *Proceedings of the Workshop in Recent Advances in Scandinavian Machine Translation*.
- Manning, Christopher D., and Schütze, Hinrich (1999) *Foundations of statistical natural language processing* The MIT Press.
- Mitamura, Teruko, and Nyberg, Eric (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* 1–12.

- 村木新次郎 (1991) 『日本語動詞の諸相』 ひつじ書房.
- 小野尚之 (2005) 『生成語彙意味論』 くろしお出版.
- Palmer, Martha, Gildea, Daniel, and Kingsbury, Paul (2005) The Proposition Bank: an annotated corpus of semantic roles. *Computational Linguistics* 31 (1) 71–106.
- Ravichandran, Deepak, and Hovy, Eduard (2002) Learning surface text patterns for a question answering system. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* 215–222.
- 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) 2937–2945.
- Sato, Satoshi, and Nakagawa, Hiroshi (Eds.) (2001) *Workshop on Automatic Paraphrasing: Theories and Applications*. NLP RS-2001 Workshop.
- Shinyama, Yusuke, and Sekine, Satoshi (2003) Paraphrase acquisition for information extraction. In *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)* 65–71.
- 竹内孔一, 乾健太郎, 藤田篤 (2006) 「語彙概念構造に基づく日本語動詞の統語・意味特性の記述」, 本書所収.
- 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) 1446–1456.
- 竹内孔一 (2004) 「語彙概念構造による動詞辞書の作成」, 『言語処理学会第10回年次大会発表論文集』 576–579.
- 竹内孔一, 乾健太郎, 藤田篤, 竹内奈央, 阿部修也 (2005) 「分類の根拠を明示した動詞語彙概念構造辞書の構築」, 『情報処理学会研究報告, NL-169-18』 123–130.
- 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) 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) 99–117.

————— (乾) 奈良先端科学技術大学院大学
(藤田) 名古屋大学

Abstract**A Lexical Semantics-Based Approach to Computational Modeling of
Lexico-Syntactic Paraphrasing****Kentaro INUI and Atsushi FUJITA**

Paraphrases are alternative ways of conveying the same content. The language technology for paraphrasing has drawn the attention of an increasing number of researchers because of its potential contribution to a wide variety of natural language applications. Given that paraphrasing can be seen as a syntactic transformation process guided by semantic equivalence, the paraphrasing technology should be able to take full advantage of recent research in lexical semantics since this linguistic subfield aims at discovering and modeling the relationship between the syntactic behavior of words and their semantic properties. Motivated by this research context, the paper provides technical background of the problem and discusses how to incorporate recent advances in lexical semantics, more specifically, findings from recent studies on Lexical Conceptual Structure (LCS), into a computational model of lexico-syntactic paraphrasing. The paper also reports on our ongoing studies addressing three major issues involved in this lexical semantics-based approach: LCS parsing, LCS-guided syntactic transformation, and candidate paraphrase selection.