

語彙概念構造に基づく言い換え生成 —機能動詞構文の言い換えを例題に

藤田 篤[†] 降幡 建太郎^{††}
乾 健太郎^{†††} 松本 裕治^{†††}

本論文では、語彙概念構造(LCS)に基づく言い換え生成モデルを提案する。LCSは語彙意味論の体系の一つであり、動詞の様々な統語・意味的な特性に対応する意味述語を組み合わせて動詞の意味を記述する枠組みである。提案するモデルは、LCSが表す意味構造の変形操作および語彙意味論的制約に基づいて言い換えを生成する。本論文では、例題として日本語の機能動詞構文の言い換えを取り上げ、言い換えの生成実験を通じてモデルおよび既存のLCS辞書の性能を検証する。

Paraphrase Generation Based on Lexical Conceptual Structure: A Case Study on Paraphrasing of Light-verb Constructions

ATSUSHI FUJITA,[†] KENTARO FURIHATA,^{††} KENTARO INUI^{†††}
and YUJI MATSUMOTO^{†††}

Lexical Conceptual Structure (LCS), a theory of lexical semantics, represents verbs as semantic structures using a small number of semantic predicates each of which denotes a particular property of verbs. The verb classification given by LCS has been used to explain regularities underlying lexical and syntactic paraphrases, such as verb alternation, compound word decomposition, and lexical derivation. This paper describes our LCS-based approach to paraphrase generation, and reports on our empirical experiments taking paraphrasing of Japanese light-verb constructions as an example class of paraphrases. Experimental results confirm that syntactic and semantic properties of verbs represented by LCS are beneficial to semantically constrain syntactic transformation in paraphrase generation.

1. はじめに

自然言語処理の様々なタスクにおいて、近年、同じ意味を表す異なる表現を扱うこと、すなわち言い換えを生成したり認識したりする処理が注目されている¹⁰⁾。言い換え生成の側面では、これまで、テキスト簡単化による読解支援への効果や機械翻訳のカバレッジや訳質の向上が報告されている。また、言い換え認識の側面では、質問応答における解の探索や複数文書要約における同義箇所同定において表現の多様性を吸収することの効果も報告されている。

言い換えの中には、単語とその同義表現、慣用表現

とその説明表現など、個々の表現に依存した現象があるが、抽象的なパターンで表現できる規則的な現象も多い。たとえば、例(1)、(2)においては、複合名詞から動詞句への言い換えという同じ変換パターンが見られる。同様に、機能動詞構文を含む文の言い換えも例(3)、(4)に見られるように類似した変形からなる。

- (1) s. いつも正確な機械操作。
t. いつも正確に機械を操作する。
- (2) s. コスト削減は企業の大きな課題だ。
t. コストを削減することは企業の大きな課題だ。
- (3) s. 友人の活躍が息子に刺激を与えた。
t. 友人の活躍が息子を刺激した。
- (4) s. 株価の変動が為替に影響を与えた。
t. 株価の変動が為替に影響した。

例文中、s., t. は各々言い換え前の文、言い換え後の文を表す。また、自動生成した言い換え文が文法的・意味的に不適格な場合は記号“*”を、文法的・意味的に適格でも言い換えとして適切でない場合は記号“≠”を記す。

[†] 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University

^{††} 株式会社東芝研究開発センター
R&D Center, Toshiba Corporation

^{†††} 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

ただし、これらの例に見られる規則性は、言い換え前後の表現の形態素情報や依存構造を参照するだけではとらえられない。たとえば、「機械翻訳」という複合名詞を例(1)、(2)と同じように「機械を翻訳する」とは言い換えられない。同様に、例文(5s)中の機能動詞構文を例(3)、(4)と同じく(5t)のようには言い換えられず、(5t')のように使役態を表す接尾辞「させる」を後続させる必要がある。さらに、(3t)と(5t')については格助詞の変更も必要である。

(5) s. その映画が彼に感動を与えた。

t. *その映画が彼に感動した。

t'. その映画が彼を感動させた。

言い換えの中には、表現を構成する個々の語の統語・意味的特性に依存して言い換えの可否や言い換え方が決まる上のような現象が少なくない。上に示した例以外にも、種々の動詞交替や語彙の派生などはその範疇に含まれる¹⁰⁾。すなわち、このような語彙構成的な言い換えを生成、認識するためには、語の統語・意味的特性の体系化と記述が欠かせない。

語の統語・意味的特性を記述する枠組みとして、近年、語彙意味論と呼ばれる理論群が発展を見せている。語彙意味論の枠組みの一つである語彙概念構造(Lexical Conceptual Structure; LCS)¹³⁾は、語の統語的振舞いに基づいてその語の意味的特性を分析し、その結果を意味述語で表す。そして、意味述語を階層的に組み合わせた述語項構造によって語全体の意味を記述する。Dorr³⁾は、Levin²⁰⁾の動詞分類に基づいてLCSの体系を定義し、約4,000語の英語動詞にLCSを付与した辞書を構築した。Habashら⁷⁾はDorrの体系に沿ってスペイン語、中国語のLCS辞書を構築し、機械翻訳や言語生成の中間言語として用いている。日本語についても、大石ら²⁴⁾、影山¹⁴⁾の包括的な分析に加え、種々の動詞交替、複合名詞の語構成、語彙の派生などの現象の分析において、語の統語的振舞いの分類や表現間の意味関係の説明に用いられている^{12),15),28)}。電子的な資源としては、竹内ら^{28),29)}が影山の分析に基づいて整備した日本語語彙概念構造辞書(T-LCS辞書、サ変動詞を中心に約1,200語を収録)が無償で利用可能である。また、さらなる体系化、大規模化に関するいくつかのプロジェクトも進行中である^{18),30)}。これらをふまえると、近い将来、実用規模の日本語LCS辞書が利用可能になると期待できる。すなわち、生成語彙論²⁷⁾、フレーム意味論¹⁾な

どなどの語彙意味論の枠組みに比べて注目に値する。以上より、本論文では、

LCSが持つ情報が言い換え生成のどの場面でどのように役立つか

を整理し、

言い換え生成という実践的側面から見たLCS体系・辞書の改善案

を論じる。ただし、様々な種類の言い換えを一度に実現することは困難なので、例題として例(3)~(5)のような機能動詞構文の言い換えを取り上げる。機能動詞構文の言い換えは、動詞の格構造や態に影響するなど、言い換え生成における一般的な問題を多く含んでいる。さらに、先行研究で扱われてきた現象とは異なり、複数の動詞要素の統語・意味的特性の相互作用によって言い換え方が定まるといった性質を持つため、LCSの効用を調査するうえでも新規の知見を期待できる。

以下、まず2章では本論文で用いる日本語LCS辞書の体系およびその体系で示されている動詞の特性について述べる。次に、3章では機能動詞構文の言い換の概形および構成語の統語・意味的特性の相互作用について述べる。4章ではLCSに基づく機能動詞構文の言い換えモデルについて、5章では言い換え生成実験と提案モデルの評価結果について述べる。6章で言い換え生成誤りの原因および提案モデル、LCS辞書の改善点を示し、7章で関連研究に言及する。最後に8章でまとめる。

2. 動詞の意味表現—語彙概念構造

2.1 動詞の振舞いの意味記述

語彙概念構造(LCS)¹³⁾は語の意味を記述する語彙意味論の枠組みの一つである。具体的には、語の統語的振舞いに基づいてその語の意味的特性を分析し、意味的特性を表す意味述語を階層的に組み合わせた述語項構造によって語全体の意味を記述する。これまでの記述の対象はほとんどが動詞であり、文献(14)、(28)などに共通して、表1に示すようなLCSが用いられている。例文中の各補語の添字は意味役割を表す。

本論文ではT-LCS辞書^{28),29)}を採用する。T-LCS辞書は影山¹⁴⁾が提案する意味述語体系および動詞の統語的振舞いを観察するための言語テストを踏襲して作成されている。任意の動詞に対するLCSは言語テストに基づく決定木に従って付与されており、辞書の一貫性を見直したり辞書を拡張したりする際に都合が良い。以下、LCSによって表される語の特性のうち、機能動詞構文の言い換えに利用する次の(a)~(c)について、表1の例を参照しながら詳述する。

表 1 T-LCS 辞書中の LCS の例
Table 1 Examples of LCS in T-LCS dictionary.

ID	LCS LCS のインスタンス	動詞の例 例文
10	[y BE AT z] [[本棚]y BE AT [僕の家]z]	位置する, 存在する, ある 僕の家に(Goal) 本棚が(Theme) ある.
6	[BECOME [y BE AT z]] [BECOME [[本棚]y BE AT [僕の家]z]]	分布する, 飽和する, 届く 僕の家(Goal)に 本棚が(Theme) 届く.
2	[x CONTROL [BECOME [y BE AT z]]] [[店員]x CONTROL [BECOME [[本棚]y BE AT [僕の家]z]]]	訳す, 紹介する, 届ける 店員が(Agent) 僕の家(Goal)に 本棚を(Theme) 届ける.
1	[x ACT ON y] [[監督]x ACT ON [選手]y]	操作する, 運転する, 褒める 監督が(Agent) 選手を(Theme) 褒める.
7	[y MOVE TO z] [[ブログ]y MOVE TO [社会]z]	移動する, 遷移する, 浸透する ブログが(Theme) 社会(Goal)に 浸透する.

- (a) 動詞が表す出来事中の要素間の作用関係
(b) 動詞が表す出来事の視点
(c) 述語項構造と表層格構造の対応関係

2.1.1 動詞が表す出来事中の要素間の作用関係

動詞の統語的振舞いを観察するために, たとえば次のような言語テストが用いられる^{14),18),24),29)}.

- どのような格構造をとるか.
- 目的語を主語として受身文を作れるか否か.
- 「～分間」などの特定の副詞と共起するか.
- 「つづける」を後続したときの解釈は動作の継続か, 動作の繰返しか.

LCS では, これらの言語テストの結果を統合して示される語の意味的特性を “CONTROL”, “BECOME”, “BE AT” などの意味述語で表す. そして, それらの意味述語を組み合わせて動詞の意味を記述する.

LCS 中の x, y, z は動詞が表す出来事に関わる項であり, それを項とする意味述語に基づいて意味役割を与えられる. たとえば, “BE AT” の第 1 項 (記述の文法上左側) は Theme を, 第 2 項 (同右側) は Goal を表す (LCS10, LCS6, LCS2). 同様に, “CONTROL”, “ACT” の第 1 項は Agent (働きかけの動作主) を表す (LCS2, LCS1). ただし, x, y などは項のラベルにすぎない.

LCS は, 単なる意味特性の束ではなく, 述語項構造の階層的な組合せによって動詞が指す概念的意味の範囲を記述する. たとえば, LCS10, LCS6, LCS2 の間には包含関係がある. LCS2 において, 動作主の働きかけ (“CONTROL”) は, 状態変化 (“BECOME BE AT”) の外側 (左側) に記述されている. このような上位事象と下位事象の階層構造によって, 動作主からそれによって引き起こされる事象への作用関係が表されている. また, 下位事象を示す意味述語に関しては, 動詞の完了相 (“BECOME BE AT”), 非完了相 (“MOVE TO”) にかかわらず, Theme を左側 (第 1

項), Goal を右側 (第 2 項) にとっており, 一貫した作用関係を示す.

2.1.2 動詞が表す出来事の視点

LCS10 の例文は (6a) のように書き換えることができ, LCS10 は (6b) のようにも表せる¹⁴⁾.

- (6) a. 僕は (僕の家に) 大きな棚を持っている.
b. [[僕]z BE WITH [[本棚]y BE AT [僕の家]z]]
LCS10 と (6b) の比較から, “[z BE WITH []]” という部分構造が z を視点に据えていると解釈できる. また, LCS6 と LCS2 を比較すると, “[x CONTROL []]” も同様に動詞に含まれる視点を表す部分構造とみなせる. ここでは x が視点を表している.

文献 16) では, 例 (7) に示すような授受の動詞の交替が LCS を用いて説明されている.

- (7) s. 先輩が後輩に合格の秘訣を教える.
[[先輩]x CONTROL
[[秘訣]y MOVE FROM [先輩]x TO [後輩]z]]
t. 後輩が先輩に合格の秘訣を教わる.
[BECOME [[後輩]z BE WITH
[[秘訣]y MOVE FROM [先輩]x TO [後輩]z]]]

「秘訣」が「先輩」から「後輩」に伝わるといった共通の出来事が, LCS においても共通の意味述語 “MOVE FROM TO” で表されている. 例文 (7s) はこの出来事を「先輩」の視点から, (7t) は「後輩」の視点から述べており, この視点の違いは, LCS のレベルでも上で示した 2 つの部分構造によって表されている.

2.1.3 述語項構造と表層格構造の対応関係

LCS は, 述語項構造によって働きかけや状態変化などの出来事に関わる要素の関係を表すだけでなく, 述語項構造と表層格構造を結び付けるための情報 (リンキングルール) も備えている.

T-LCS では, Agent の項を持つか否かに基づいて項と格を対応付ける²⁹⁾. たとえば, “CONTROL”, “ACT” などの意味述語は Agent を項に持つ. そのよ

うな LCS については、上位事象の第 1 項から順に、ガ格、ヲ格、ニ格に対応付ける (LCS2, LCS1)。一方、LCS が Agent を含まない場合は同様に、ガ格、ニ格に対応付ける (LCS10, LCS6, LCS7)。このリンキングルールは要素間の作用関係、動詞が必ずガ格をとることなどの性質を反映している。

2.2 種々の曖昧性と LCS

所与の動詞句を LCS で表現するには、語義、格構造および格要素の意味役割に関する曖昧性を解消する必要がある^{3),13),28),29)}。ある動詞が複数の語義を持つ場合、それらの統語的振舞いが異なるならば、それに基づいて付与される LCS も異なる。たとえば、文献 30) では次の 2 文の違いが解説されている。

(8) a. 杖を握った。

b. おにぎりを握った。

また、同じ動詞の同じ語義であっても、格構造が異なれば異なる LCS で表現される。たとえば、格構造が異なる例 (9) の 2 文には壁全体を塗ったか否かという微妙な解釈の違いがある。文献 19) では、これは格要素の意味役割の違いによって生じる違いとし、各文を異なる LCS で表している。

(9) a. 壁に (Goal) ペンキを (Theme) 塗った。

b. 壁を (Theme) ペンキで (Instrument) 塗った。

さらに、格要素の意味役割の曖昧性もある。日本語ではニ格に対応する意味役割の曖昧性がしばしば問題となる。たとえば、例文 (10a) の「正午に」は Goal の意味役割を持つ補語であるが、(10b) の同じ格は補語ではないため、LCS では表示されない。

(10) a. 出発時間を正午に (Goal) 変更した。

b. 先方に正午に (Time) メールを送った。

近年、FrameNet¹⁾ や Proposition Bank²⁶⁾ など、意味的なタグを付与したコーパス、語彙資源が利用可能になるにつれて、上で示したような曖昧性を解消する手法がさかんに研究されるようになってきた^{2),6)}。セマンティックパーズング、深層構造解析と呼ばれるこの分野の研究成果は目覚ましく、将来日本語に関しても高精度なアルゴリズム、ツールが利用可能になると期待できる。そこで本論文では、この技術を仮定して、言語生成の側面に焦点を当てる。

3. 機能動詞構文の言い換え

規則性の高い様々な種類の言い換えが語の統語・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。たとえば、動詞の自他交替や場所格交替、複合名詞の動詞句への言い換えを LCS に基づいて説明する試みがある^{14),19),28)}。しかしながら、LCS が表すどの特

性がどのように関わっているかは言い換えの種類ごとに異なるため、それぞれに有用な特性とその利用方法を同定する作業が欠かせない。

本論文では、例 (3)~(5) に示したような機能動詞構文の言い換えを取り上げる。機能動詞構文の言い換えは、動詞の格構造や態に影響するなど、言い換え生成における一般的な問題を多く含んでいるため、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といえる。さらに、先行研究で扱われてきた現象とは異なり、複数の動詞要素の統語・意味的特性の相互作用によって言い換え方が定まるといった性質を持つため、LCS の効用を調査するうえでも新規の知見を期待できる。

3.1 機能動詞の文法的機能

村木²²⁾ は「実質的な意味を名詞にあずけて、みずからはもっぱら文法的な機能をはたす動詞」を機能動詞と呼んでいる。一方本論文では、表層的な特徴に基づき、動詞から派生した名詞 (動作性名詞) を格要素に持ちうる動詞を機能動詞構文を構成しうる動詞と見なし、これを便宜的に機能動詞と呼ぶ。次の例文 (11s), (12s), (13s) においては、動作性名詞「感動」「攻撃」「逆転」を格に持つ「与える」「出る」「狙う」を機能動詞と呼ぶ。

(11) s. その映画が彼に感動を与えた。

t. その映画が彼を感動させた。

(12) s. 思い切った攻撃に出る。

t. 思い切って攻撃し始める。

(13) s. 速攻をしかけて逆転を狙う。

t. 速攻をしかけて逆転しようとする。

機能動詞構文において機能動詞はいくつかの文法的機能を果たす。例文 (11s)~(13s) における機能動詞はそれぞれ (11t)~(13t) のように言い換えられることから、ヴォイス、アスペクト、ムードに相当する機能を果たしているといえる。このうち、アスペクト、ムードに相当する機能は、機能動詞が果たす機能をそのまま機能動詞構文全体の意味と見なせる。したがって、機能動詞のそれぞれに対して、

「出る」=「し始める」

「狙う」=「しようとする」

のような文法的機能のラベルを付与すれば言い換えを生成できる可能性がある。一方、ヴォイスに相当する機能を持つ機能動詞構文を言い換える場合、機能動詞と動作性名詞の組合せを考慮しなければ、ヴォイス表現 (「する」「される」「させる」) を定めることができない^{22),25)}。たとえば、例 (3) では「与える」を能動態に言い換えるべきだが、例 (11) では同じ「与える」

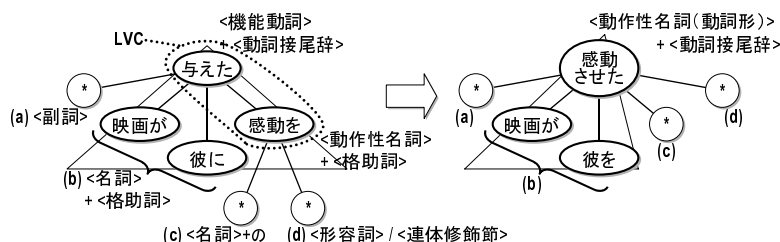


図 1 機能動詞構文の言い換えが影響する依存構造上の範囲

Fig. 1 The pair of dependency subtrees that are involved in paraphrasing of light-verb constructions.

を言い換える際に受動態にしなければならない。本論文では、このようなヴォイスに相当する機能を持つ機能動詞構文を対象を絞って LCS の効用を調査する。

3.2 扱う問題の範囲

機能動詞構文の言い換えの統語的な（依存構造上の）概形を図 1（例 (11)）に示す。図中の楕円は文節を表す。機能動詞構文の言い換えでは、言い換え生成においても一般的な次の 4 つの処理が必要である。

依存関係の変更：機能動詞がなくなるのに応じて図中の要素 (a), (b) の係り先を変更する。

連体修飾表現の変換：動作性名詞の品詞が動詞に変わるのに応じて動作性名詞の連体修飾要素 (c), (d) を連用修飾に変換する。

ヴォイス表現の選択：能動態「する」、受動態「される」、使役態「させる」の中から適切なヴォイス表現を選択する。

格助詞の再割当て：動作性名詞の動詞形が新しく節の主辞となるのに応じて要素 (b), (c) に格助詞を割り当てる。

1 つ目の処理は統語構造の操作によって容易に実現可能である。2 つ目の処理では、活用形の修正だけでなく、「ことを」、「ように」などの形式名詞の挿入や、不適切なコロケーションを避けるための類義語への変換が必要な場合もある²³⁾。ただし、これらは機能動詞構文の言い換えとは独立に処理できる形態素生成の問題と見なせる。本論文では、言い換えと独立に形態素情報や依存構造を参照するだけでは特定できない、3 つ目のヴォイス表現の選択および 4 つ目の格助詞の再割当ての処理に LCS の情報を利用する。

村木²²⁾ は、ヴォイス表現の選択に関わる個々の動詞の特性として、主体から見た行為の遠心性（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）の 2 つの方向性をあげている。この概念は、2.1 節で述べた LCS の項間の作用関係や視点の情報を用いて計算できる可能性がある。また、例 (3), (11) のような格助詞の変更も、リンキングルールを参

照すれば実現できると期待してよい。

なお、例 (14) の「消費者の」のように要素 (c) を格要素とすべき場合もある。

(14) s. 「無農薬」表示が消費者の誤解を招く。

t. 「無農薬」表示が消費者を誤解させる。

この要素については、まず動詞性名詞を動詞に言い換える際に格要素とすべきか、他の連用修飾表現とすべきかの判定が必要であり、要素 (c) を格要素とすべき場合は少なくない。ただし、事前に分析した範囲では、要素 (c) の有無で動作性名詞の兄弟格要素 (b) に割り当てられる格助詞が変化する事例はなかった。そこで、本論文では、ヴォイス表現の選択や要素 (b) の格助詞の再割当てと独立に要素 (c) を処理できると仮定し、機能動詞構文および要素 (b) のみ（図中の三角形の範囲）を処理対象とする。

4. LCS に基づく非決定性解消

機能動詞構文を言い換えた後の動詞句のヴォイス表現および補語の表層格を、LCS が示す動詞の統語・意味的特性に基づいて次の 3 ステップで計算する。図 2 も参照されたい。

- (ステップ 1) 意味解析：まず、入力文中の機能動詞構文を、統語構造解析結果に基づいて LCS で表現する。具体的には、LCS 辞書から機能動詞の LCS (LCS_{V0}) を取り出し、リンキングルールによって項を埋めて LCS_{V1} を得る。
- (ステップ 2) LCS の対応付け（意味構造変換）：次に、LCS 辞書から動作性名詞の動詞形に対応する LCS (LCS_{N0}) を取り出し LCS_{V1} と対応付ける。そして、項が埋まった LCS_{N1} および対応付けの際に余った部分構造 LCS_S を出力する。
- (ステップ 3) 表層生成：リンキングルールによって LCS_{N1} の各項に対応する格助詞を定め、最後に、 LCS_S に基づいてヴォイス表現を決定するとともに格交替を施す。

言い換え生成の各ステップは、2.1 節で述べた LCS

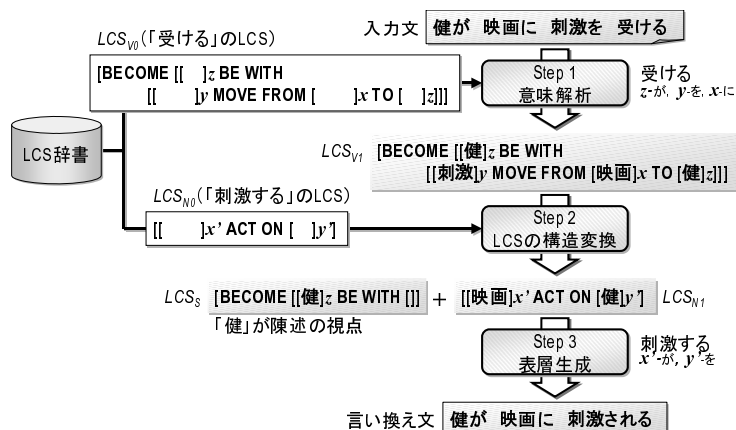


図 2 LCS の対応付けに基づく言い換え生成

Fig. 2 LCS-based paraphrase generation.

の諸性質に基づいて説明可能な少数の規則によって実現されており、実用規模の LCS 辞書があれば頑健に作用する。この節では以下、言い換え生成の各ステップについて図 2 の例を用いて説明する。なお、機能動詞 V の LCS の項を x, y, z 、動作性名詞 N の動詞形の項を x', y', z' と区別する。

4.1 意味解析

まず、LCS 辞書から機能動詞の LCS (LCS_{V0}) を取り出す。そして、格要素を LCS の項に対応付けて LCS_{V1} を得る。2.2 節で述べたような曖昧性の問題はありますが、言い換え生成の処理全体を自動化するため、2.1.3 項で示したリンキングルールを用いる。図 2 の例では、機能動詞「受ける」の LCS (LCS_{V0}) は Agent の項を持っている。ゆえに最も外側の項 z をガ格の名詞句「健」で埋める。また、左側から 2 番目、3 番目の項 y, x をそれぞれヲ格の名詞句「刺激」、二格の名詞句「映画」で埋める。

4.2 意味構造変換

次に、機能動詞構文の解析結果 LCS_{V1} の項を、動作性名詞の動詞形に対応する LCS のテンプレート LCS_{N0} の項と対応付けて、言い換え後の動詞の項構造 LCS_{N1} を得る。ここで、LCS が、上位事象と下位事象の包含関係および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、出来事の視点を明示的に表すことに着目する。そして、動詞の遠心性、求心性を近似的にとらえ、それらを変化させないように、出来事の視点以外の命題的な意味を担う部分から順に 2 つの LCS を対応付ける。すなわち、上位事象より下位事象を先に、状態変化の対象よりその状態変化の結果状態を先に対応付ける。この操作は LCS の右端の項から順に対応付け

る作業に相当する。図 2 の例では、最初に LCS_{V1} の z を LCS_{N0} の y' と対応付け、 z の要素「健」を y' に埋める。このような操作を順次施して言い換え後の動詞「刺激する」の項構造 LCS_{N1} を得る。 LCS_{V1} によって表される意味の範囲が LCS_{N1} よりも広い場合、左端の部分構造が対応付けられずに残る。この視点に関する情報 LCS_S も後続の表層生成のステップで用いるために出力する。

以下、(i) 意味述語の対応付け、(ii) 項の対応付け、(iii) 余った述語項構造の付加、の 3 段階の処理の各々について詳述する。図 3 も参照されたい。

4.2.1 (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 の働きかけ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 種に分類した。この分類に基づき、同じ分類の意味述語を対応付けるが、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”

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”

意味述語 “MOVE FROM TO” は例 (7) の「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用いられている¹⁵⁾。竹内らの LCS 辞書^{28), 29)} 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、LCS を機能動詞構文の言い換えに応用するにあたって “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。



図 3 LCS の対応付けの例

Fig. 3 An example of LCS transformation.

図 2 の例では, LCS_{V1} の “MOVE FROM TO” と LCS_{N0} の “ACT ON” を対応付けている.

4.2.2 (ii) 項の対応付け

意味述語を対応付けた後は, 意味述語の対応に基づいて LCS_{V1} と LCS_{N0} の項を作用の受け手あるいは状態変化の結果状態の項から順に対応付け, LCS_{N0} の値を埋める. ただし, 次に示す条件に該当しない場合に限る.

- 対象となる動作性名詞 (図 2 では「刺激」) の意味はすでに LCS_{N0} によって表されているので, LCS_{N0} のどの項とも対応付けない.
- 動詞が暗示的な格要素 (定項) を持つ場合 (たとえば動詞「署名する」のヲ格は必ず「名前」), LCS では “FILLED” (以下 “F”) という意味述語を項に埋めて表す. LCS_{V1} , LCS_{N0} 中の対応付け対象の項が “F” であるときは, 対応付けの対象を 1 つ上位 (左側) の項にシフトする.

図 2 の例では, 意味述語の対応に基づいて z と y' , x と x' が各々順に対応付けられ, 「健」と「映画」が y' , x' を埋めている.

4.2.3 (iii) 余った項構造の付加

上述の手続き (i), (ii) を終えた時点で LCS_{V1} の部分構造が LCS_{N1} 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_{N1} が示す動詞 1 語の意味には含まれないが, 元の表現の意味を保存するように言い換えるためには, なん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, 転じてはヴォイス表現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すべきではない. ゆえに提案モデルでは, この余った部分構造 LCS_S を LCS_{N1} とともに出力し, 後続の表層生成のステップで用いる. 図 2 の例では, LCS_{V1} の “[BECOME [[健]z BE WITH []]]” の部分が LCS_S となる.

4.3 表層生成

最後に, LCS_{N1} と LCS_S を言語表現生成の制約として用い, 統語構造上は未指定であったヴォイス表現と補語の表層格を決定する. 具体的にはまず, 4.1 節と同様にリンキングルールを用いて LCS_{N1} から言

語表現を生成する. 次に, LCS_S に基づいてヴォイス表現を決定し, 態交替を施す. 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.

図 2 の例では, まず, LCS_{N1} から能動態の文 (15) を生成する.

(15) 映画が健を刺激する

次に, 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し, 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. そして, 言い換え文 (16) を得る.

(16) 健が映画に刺激される

このような規則を 5 種類作成した (表 2). 上の規則の例は表中 2 番目の規則に該当する. この規則は長谷部⁸⁾ が述べた統語的な態交替の解釈に等しい.

5. 言い換えの生成実験

提案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について 5.1 節で述べ, 5.2 節で実験結果を示す.

5.1 語彙および評価用データ

5.1.1 動作性名詞

T-LCS 辞書 (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複数の異なる LCS が付与されている.

5.1.2 機能動詞構文パターン (動詞と格助詞の組)

村木²²⁾ 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あるいは平易語とは限らない. また, 同じ動詞でも, 動作性名詞の表層格が異なれば機能動詞構文とならない場合, 文法的機能が異なる場合がある. そこで, 「を与える」「を受ける」「に組み込む」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 1 つの機能動詞構文パターンとし, コーパス中の出現頻度を参照して新たに列挙した. パターン収集の詳細な

表 2 態交替の決定リスト
Table 2 Decision list for voice alternation.

適用順序	適用条件	ヴォイス表現	態交替の種類
1	LCS_S の最左の項と LCS_{N1} の最左の項が同じ要素を持つ	する	なし
2	LCS_S の最左の項と LCS_{N1} の Agent 以外の項が同じ要素を持つ	される	受動化
3	LCS_S が “BE WITH” を持ち、その第 1 項と LCS_{N1} の Agent の項が同じ要素を持つ	させる	使役化
4	LCS_S が Agent の項を持ち、 LCS_{N1} のいずれの項とも異なる要素を持つ	させる	使役化
5	なし	する	なし

手順は次のとおりである。

(ステップ 1) 新聞記事 19 年分のテキストの係り受け解析結果 から 876,101 組の $\langle v, c, n \rangle$ を収集した。 $\langle v, c, n \rangle$ は、上述の動作性名詞 n が格助詞 c を介して動詞 v に係っている関係を表す。

(ステップ 2) 頻度上位 50 組の $\langle v, c \rangle$ のそれぞれについてそれを含む $\langle v, c, n \rangle$ の頻度上位 10 組を取り出し、個々の $\langle v, c, n \rangle$ についてそれが機能動詞構文か否かを人手で判定した。機能動詞構文となる $\langle v, c, n \rangle$ が 1 組でも含まれる場合に $\langle v, c \rangle$ を機能動詞パターンとした結果、40 組の $\langle v, c \rangle$ が得られた。

(ステップ 3) T-LCS 辞書の作成に用いられたのと同じ言語テストおよび決定木に従って、各 $\langle v, c \rangle$ に LCS を付与した。機能動詞に多義性はなく、各々唯一の LCS が付与できた。

5.1.3 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を以下の手順で収集した。

(ステップ 1) 上で集められた 876,101 組の $\langle v, c, n \rangle$ から上述の機能動詞を含む $\langle v, c, n \rangle$ (23,608 組) を取り出し、頻度上位 245 組の $\langle v, c, n \rangle$ のそれぞれについて、それを含むような動詞格構造の頻度上位 3 例を取り出した。各々は単文と見なせる。

(ステップ 2) 2 名の日本語話者(大学院生)が言い換え文を人手で作成した。2 名の意見が一致するまで議論した結果、735 文中 711 文が機能動詞構文として言い換えられ、残りの 24 文は機能動詞構文ではなく、言い換えられないとされた。

コーパス中の機能動詞構文のトークンに対する上記 245 組の $\langle v, c, n \rangle$ の被覆率は、あらゆる $\langle v, c, n \rangle$ が機能動詞構文であると仮定したときを下限、上述の 40 組のみが機能動詞であると仮定したときを上限とすると、6.47% (492,737 / 7,621,089) から 24.1% (492,737 / 2,044,387) と推定される。

5.2 言い換え生成と評価

高頻度の語から優先的に機能動詞を収集したため、評価用データ中の機能動詞構文とそうでない表現の分布には偏りがある。そこで、機能動詞構文かそうでないかの分類ではなく、言い換え生成の性能評価に焦点を当てる。評価用データ 735 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、それぞれの格要素に格助詞が正しく付与されていれば格要素の語順が異なっていても適格と判定する。

以下、いくつかの言い換え生成モデルについて述べ、性能を比較する。

5.2.1 提案モデル (LCS)

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、提案モデルは機能動詞の LCS と動作性名詞の LCS のあらゆる組合せについて、4 章で述べた手続きを試行し、可能な言い換え候補をすべて出力する。

5.2.2 ベースライン (LM)

機能動詞構文を言い換えるタスクは、言い換え先の動詞(動作性名詞の動詞形)のヴォイス表現を「する」、「される」、「られる」の 3 つから選択し、その格となる名詞に格助詞「が」、「を」、「に」を重複がないように付与するという、言語生成の選択問題とみなせる。そこで、言い換え先の動詞 n_v および m 個の格要素 n_1, \dots, n_m に対して、ヴォイス表現 s および格助詞 $c = c_1, \dots, c_m$ の最適な組合せを確率的に推定する言語モデル (LM) を構築し、評価のベースラインとする。このモデルは次の式に従ってヴォイス表現および格助詞を決定する。

$$(s, c) = \operatorname{argmax}_{(s', c')} \prod_{i=1}^m P(\langle n_v, s' \rangle, c'_i, n_i)$$

$\langle n_v, s' \rangle$ は言い換え先の動詞 n_v をヴォイス表現 s' に合わせて活用変化させたもの(たとえば \langle 誘う, られる \rangle は「誘われる」)を表す。動詞と名詞の格関係に対する確率 $P(\cdot)$ としては、分布クラスタリングによ

表 3 ベースラインおよび提案モデルの性能

Table 3 Performance of the models.

モデル	LCS	LM	LCS+LM
自動生成された言い換え文の数	822	547	741
適格な言い換え文の数	624	322	609
不適格な言い換え文の数	198	225	132
再現率	0.878	0.453	0.857
精度	0.759	0.589	0.822
F 値 ($\alpha = 0.5$)	0.814	0.512	0.839

てスムージングされた共起確率⁵⁾を用いる。確率モデルは大規模なテキストコーパスを用いて構築するが、自動生成された表現がこのモデルの語彙に入らない場合もある。そのような場合、モデルは入力に対していかなる言い換えも出力しない。

5.2.3 統計的フィルタリング (LCS+LM)

提案モデルによって生成した言い換え文中の $\langle v, c, n \rangle$ を文献 5) で提案したモデルを用いて正例と負例に分類する。ここでは、正例に基づくモデルと負例に基づくモデルの混合モデルを用い、分類の閾値は文献 5) で用いたデータに基づいて経験的に 0.6977 とする。

5.2.4 評価結果

各モデルによって生成された言い換え文を次の 3 つの指標で評価した結果を表 3 に示す。

$$\text{再現率} = \frac{\text{適格な言い換え文の数}}{\text{入手で作成した言い換え文の数}}$$

$$\text{精度} = \frac{\text{適格な言い換え文の数}}{\text{自動生成された言い換え文の数}}$$

$$F \text{ 値} = \frac{1}{\alpha(1/\text{再現率}) + (1-\alpha)(1/\text{精度})}$$

提案モデルはベースラインよりも顕著に高い性能を示した。ゆえに、ヴォイス表現と格助詞の選択に関して、LCS が表す構成語の統語・意味的特性が有効であるといえる。また、統計的フィルタリングによって提案手法の誤りを軽減できることが確認できた。これについては、2 群の比率差の検定、有意水準 5% で、統計的フィルタを導入することの有意さを確認した。

6. 考 察

6.1 誤り分析

言い換え生成実験における、提案モデル (LCS) およびそれと統計的フィルタリングの組合せ (LCS+LM) の言い換え誤りの分布を表 4 に示す。

我々は、所与の $\langle v, c, n \rangle$ を正負例に分類するタスクを取り上げ、確率モデルと負例を用いた尤度関数を組み合わせることで確率モデル単体よりも高い分類性能が得られることを明らかにした⁵⁾。ただし、文献 5) が分類問題を扱ったのに対し、本論文で扱うのはヴォイス表現と格助詞の組合せの選択問題である。ゆえに、任意の組合せを上式のように確率の積で表現し、相対的に最適な組合せを選択する。

表 4 言い換え誤りの分布

Table 4 Error distribution.

モデル	LCS	LCS+LM
言い換え誤りの数	198 (100%)	132 (100%)
LCS の定義誤り	30 (15%)	19 (14%)
機能動詞の LCS	24	14
動作生名詞の LCS	6	5
言い換え生成モデルの誤り	61 (31%)	38 (29%)
LCS の構造変換誤り	59	36
動詞「する」の扱い	2	2
曖昧性による誤り	107 (54%)	75 (57%)
二格の曖昧性の扱い	78	47
機能動詞構文の誤認識	24	24
自動詞と他動詞の選択誤り	5	4

6.1.1 二格の曖昧性による誤り

最も多かったのは 2.2 節で述べたような二格名詞句の意味役割の曖昧性による誤りであった。提案モデルに例文 (17s) を入力すると、副詞的な役割を担う二格名詞が LCS_{V1} の項に埋められてしまう。それらが LCS の項の対応付けおよび態交替によって LCS_{N1} のラ格に対応する項に移され、不適格な言い換えが生成された。

(17) s. (金曜日を) めどに作業を進める。

LCS_{V1} [[ϕ]x CONTROL
[BECOME [[作業]_y BE AT [めど]_z]]]
 LCS_{N1} [[ϕ]x' ACT ON [めど]_y']

t. *(金曜日を) めどを作業する。

この入力文に対する正しい解析結果 ($18LCS_{V1}$) が得られれば、LCS の構造変換で正しい LCS_{N1} および言い換え文が生成できる。2.2 節で述べたとおり、LCS の実応用のためにセマンティックパーズングの技術の確立は急務である。

(18)

LCS_{V1} [[ϕ]x CONTROL
[BECOME [[作業]_y BE AT [めど]_z]]]
 LCS_{N1} [[ϕ]x' ACT ON [ϕ]y']

t. (金曜日を) めどに作業する。

6.1.2 機能動詞構文の誤認識

提案モデルは所与の入力を機能動詞構文と見なして必ず言い換えを生成するため、機能動詞構文ではない 24 例をすべて誤って言い換えてしまった。これらは統計的フィルタリングを適用しても棄却できなかったことから、機能動詞構文とそうでない表現は事前に区別しておく方が良いと考えられる。ただし、所与の $\langle v, c, n \rangle$ が機能動詞構文であるか否かは $\langle v, c, n \rangle$ 単独では必ずしも判断できない。たとえば、評価用データ

括弧内の表現は説明のために追加した恣意的な文脈である。

中の例文 (19s) では「意味がある」を「意味する」言い換えるべきではないとしたが、作例 (20s) では同じ表現を言い換えることができる。

- (19) s. (参加すること) 自体に意味がある。
t. *(参加すること) 自体を意味する。
(20) s. その言葉には尊敬の意味がある。
t. その言葉は尊敬を意味する。

これは、動作性名詞「意味」の多義性に起因している。したがって、機能動詞構文か否かの判定のためには、2.2 節で述べたように語義曖昧性解消が必要である。ただし、我々が収集した機能動詞 40 表現に限っては統語的振舞いや LCS の曖昧性はなかったため、まずは動作性名詞のみを対象と考えて差し支えない。

6.1.3 意味述語の対応付け誤り

誤りの半分以上は曖昧性によるものだったが、提案モデルにおける LCS の不適切な対応付けの問題も実験を通じて明らかになった。たとえば例文 (21s) に対しては、(21t) のような誤った言い換えが生成された。この際に生成された LCS_{V1} , LCS_{N1} は (21) に示すとおりである。

- (21) s. 監督が指示を出す。
 LCS_{V1} [[監督] x CONTROL
[BECOME [[指示] y BE AT [ϕ] z]]]
 LCS_{N1} [[ϕ] x' CONTROL
[[ϕ] y' MOVE FROM [監督] z' TO [F]]]
t. ≠ 監督に指示する。

正しくは例文 (22t) のように言い換えるべきであり、このためには (21 LCS_{V1}) から (22 LCS_{N1}) が構成されるべきであった。

- (22) t. 監督が指示する。
 LCS_{N1} [[監督] x' CONTROL
[[ϕ] y' MOVE FROM [ϕ] z' TO [F]]]

この誤りの原因は意味述語“CONTROL”と“MOVE FROM TO”を対応付けたことにある。図 3 の例では Agent が関与する意味述語との対応付けが可能であったので比較したところ、(21 LCS_{N1}) では Source の項 (“MOVE FROM TO”の第 2 項) と Agent が異なり、この項が Agent の働きかけを含意していないことが明らかになった。“MOVE FROM TO”そのものの定義についても問題があるので、6.2 節で議論する。

6.1.4 項の対応付け誤り

項の対応付けにおいても誤りを生じた。例文 (23s) に対して不適格な文 (23t) が生成されたが、正しくは (24t) のように言い換えるべきであった。

- (23) s. 時間に制限がある。
 LCS_{V1} [[制限] y BE AT [時間] z]

- LCS_{N1} [[ϕ] x' CONTROL
[BECOME [[ϕ] y' BE AT [時間] z']]]
t. *時間に制限する。

- (24) t. 時間を制限する。
 LCS_{N1} [[ϕ] x' CONTROL
[BECOME [[時間] y' BE AT [ϕ] z']]]

LCS_{N1} 中の項が“F”の場合に限って対応付けの対象をシフトした。しかし、例 (23) では上述の条件を満たしていなくても対象のシフトが必要であった。現状の項の対応付け規則はきわめて単純なので、今後、評価用データ中の言い換え前後の文に人手で適切な LCS_{V1} , LCS_{N1} を付与し、それらを比較して帰納的に規則を精緻化する必要がある。

6.2 LCS 体系へのフィードバック

我々は、少数の作例と LCS の情報を手がかりに提案モデル中の各種規則集合を設計した。これらの規則集合は、評価用データに対してある程度適切に作用したが、上述のような不適格な誤り事例を生成しないためには精緻化が欠かせない。ただし、個々の規則は T-LCS 辞書における LCS の体系に依存しているため、まずはこの体系を改善すべきと考えている。現時点では次に示す 3 つの改善点がある。

統語的振舞いと意味述語の整理：“MOVE FROM TO”は Agent が関与しない意味述語“MOVE TO”に移動や状態変化の源泉を表す項 (Source) を付加して作成された意味述語であるが、Source の必須性や動作主性、アスペクトに関する分析を経ず、厳密な定義はない。この意味述語に限らず、(i) LCS で表現する語の振舞い、(ii) 判別に用いる言語テスト、および (iii) テストの結果と意味述語の対応付けについて包括的に整理し、一貫性のある体系を構築する必要がある。

語義の網羅：T-LCS 辞書では複数の語義を網羅的には区別していないため、言い換え生成時に所与の語義に対して適切な LCS が得られない場合がある。語義ごとに語の統語的振舞いや LCS が異なる可能性があるため、語義を枚挙し、各語義に対して適切な例文を用意したうえで LCS を付与する必要がある。

格構造の多様性と項の意味役割の多様性の網羅：上の語義の網羅と同様に、ある動詞がとりうる格構造や述語項構造で扱うべき (必須格に対応する) 項の意味役割を網羅できていない。同じ語義でも例 (9) の場所格交替で見たように格構造、意味役割などが異なる可能性があるため、今後分析対象を広げて網羅性を高める必要がある。

LCS 辞書のような高価な資源を構築する際には、資源そのものの評価が欠かせない。すなわち、本論文で取り上げたように応用を通じて体系の不整合や収録すべき情報の不足などを明らかにする作業には意義がある。

7. 関連研究

7.1 言い換えの実現方法

LCS を言い換え生成に用いる試みは Dorr の研究³⁾にも見られる。ただし、彼女のモデルでは、動詞の構文ごとに LCS を記述し、それらを直接対応付けて同じ動詞の種々の交替現象を表現している。しかし、「売る」と「買う」のような異なる動詞間の交替は、より抽象的が高いレベルの意味記述言語によって実現されている⁴⁾。一方、日本語の交替現象に関する言語学的分析^{12),14),15)}では、例(7)のような語彙の派生のメカニズムや動詞の自他交替など、動詞が変わる現象にも言及している。本論文ではこれらを踏まえ、LCS 上の記号操作という一般化したレベルで言い換えを生成する枠組みを提案した。

言い換え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)^{11),21)}があげられる。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えの関係にある。

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。

変形規則と言い換え規則の語彙依存性、すなわち個々の語の交替、派生などの共起関係は、任意の表現レベルの構造を引数/返り値とする語彙関数 (lexical function) を用いて記述される。たとえば機能動詞構文の言い換えは、ある動詞 X に対してその名詞形を返す $S_0(X)$ ($S_0(\text{affect}) = \text{effect}$) と、ある動詞の名詞形 X 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 $Oper_1(X)$ ($Oper_1(\text{attempt}) = \text{make}$) という2つの語彙関数を組み合わせた(25)のような言い換え規則によって実現されている。

$$(25) X \leftrightarrow S_0(X) + Oper_1(S_0(X))$$

MTT では、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表現できるとされており、

それらの関数を組み合わせて表現する(25)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される。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インスタンスが必要となる。たとえば、機能動詞構文の言い換えに必要な語彙関数を実現するには、動作性名詞(引数)と機能動詞(返り値)の組合せの各々について項がどのように引き渡されるかを記述しなければならず、記述量はオープンクラスの語の組合せのオーダになってしまう。

一方 LCS では、同じように振る舞う動詞は同一の意味構造で表現される。そこで、我々は、動作性名詞と機能動詞の様々な組合せからなる機能動詞構文の言い換えを、LCS が提供する視点の情報やリンキングルールに基づく比較的少数の手続きによって表現した。このアプローチには、MTT に基づくアプローチと異なり、記述の一貫性を保ちつつ資源開発のコストを抑えられるという利点がある。

7.2 機能動詞構文の言い換え

鍛冶ら¹⁷⁾は、国語辞典の語釈文との表層的なパターンマッチに基づいて、「動作性名詞+格助詞+動詞」という形の表現の中から機能動詞構文(論文中は迂言表現)を認識し、言い換える手法を提案している。このアプローチには、(i) 機能動詞構文とそうでない表現の分類問題も扱う、(ii) 既存の資源を利用するため人的なコストを必要としない、(iii) ヴォイス、アスペクトだけでなく例(26)のように副詞表現などを用いた多様な言い換えが生成できる、という利点がある。

(26) s. 守りを固める。

t. しっかりと守る。

ただし、ヴォイスおよびアスペクトに相当する機能を持つ機能動詞のそれぞれに「受身」「開始」などのラベルを付与して用いているため、例(3)と(5)のような違いをとらえることはできない。奥²⁵⁾は、動作性名詞を自動詞、感情を表す他動詞、それ以外の他動詞の3種類に分類し、各群と機能動詞の組に対して、機能動詞構文を解析するための制約と変形操作を経験的に記述した。一方、本論文は言い換えにおいて個々の語がどのように振る舞うかを LCS が示す情報を用いてとらえることを主眼としており、その観点では良好な実験結果と有益な知見を得た。LCS に基づくアプローチには、国語辞典の語釈文と異なり、辞書の編纂の方針によって記述情報が揺れることが比較的少ないという利点がある。また、動詞の分類体系が言語テストに基づいて緻密に設計できれば、他の自然言語処理タスクへの応用可能性も見込まれる。

大竹²³⁾は、コーパス中に動詞の直近の格として現

れる動作性名詞の分布に基づいて機能動詞の候補を抽出し、動作性名詞の連体修飾表現の変換(図1の(c),(d))に焦点を当てた分析を行っている。本論文の提案モデルがLCSを用いてとらえられる情報は、2.1節で述べたような動詞の統語・意味的特性のみであるため、周辺文脈との整合性を保つためにはこのような分析と後処理の実現も欠かせない。

8. おわりに

本論文では、言い換え生成における個々の語の統語・意味的な特性に依存した問題の解決法として、語彙意味論の体系の一つである語彙概念構造(LCS)に基づくアプローチを提案した。我々は、既存のLCS辞書の体系および既存の理論的分析に基づいて、LCSが表す意味構造の変形操作および語彙意味論的制約に基づく言い換え生成モデルを構築した。そして、機能動詞構文の言い換えを取り上げた言い換の生成実験を通じて、LCSが持つ情報の言い換え生成における効用を確認した。また、既存のLCS辞書の問題を明らかにし、改善点を整理した。

LCSの体系化・辞書構築に関していくつかのプロジェクトが進行中である。これにともない我々は、言い換の生成実験で得られた知見をLCSの体系化・辞書構築のプロセスにフィードバックし、語彙や格要素の意味役割などの曖昧性解消を含めた実用的なモデルに発展させるための研究に取り組んでいる^{9),30)}。今後は、LCS辞書の整備にともない、機能動詞構文の言い換え規則の精緻化、他の様々な種類の言い換えへのLCSの応用可能性の調査を進めたい。

謝辞 岡山大学の竹内孔一氏に、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。本論文の査読者の方々には、論文を精読していただき、有益なご指摘をいただきました。皆様に深く感謝いたします。

参 考 文 献

- 1) Baker, C.F., Fillmore, C.J. and Lowe, J.B.: The Berkeley FrameNet project, *Proc. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL)*, pp.86-90 (1998).
- 2) Carreras, X. and Màrques, L.: Introduction to the CoNLL-2004 shared task: Semantic role labeling, *Proc. of 8th Conference on Natural Language Learning (CoNLL)*, pp.89-97 (2004).
- 3) Dorr, B.J.: Large-scale dictionary construction for foreign language tutoring and interlingual machine translation, *Machine Translation*, Vol.12, No.4, pp.271-322 (1997).
- 4) Dorr, B.J., Green, R., Levin, L., Rambow, O., Farwell, D., Habash, N., Helmreich, S., Hovy, E., Miller, K.J., Mitamura, T., Reeder, F. and Siddharthan, A.: Semantic annotation and lexico-syntactic paraphrase, *Proc. 4th International Conference on Language Resources and Evaluation (LREC) Workshop on Building Lexical Resources from Semantically Annotated Corpora*, pp.47-52 (2004).
- 5) 藤田 篤, 乾健太郎, 松本裕治: 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*, Vol.45, No.4, pp.1176-1187 (2004).
- 6) Gildea, D. and Jurafsky, D.: Automatic labeling of semantic roles, *Computational Linguistics*, Vol.28, No.3, pp.245-288 (2002).
- 7) Habash, N., Dorr, B.J. and Traum, D.: Hybrid natural language generation from Lexical Conceptual Structures, *Machine Translation*, Vol.18, No.2, pp.81-127 (2003).
- 8) 長谷部郁子: 日英語の非対格構造の統語的使役化, *レキシコンフォーラム No.1*, 影山太郎(編), pp.103-131, ひつじ書房 (2005).
- 9) 平野 徹, 飯田 龍, 藤田 篤, 乾健太郎, 松本裕治: 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*, pp.396-399 (2005).
- 10) 乾健太郎, 藤田 篤: 言い換え技術に関する研究動向, *自然言語処理*, Vol.11, No.5, pp.151-198 (2004).
- 11) Iordanskaja, L., Kittredge, R. and Polguère, A.: Lexical selection and paraphrase in a meaning-text generation model, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, Paris, C.L., Swartout, W.R. and Mann, W.C. (Eds.), pp.293-312, Kluwer Academic Publishers (1991).
- 12) 伊藤たかね(編): 文法理論: レキシコンと統語, 東京大学出版会 (2002).
- 13) Jackendoff, R.: *Semantic structures*, The MIT Press (1990).
- 14) 影山太郎: 動詞意味論—言語と認知の接点, くろしお出版 (1996).
- 15) 影山太郎(編): 日英対照 動詞の意味と構文, 大修館書店 (2001).
- 16) 影山太郎: 非対格構造の他動詞—意味と統語のインターフェイス, 文法理論: レキシコンと統語, 伊藤たかね(編) pp.119-145, 東京大学出版会 (2002).
- 17) 鍛冶伸裕, 黒橋禎夫: 迂言表現と重複表現の認識

- と言い換え, 自然言語処理, Vol.11, No.1, pp.81-106 (2004).
- 18) 加藤恒昭, 畠山真一, 坂本 浩, 伊藤たかね: 実証的な語彙概念構造辞書の構築に向けて—日本語和語動詞に関する語彙概念構造辞書構築の試み, 東京大学 21 世紀 COE プログラムシンポジウム: 語彙概念構造辞書の構築と応用, pp.21-26 (2005).
 - 19) 岸本秀樹: 壁塗り構文, 日英対照動詞の意味と構文, 影山太郎 (編), pp.100-126, 大修館書店 (2001).
 - 20) Levin, B.: *English verb classes and alternations: A preliminary investigation*, Chicago Press (1993).
 - 21) Mel'čuk, I. and Polguère, A.: A formal lexicon in meaning-text theory (or how to do lexica with words), *Computational Linguistics*, Vol.13, No.3-4, pp.261-275 (1987).
 - 22) 村木新次郎: 日本語動詞の諸相, ひつじ書房 (1991).
 - 23) 大竹清敬: 機能動詞結合の換言に伴う連体修飾表現の変換, 言語処理学会第 11 回年次大会発表論文集, pp.337-340 (2005).
 - 24) 大石 亨, 松本裕治: 格パターン分析に基づく動詞の語彙知識獲得, 情報処理学会論文誌, Vol.36, No.11, pp.167-180 (1995).
 - 25) 奥 雅博: 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, Vol.31, No.12, pp.1727-1734 (1990).
 - 26) Palmer, M., Gildea, D. and Kingsbury, P.: The Proposition Bank: an annotated corpus of semantic roles, *Computational Linguistics*, Vol.31, No.1, pp.71-106 (2005).
 - 27) Pustejovsky, J.: *The generative lexicon*, The MIT Press (1995).
 - 28) 竹内孔一, 内山清子, 吉岡真治, 影浦 峯, 小山照夫: 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, Vol.43, No.5, pp.1446-1456 (2002).
 - 29) 竹内孔一: 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, pp.576-579 (2004).
 - 30) 竹内孔一, 乾健太郎, 藤田 篤, 竹内奈央, 阿部修也: 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, pp.123-130 (2005).

(平成 17 年 9 月 15 日受付)

(平成 18 年 3 月 2 日採録)



藤田 篤 (正会員)

2005 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。京都大学情報科学研究科産学官連携研究員を経て, 2006 年より名古屋大学大学院工学研究科助手。現在に至る。博士 (工学)。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。2005 年度本学会山下記念研究賞受賞。言語処理学会, ACL 各会員。



降幡建太郎

2005 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。同年 (株) 東芝入社。現在に至る。自然言語処理の研究に従事。



乾 健太郎 (正会員)

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 ~ 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。2004 年文部科学省長期在外研究員として英国サセックス大学に滞在。現在に至る。博士 (工学)。自然言語処理の研究に従事。人工知能学会, 言語処理学会, ACL 各会員。



松本 裕治 (正会員)

1979 年京都大学大学院工学研究科修士課程情報工学専攻修了。同年電子技術総合研究所入所。1984 ~ 1985 年英国インペリアルカレッジ客員研究員。1985 ~ 1987 年 (財) 新世代コンピュータ技術開発機構に。京都大学助教授を経て, 1993 年より奈良先端科学技術大学院大学情報科学研究科教授。現在に至る。工学博士。専門は自然言語処理。人工知能学会, 日本ソフトウェア科学会, 言語処理学会, 認知科学会, 計量国語学会, AAAI, ACL, ACM 各会員。