

「サ変名詞+する」から動詞相当句への言い換え

近藤 恵子[†] 佐藤 理史^{†,††} 奥村 学[†]

本稿では、「サ変名詞+する+接尾辞」からより平易な動詞相当句への言い換えを機械的に実現する方法を提案する。この方法は、5つの言い換え規則と4つの辞書とアルゴリズムから成っている。もし与えられた入力に特別な接尾辞がなければ、最も基本的な置換の規則が適用され、そのために、サ変名詞-動詞相当句対応辞書と動詞活用辞書が使用される。もし、テイル形、使役表現、受動表現、可能表現を表す接尾辞があれば、アルゴリズムにより最も適切な規則が適用される。その規則は、自動詞-他動詞対応辞書と有情/非情名詞辞書から得られる構文と意味の情報によって選択される。我々はこの方法に基づいた言い換えシステムを作成し、その有効性を確認した。

Paraphrasing of “sahen-noun + SURU”

KEIKO KONDO,[†] SATOSHI SATO^{†,††} and MANABU OKUMURA[†]

This paper proposes a method of automatically paraphrasing a Japanese expression “sahen-noun + SURU + suffixes” into a simpler expression. The “sahen-noun” is a verbal noun in Japanese and it works as a verb in the form of “sahen-noun + SURU”. The proposed method consists of five rules, four dictionaries, and an algorithm. If there is no special suffix in the given input, the simplest replacement rule is applied. Two dictionaries are used for the replacement and adjustment: the dictionary that records sahen-nouns and their corresponding verbs, and the dictionary for verb-inflection. If there is one of four special suffixes, “-SERU” (compulsion), “-RERU” (passiveness), “-DEKIRU” (possibility), and “-IRU” (continuous), the most appropriate rule is applied. The rule is selected by syntactic and semantic information that is provided by other two dictionaries: the dictionary that records intransitive verbs and their corresponding transitive verbs, and the dictionary that records animateness/inanimateness of nouns. We have constructed the paraphrase system based on the method, and showed the effectiveness of the method.

1. はじめに

言い換えとは、ある表現に対して、「内容を変えずに表現を変える」ことである。

我々は日常的に言い換えを行っている。相手が専門用語を知らなければ、一般用語に言い換える。表現が難解であるために理解されなければ、平易な表現に言い換える。内容が伝わらなければ、別な言い方に言い換える。こうした言い換えは、円滑なコミュニケーションのために重要である。

こうした言い換えを機械的に実現することは、どのくらい困難な問題なのであろうか。また、そのためには、どのような知識が必要になるのであろうか。ある

言語を別の言語に言い換える「翻訳」がさかんに研究されてきたのに対し、同一言語内の「言い換え」は、あまり研究されてこなかった。過去の言い換えに関する研究には、テキストを言い換えるモデル¹⁾や、構文木を用いた言い換え法²⁾、変換テーブルを用いた部分的な置換⁹⁾などがある。しかし、上記の疑問に対する答えはほとんど分かっていない。

言い換えといっても、そのカバーする範囲はかなり広く、いろいろな種類の言い換えが存在する。大きくは、構文的言い換え、意味的言い換え、プラグマティック的言い換への、3つのクラスを考えることができる。この中でも、最も機械化が容易だと考えられるのは、構文的言い換えである。構文的言い換えは、構造のマッピングに基づくもので、語の意味を文脈やその他の知識から判断し言い換えるようなものは、このクラスには含まれない。単語の同義語・類義語への置換や構造の変形は、この構文的言い換えに含まれる⁸⁾。

本稿では、このような構文的言い換への1つとして、

[†] 北陸先端科学技術大学院大学

Japan Advanced Institute of Science and Technology, HOKURIKU

^{††} 科学技術振興事業団さきがけ研究 21「情報と知」領域グループ
“Information and Human Activity”, PRESTO, JST

「サ変名詞+する」をより簡単な動詞（あるいは動詞相当句）に言い換える問題を取り上げる。「サ変名詞+する」は、活用変化や接尾辞をともなって、述語を構成する。多くの場合、「サ変名詞+する」を動詞に置き換えても、活用形や接尾辞などを変更する必要はない。しかし、テイル形（動きの継続や状態を表す）、使役表現、受動表現、可能表現をとる場合には、活用形や接尾辞なども変更する必要が生じる。このため、単に「サ変名詞+する」とそれに対応する動詞（もしくは動詞相当句）の辞書を用意し、それを用いて「サ変名詞+する」の部分だけを置き換えるのでは、正しい言い換えを実現できない。

以下、2章では「サ変名詞+する」の言い換えの具体例を検討し、それに基づき機械的に言い換えを実現する方法を提案する。3章ではその方法に基づく言い換えシステムについて述べる。4章ではその言い換えシステムに必要な各種知識を既存辞書から獲得する方法について述べる。5章では実験とその結果の検討について述べ、6章で議論を行い、7章でまとめを述べる。

なお、本稿では、用いる知識を、おおよそ文法に関する知識に限定し、その範囲で解くべき問題を扱う。文脈などを考慮に入れて、言い換え先の語を決定する問題などは扱わないこととする*。

2. 言い換えの分析

本章では、サ変名詞を動詞的に使用している「サ変名詞+する+接尾辞*」と「サ変名詞+できる+接尾辞*」をより簡単な表現に言い換える具体例を検討し、言い換えのための規則と、それを適用するアルゴリズムを提案する。なお、本研究の品詞、および動詞と接尾辞の活用型と活用形は、文献3)を参考にし、JUMAN¹⁰⁾に準拠している。また、“*”は正規表現に従い0個以上の繰返しを意味するものとして使用する。

2.1 基本的な言い換え

まず「サ変名詞+する+接尾辞*」を動詞に言い換える、ごく単純な例文を考える。

作例1 入学式に桜が開花する → 入学式に桜が咲く

作例2 入学式には桜はまだ開花しません → 入学式には桜はまだ咲きません

このような単純な言い換えは、図1に示した規則1で実現される。すなわち、

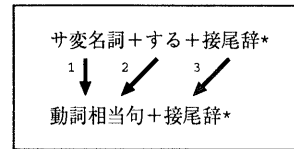


図1 規則1：基本的な言い換え規則
Fig. 1 Rule 1: Default rule.

- (1) サ変名詞を動詞相当句に置き換える。
- (2) 動詞相当句の活用形を「する」の活用形に一致させる。
- (3) 接尾辞をコピーする。

「サ変名詞+する」は、動詞1語に言い換えられることが多く、作例1でも「開花する」は「咲く」という動詞1語に言い換えられている。しかし、それ以外にも、「形容詞+する」や「形容詞+なる」、「手に入れる」などの熟語、「だんだんに増える」などの動詞に係る句を含んだものに言い換えられる場合もある。本稿ではこれらを動詞に相当する句という意味で、動詞1語も含め、動詞相当句と呼ぶ。

動詞相当句の活用形は、「する」の活用形に基づいて決定する。通常は、動詞相当句の活用形を、「する」の活用形と一致させればよい。たとえば、作例2の「開花しません」において、「する」は「し（基本連用形**）」として現れているため、動詞相当句「咲く」の活用形をこれに一致させ、「咲き（基本連用形）」とすればよい。しかし、「する」の活用形と同一の活用形が、必ずしも動詞相当句の活用形として存在するとは限らない。たとえば、「する」には文語命令形「せよ」が存在するが、「咲く」の活用型である子音動詞カ行には、文語命令形が存在しない。このような場合は、言い換え先の動詞相当句の活用形を何形にすればよいか、決定する必要がある。この決定のための規則を、表1に示す。たとえば、「開花せよ（文語命令形）」を言い換える場合、「咲く」の活用形を表1に従い命令形とすることにより、「咲け（命令形）」と言い換える。なお、動詞相当句の活用型、活用形とは、動詞相当句の末尾に含まれている動詞、もしくは接続している「なる」や「する」の活用型と活用形を指すこととする。

「する」の後に接尾辞が接続している場合、接尾辞はそのまま動詞相当句にコピーすればよい。しかし、「サ変名詞+し+ない+接尾辞*」のように「ない」が付く場合は、動詞相当句の活用形は例外的に次のように変更される。

* 1つの「サ変名詞+する」には、一般に複数の言い換え先の動詞（あるいは動詞相当句）が考えられる。これらの中から適切なものを選ぶ問題は、文脈などを考慮に入れて、語義の決定やコロケーションの問題を扱う必要があるという点で、機械翻訳の動詞の訳語選択の問題とはほとんど同じ問題となる。

** JUMANは、「する」の連用形を「し（基本連用形）」の他に、「して（タ系連用テ形）」「したり（タ系連用タリ形）」「しちゃ（タ系連用チャ形）」に分類している。

表1 動詞相当句活用形決定規則
Table 1 Mapping Rules to the other inflection.

| 「する」の活用形 | 動詞相当句の活用形 (同一活用形が存在しない場合) |
|----------|------------------------------|
| 文語基本形 | 基本形 |
| 文語未然形 | 未然形 |
| 文語命令形 | 命令形 |
| 音便条件形 | 基本条件形 |

作例3 入学式に桜が開花しない → 入学式に桜が咲かない

この用法の「ない」は、従来の国文法では助動詞に分類され、動詞の未然形に接続するとされている。一方、JUMANにおける「ない」は、動詞の活用型がサ変動詞ならば基本連用形、それ以外の動詞の場合には未然形に接続する接尾辞として扱われている。「する」はサ変動詞であるため、JUMANは「し+ない」の「し」を基本連用形として扱う。そのため、「サ変名詞+する」に接尾辞「ない」が接続している場合には、動詞相当句の活用形は、活用型がサ変動詞ならば同じ基本連用形にし、それ以外の活用型の場合は未然形に変更する必要がある。

2.2 特別な言い換え

接尾辞がアスペクトやヴォイスを表現している場合、規則1では自然な言い換えにならない場合がある。

2.2.1 テイル形

動きの継続や結果の状態などを意味する表現に、「テ形+接尾辞「いる」」がある。この表現は「テイル形」とも呼ばれる。「テ形+いる」は、動詞相当句にコピーできる場合と、できない場合がある。

作例4 入学式には桜が開花しているだろう → 入学式には桜が咲いているだろう

作例5 文明が存在していた → ×文明があっていた

作例4の言い換えはテイル形をコピーできるが、作例5の言い換えはテイル形をコピーできない。これは、作例5の動詞「ある」が、存在・所有を意味する状態動詞であり、テイル形が存在しないためである。このような動詞として他に「いる」「おる」がある。

以上より、テイル形の言い換えアルゴリズムは、次のように整理される。

- 動詞相当句が「ある」「いる」「おる」→ 言い換えない(テイル形と分かった時点で言い換えを中止する)
- 上記以外 → 規則1

2.2.2 使役表現

「サ変名詞+する」の使役は「サ変名詞+さ+接尾辞「せる」」で表現される。

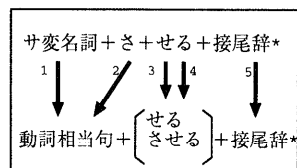


図2 「せる」を接続する使役表現の言い換え

Fig. 2 Mapping of "SERU".

作例6 形勢が不利になったので、味方を退却させる → 形勢が不利になったので、味方を退かせる

作例7 嘘をついて信用させた → 嘘をついて信じさせた

作例6, 7の言い換えは、図2に示した図式により実現される。すなわち、

- (1) サ変名詞を動詞相当句に置き換える。
- (2) 動詞相当句の活用形を「する」の活用形に一致させる。
- (3) 動詞相当句の活用型がサ変動詞か子音動詞ならば「せる」をそのまま接続させ、それ以外の場合は「させる」に変更して接続させる。
- (4) 動詞相当句の「せる」(もしくは「させる」)の活用形を、元の「せる」の活用形に一致させる。
- (5) 接尾辞をコピーする。

これは、「せる」が「させる」に変更されることがあるという点を除くと、規則1に従った言い換えと見なすことができる。

しかし、次のように自動詞の使役表現を他動詞に言い換えるものも見られる。

作例8 AとBの問題を直結させないように → AとBの問題を結び付けないように

作例8の言い換えは、図3に示した規則2で表現される。すなわち、

- (1) 「サ変名詞+さ+せる」を他動詞の動詞相当句に置き換える。
- (2) 動詞相当句の活用形を元の「せる」の活用形に一致させる。
- (3) 接尾辞をコピーする。

ここでは、「直結させる」という自動詞の使役表現が、「結び付ける」という他動詞に言い換えられている。しかし、自動詞の使役表現がつねに他動詞になるわけではない。たとえば、作例6も自動詞の使役表現であるが、他動詞への言い換えは行われていない。

規則1で言い換えるか、規則2で言い換えるかは、サ変名詞によって決定できることが多い。たとえば、作例6の「退却させる」は「退ける」にはならず、つねに規則1で言い換えた「退かせる」になる。また作

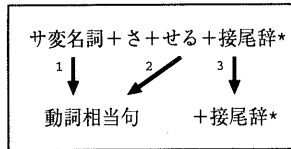


図3 規則2: 他動詞化する使役表現の言い換え規則
Fig.3 Rule 2: Mapping to the transitive verb.

例8の「直結させないように」は「結び付けさせないように」ではなく、つねに規則2で言い換えた「結び付けないように」になる。

しかし、サ変名詞だけでは決定できない場合もある。その例を次に示す。

作例9 私は車を前進させる → 私は車を進める

作例10 指揮者が彼を前進させる → 指揮者が彼を進ませる

作例9は規則2で言い換えられており、作例10は規則1で言い換えられている。

この「進める」と「進ませる」への言い換への違いは、一般に、対象への作用が直接作用か間接作用かによると考えられている¹³⁾。直接作用は、ある事象に対して第三者が直接的に力を加え、操作することである。これに対し間接作用は、第三者は間接的にしかかわらず、第三者に指示されたにしろ、事象自体は対象が行うものである。

直接作用か間接作用かは、厳密には文脈から読みとらねばならない。しかし、多くの場合は、「車」と「彼」が自発的に動くかどうかで近似的に決定できる。たとえば、作例9では「私」が直接的に「車」を操作して「前進させた」のだから、直接作用である。一方、作例10では、「指揮者」は「彼」に指示を出すなどして、彼に自分から「前進させた」のであるから、間接作用である。ここでの「車」と「彼」の違いは、自発的に動くかどうかという点にある。「車」は意志を持たず自分から動かないものである。これを「非情」という。一方「彼」は意志を持ち自力で動けるものである。これを「有情」という³⁾。

このように、直接作用か間接作用かは、対象が有情であるか非情であるかに強く影響される。そのため、規則1と規則2のどちらで言い換えるかは、表層ヲ格の名詞が有情か非情かで近似的に判定できる。

以上をまとめると、使役表現の言い換えアルゴリズムは、次のように整理される。

- 必ず動詞相当句に「せる」(もしくは「させる」)を接続するサ変名詞 → 規則1
- 必ず他動詞に言い換えるサ変名詞 → 規則2
- それ以外のサ変名詞

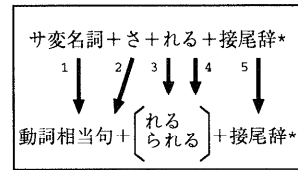


図4 「れる」を接続する受動表現の言い換え
Fig.4 Mapping of "RERU".

－ 表層ヲ格の名詞が有情 → 規則1

－ 表層ヲ格の名詞が非情 → 規則2

2.2.3 受動表現

「サ変名詞+する」の受動は「サ変名詞+さ+接尾辞“れる”」で表現される。

受動表現は直接受動表現と間接受動表現に分けることができる。まず直接受動表現の場合を考えよう。

作例11 彼はコーチに選手として育成された → 彼はコーチに選手として育てられた

作例12 犯人は警察に逮捕された → 犯人は警察に捕まえられた

作例11, 12の言い換えは図4の図式で実現される。すなわち、

- (1) サ変名詞を動詞相当句に置き換える。
- (2) 動詞相当句の活用形を「する」の活用形に一致させる。
- (3) 動詞相当句の活用型がサ変動詞か子音動詞ならば「れる」をそのまま接続させ、それ以外の場合は「られる」に変更して接続させる。
- (4) 「れる」(もしくは「られる」)の活用形を元の「れる」に一致させる。
- (5) 接尾辞をコピーする。

これは、「れる」が「られる」に変更されることがあるという点を除くと、規則1に従った言い換えと見なすことができる。

しかし、次のように他動詞の受動表現を自動詞に言い換えるものも見られる。

作例13 彼女の新刊が出版される → 彼女の新しい本が出る

作例13の言い換えは、図5の規則3で実現される。すなわち、

- (1) 「サ変名詞+さ+れる」を自動詞の動詞相当句に置き換える。
- (2) 動詞相当句の活用形を「れる」の活用形に一致させる。
- (3) 接尾辞をコピーする。

ここでは、「出版する」という他動詞の受動表現が、「出る」という自動詞に言い換えられている。

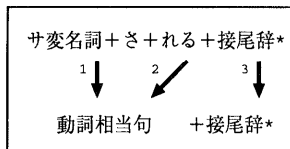


図5 規則3:自動詞化する受動表現の言い換え規則

Fig.5 Rule 3: Mapping to the intransitive verb.

ある他動詞に対応する自動詞と、その他動詞の受動表現とは、元の他動詞との関係において並行性があるという指摘がある。さらに文献13)において野村は、そのような場合、主体が非情である場合には自動詞による表現が好まれ、主体が有情である場合には受動の表現が好まれるという相補的な傾向があると述べている。

作例11の主体は「彼」であり、有情である。「育成された」は「育てられた」には言い換えられるが、自動詞「育った」には言い換えられない。しかし、作例12において、「逮捕された」は、文脈によっては自動詞「捕まった」になる可能性も考えられる。作例13の主体は「本」であり、非情である。「出る」は自動詞であり、他動詞の受動表現ならば「出された」となるが、ここでは自動詞「出る」に言い換える方が、より自然である。

このように、主体が有情であれば受動表現になり、非情であれば自動詞による表現になるという傾向が、例においても見られる。しかし、もし主体が非情であっても、動詞相当句の他動詞に対応する自動詞が存在しなければ、「れる」(もしくは「られる」)によって言い換えられる。また、自動詞に言い換えるとかえって不自然になることもある。以下にその例を示す。

作例14 大坂城は豊臣秀吉によって建設された → 大坂城は豊臣秀吉によって建てられた(×建った)

このように、主体が有情であるか非情であるかだけで、厳密に判定することはできないが、傾向としてこれを参考にすることはできる。

次に、間接受動表現について見る。

作例15 私は家を地震により破壊された → 私は家を地震により壊された

作例16 私は不況のために会社に労働時間を短縮された → 私は不況のために会社に労働時間を短くされた

作例15, 16はどちらも規則1で言い換えられている。間接受動表現は、有情の主体が間接的に影響、多くは被害をこうむることから、「迷惑の受動」ともいわれる³⁾。そのため、被害のニュアンスを残す規則1で

言い換えられる。

以上をまとめると、受動表現の言い換えアルゴリズムは、次のように整理される。

- 直接受動表現
 - 主体が有情 → 規則1
 - 主体が非情
 - * 対応する自動詞がある → 規則3
 - * 対応する自動詞がない → 規則1

- 間接受動表現 → 規則1

このアルゴリズムでは、まず、直接受動表現か間接受動表現かを判定する必要がある。判定には、表層的な格情報だけでなく意味的な解析が要求され、処理が複雑になる。しかし、間接受動表現は主体が有情であることに基づき、主体が有情であるか非情であるかに注目すると、受動表現の言い換え規則適用のアルゴリズムは、次のようにまとめ直すことができる。

- 主体が有情 → 規則1
- 主体が非情
 - 対応する自動詞がある → 規則3
 - 対応する自動詞がない → 規則1

このアルゴリズムを用いれば、直接受動と間接受動を判別する必要がなくなる。

2.2.4 可能表現

「サ変名詞+する」の可能は「サ変名詞+できる」で表現される。

作例17 彼女は英語で会話できる → 彼女は英語で話せる

作例18 彼を信用できますか → 彼を信じられますか

作例19 発表時間を短縮できない → 発表時間を短くできない

作例17, 18の言い換えは図6に示した規則4で実現される。すなわち、

- (1) サ変名詞を動詞相当句に置き換える。
- (2) 動詞相当句の活用形を決定し、動詞相当句の活用型が子音動詞ならば「-ERU」を接続させ、それ以外ならば「られる」を接続させる。
- (3) 「-ERU」(もしくは「られる」)の活用形を「できる」の活用形に一致させる。
- (4) 接尾辞をコピーする。

動詞相当句の活用形は活用型によって決定され、「できる」も動詞相当句の活用型により「-ERU」か「られる」に変更される。これらの決定と変更の規則は、表2のように整理される。

作例19の言い換えは図7に示した規則5で実現される。すなわち、

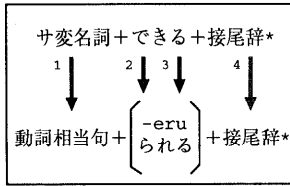


図6 規則4:「-ERU/られる」を用いた可能表現の言い換え規則
Fig. 6 Rule 4: The verb in the potential mood.

表2 「-ERU/られる」への変更規則
Table 2 Mapping to “-ERU/RARERU”.

| 動詞相当句 | | 「できる」が 変更される形 |
|-------|------------|------------------|
| 品詞 | 活用型 活用形 | |
| 動詞 | 子音動詞 語幹 | -ERU |
| 動詞 | 母音動詞 未然形 | られる |
| 動詞 | カ変動詞 未然形 | られる |
| 動詞 | ザ変動詞 文語未然形 | られる |

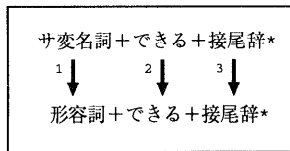


図7 規則5:「できる」のコピーによる可能表現の言い換え規則
Fig. 7 Rule 5: The adjective in the potential mood.

- (1) サ変名詞を形容詞に置き換える。
- (2) 形容詞の活用形を決定し、「できる」をコピーする。
- (3) 接尾辞をコピーする。

形容詞の活用形は、次に「できる」が接続することから、形容詞の活用形がイ形容詞かタル形容詞であれば基本連用形に、それ以外の活用形ではダ列基本連用形となる。

規則4で言い換えるか、規則5で言い換えるかは、動詞相当句によって決定される。動詞は「-ERU」（もしくは「られる」）が接続することにより、可能を表現する。これは規則4により言い換えられる。しかし、「-ERU」と「られる」は形容詞には接続しない。形容詞の可能は「サ変名詞+する」と同様、「できる」の接続により表現される。そのため、形容詞の場合は規則5を用いる。

以上をまとめると、可能表現の言い換えアルゴリズムは、次のように整理される。

- 言い換え先が形容詞を含む動詞相当句 → 規則5
- 上記以外 → 規則4

3. 言い換えシステム

前章の規則とアルゴリズムの有効性を確認するため、

言い換えシステムを作成した。

3.1 システム構成

システム構成を図8に示す。本システムは、言い換え部と、言い換え部が参照する「有情/非情名詞辞書」「サ変名詞-動詞相当句対応辞書」「自動詞-他動詞対応辞書」「活用辞書」の4つの辞書から構成される。

本システムは入力として「サ変名詞+する+接尾辞*」、「表層ヲ格の名詞」、「ガ格の名詞」をとり、2章で述べた方法に基づき、入力を言い換えたものを出力する。

3.2 サ変名詞-動詞相当句対応辞書

サ変名詞-動詞相当句対応辞書の一部を表3に示す。この辞書の1レコードは1つのサ変名詞の言い換えに関する情報に対応しており、(1)サ変名詞、(2)「サ変名詞+する」のときに自動詞か他動詞か、(3)対応する動詞相当句、(4)動詞相当句の活用型、(5)使役のときにどのように言い換えるか、の5つのフィールドから構成される。このうち、(5)は前章の分析から次の3つのいずれかで与える。

- a. 自動詞を「せる」（もしくは「させる」）で活用
- b. 他動詞に言い換え
- c. 表層ヲ格が有情か非情かで a, b のどちらかを選択

3.3 自動詞-他動詞対応辞書

自動詞-他動詞対応辞書の一部を表4に示す。この辞書の1レコードは、(1)自動詞、(2)自動詞の活用型、(3)自動詞に対応する他動詞、(4)他動詞の活用型、の4つのフィールドから構成される。この辞書は使役の型が「b. 他動詞に言い換え」か「c. 表層ヲ格が有情か非情かで a, b のどちらかを選択」で b が選択された場合と、受動表現の他動詞を自動詞に言い換える場合に使用される。

3.4 活用辞書

活用辞書は、単語の活用形を変更するときに使用する。本システムでは、形態素解析システム JUMAN¹⁰⁾に含まれる活用辞書を、そのまま利用した。

3.5 有情/非情名詞辞書

有情/非情名詞辞書は、それぞれの名詞について有情/非情を 0/1 で記述したものである。この辞書は、名詞が有情か非情かを判断する必要が生じたときに使用される。

4. 言い換えシステムのための知識獲得

本章では、言い換えシステムの「サ変名詞-動詞相当句辞書」「自動詞-他動詞対応辞書」「有情/非情名詞辞書」をどのように作成したかについて述べる。

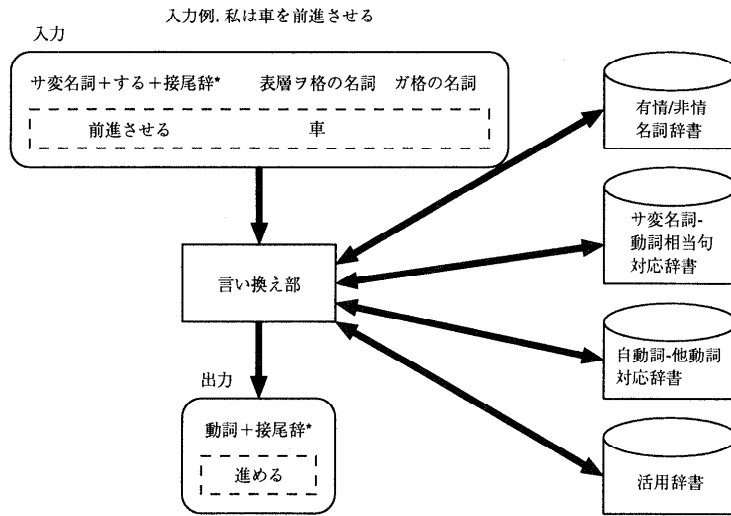


図8 言い換えシステム構成図
Fig. 8 Configuration of the paraphrase system.

表3 サ変名詞-動詞相当句対応辞書 (一部)
Table 3 Examples of the sahen-noun-corresponding verb dictionary.

| サ変名詞 | 自/他動詞 | 動詞相当句 | 活用例 | 使役の型 |
|------|-------|-------|--------|----------|
| 直結 | 自動詞 | 結び付く | 子音動詞カ行 | 他動詞 |
| 参加 | 自動詞 | 加わる | 子音動詞ラ行 | 「せる」で活用 |
| 前進 | 自動詞 | 進む | 子音動詞マ行 | ヲ格に依存 |
| 欠席 | 他動詞 | 休む | 子音動詞マ行 | 「せる」で活用 |
| 入手 | 他動詞 | 手に入れる | 母音動詞 | 「させる」で活用 |
| 根絶 | 他動詞 | 絶やす | 子音動詞サ行 | 「せる」で活用 |

表4 自動詞-他動詞対応辞書 (一部)
Table 4 Examples of the intransitive-transitive verb dictionary.

| 自動詞 | 活用例 | 他動詞 | 活用例 |
|------|--------|-------|--------|
| 結び付く | 子音動詞カ行 | 結び付ける | 母音動詞 |
| 進む | 子音動詞マ行 | 進める | 母音動詞 |
| 絶える | 母音動詞 | 絶やす | 子音動詞サ行 |

4.1 サ変名詞-動詞相当句対応辞書

サ変名詞に対応する動詞相当句は、機械的に作成したものを参考に決定した。機械的な作成では、まずEDR日本語単語辞書¹⁷⁾からサ変名詞とそれに対応する動詞の組を抽出した。

EDR日本語単語辞書には、単語見出しに対して日本語概念説明が付けられている。日本語概念説明は、単語見出しをやさしく言い換えたものと考えられることができる。そこで、サ変名詞を単語見出しとするレコードに対して、単語見出しと日本語概念説明の末尾の動詞の組を抽出した。抽出の例を以下に示す。

例「挑戦する - 困難な物事に挑む」→「挑戦 - 挑む」

このようにして、EDR日本語単語辞書から16,479組を抽出したが、これらの中には意味の合わないものも含まれていた。そのため、サ変名詞と動詞の意味的

な整合を、EDR概念体系辞書¹⁶⁾、分類語彙表⁷⁾、角川類語新辞典¹⁵⁾を使用して判定し、4,802組を得た。

この組を参考にしながら、「サ変名詞-動詞相当句対応辞書」に登録する動詞相当句を選定した。動詞が「サ変名詞+する」に対して意味的に欠落がないと思われるものは、そのままの形で登録した。意味的な欠落が見られるものは、必要と思われる形容詞や副詞などの修飾語、もしくは格を、手で補った。たとえば、「漸増する」というサ変名詞に対して、辞書からは「増える」という動詞が得られる。しかし、「漸増する」の「漸」には「だんだん」という「増える」を修飾する意味があり、「漸増する」を単に「増える」と言い換えたのでは、この「漸」の意味が欠落してしまう。そのため、人手で副詞を補い、「だんだん増える」を動詞相当句として登録する。「増加する」にはそうした関係が見られ

ないため、「増える」をそのまま登録する。

こうした判断は人手により行ったため、個人の揺らぎという点で問題が残る。本稿では、言い換えシステム実験のための辞書作成という観点から、便宜的に上記の基準により判断を行った。より精度の高い辞書を作成するためには、たとえば上記の「漸増」と「増加」の違いを語構成の違いと捉え⁴⁾、サ変名詞を語構成により分類し、それを手がかりとしてどのように言い換えるかを決定する方法などが考えられる。

登録するサ変名詞と動詞の組は、予備実験により選択した。その際、頻繁に現れるが、対応する動詞が辞書から得られなかったサ変名詞に対して、相応の動詞相当句が考えられるならば、それも追加登録した。多義の語で、辞書から複数の組がとれたものについては、一般的に多く使用される語義を考え、人手で一意に決定した。1つの意味の組しかとれなかった多義語についても、それが一般的に多く使用される語義でなく、予備実験の結果、他に相応な動詞相当句が考えられた場合には、そちらを採用した。このように、多義の「サ変名詞+する」の語義は便宜的に人手により決定し、辞書には1つのサ変名詞に対して1つの動詞相当句のみを登録した。

以上の方法により、361 エントリの「サ変動詞-動詞相当句対応辞書」を作成した。

「サ変名詞+する」が自動詞であるか他動詞であるかは岩波国語辞典¹⁴⁾により決定し、自動詞と他動詞の両方の用法がある語は便宜的に予備実験により使用頻度の高い方に決定した。活用型はJUMANに準拠し、使役の型は人手で判断して与えた。

4.2 自動詞-他動詞対応辞書

自動詞-他動詞対応辞書は、「サ変名詞-動詞相当句対応辞書」に登録した動詞について、対応する自動詞もしくは他動詞をIPAL 動詞辞書¹¹⁾を参考に決定し、作成した。ただし、自動詞的用法を持つ「形容詞+なる」と他動詞的用法を持つ「形容詞+する」は、対応する自動詞-他動詞の組として登録した。

4.3 有情/非情名詞辞書

有情/非情名詞辞書は、IPAL 名詞辞書¹²⁾とJUMANの辞書より作成した。

IPAL 名詞辞書で使用されている意味素性はNAT(自然物)、MIN(精神作用)など56種類がある。意味素性は名詞の性質を表していることから、有情か非情かの判断に利用できる。多義の名詞はそれぞれの語義について意味素性が付与されており、1つの語義に複数の性質があれば、複数の意味素性が与えられている。

本研究では便宜的に個々の語における有情の意味素

性の数と非情の意味素性の数を比較し、数の多い方を採用した。有情の意味素性としては、56種類中、次の6種類を選定した。

ORG(組織)、GAT(人間の集団)、HUM(人)、REL(関係)、AML(人を含まない動物)、ANI(人を含む動物)

一方、JUMAN 辞書は名詞辞書が「普通名詞」「副詞的名詞」「形式名詞」「固有名詞」「サ変名詞」「数詞」「時相名詞」の7種類に分かれている。この中で「サ変名詞」は何らかの「動作」を意味する語であり、ヲ格にはなりうるが自発的に動作することはつねに不可能な名詞である。よって「サ変名詞辞書」の語は非情であると判断した。

以上の方法でIPAL 名詞辞書から565 エントリ、JUMAN 辞書から17,237 エントリを得た。これらの異なりをとり、17,271 エントリの有情/非情辞書を作成した。

5. 実 験

前章の言い換えシステムを組み込んだ実験システムを作成し、郵政省発行の「通信に関する現状報告」^{20),21)}を対象に実験を行った。

5.1 実験システム

実験システムは、文章を入力とし、そこに含まれる「サ変名詞+する+接尾辞*」に相当する部分を言い換え、出力する。言い換えシステム中で必要になる表層ヲ格とガ格を文から抽出する部分も、実験のために作成し、システムに組み込んだ。

5.2 表層ヲ格決定部

表層ヲ格決定部は、言い換えるサ変名詞から文頭に向かい助詞「を」を探し、見つからなければ助詞「は」と助詞「も」を探し、表層ヲ格を決定する。ただし、次の例のような場合は数詞の前の名詞を選択する。

例. 試合で選手 2 名を退場させた。

このアルゴリズムの精度を確認するための予備実験を行った。生活白書^{5),6)}、犯罪白書^{18),19)}から使役表現の文を各50文ずつ、合計100文を選択し対象とした。正解を人手で与え評価した結果、精度は75%であった。なお精度には表層ヲ格がないものをないと正しく判定できたケースも含んでいる。

5.3 ガ格決定部

ガ格決定部は、言い換えるサ変名詞から文頭に向かい助詞「が」を探し、見つからなければ助詞「は」と助詞「も」を探す。それでも見つからない場合には言い換え部分の直後にある名詞を探し、ガ格を決定する。直後が名詞句や複合名詞の場合には、最後の名詞だけ

表5 実験結果
Table 5 Evaluation.

| | 言い換え | | 言い換え なかった | 合計 | 精度 (%) |
|------|------|----|--------------|-----------|-----------|
| | 成功 | 失敗 | | | |
| 8年度版 | 276 | 53 | 63 (31) | 392 (148) | 70.4 |
| 9年度版 | 317 | 53 | 72 (34) | 442 (158) | 71.7 |

(異なり数)

表6 誤りの原因の分析
Table 6 Error analysis.

| 原因 (重複を含む) | 8年度版 | 9年度版 |
|---------------------|------|------|
| 語義が合っていない | 45 | 48 |
| 自動詞と他動詞の違い | 5 | 5 |
| 動詞相当句の格と文中の格の重複 | 5 | 2 |
| 動詞相当句の格と文中の修飾語が合わない | 0 | 0 |
| 動詞相当句と文中の格が合わない | 0 | 2 |
| 動詞相当句と文中の修飾語が合わない | 0 | 0 |
| 合計 | 53 | 53 |

を参照する。

このアルゴリズムについても表層ヲ格と同様に生活白書と犯罪白書を対象に予備実験を行った結果、精度は79%であった。

5.4 対象テキスト

対象テキストとして、サ変名詞の使用頻度の高い公文書の中から郵政省発行「通信に関する現状報告」平成8年度版、平成9年度版を選んだ。これらの文書は、WWW上で公開されているものを入手し、タイトルや目次など文章でない部分を除き、半角を全角に換えるなど整形を行い使用した。

5.5 結果

人手で与えた正解と比較した結果を表5に、その失敗の原因の内訳を表6に示す。

表5における「言い換えなかった」は、動詞として使用されているサ変名詞が、「サ変名詞-動詞相当句対応辞書」に登録されていなかったものである。また、表6に示した失敗の原因の項目は、本実験システムにより事前に行った予備実験において見られた問題を基に定めたため、結果的に本実験では発生していない問題も含まれている。また、語義の誤りと、自動詞/他動詞の誤りが同時に発生した場合、原因はそれぞれの項目でカウントしているが、合計では1と数えている。

6. 議論

失敗の90%が、本研究で扱わなかった多義の問題と、自動詞か他動詞かの決定の問題で占められた。本システム作成においても、語義や、自動詞と他動詞の決定は辞書作成の段階で行い、これらの問題を扱わなかったことから、その失敗についても議論しない。本

章では、本実験で見られたそれ以外の問題と、本実験では見られなかったが予備実験で見られた問題を大きくコロケーションの問題としてまとめ、ここであわせて議論する。

6.1 格のコロケーション

本実験で使用したサ変名詞-動詞相当句対応辞書では、意味の欠落を防ぐために必要と思われる場合には、動詞に副詞や形容詞などの修飾語や、格を付属させた形で登録した。そうした動詞相当句に言い換えた結果、動詞相当句に含まれる格が文の他の要素と合わず、不自然になることがあった。

6.1.1 格と格

例1 日本経済ではこれまで新卒者ないし若年層が新たな成長分野に多く職に就く(就職する)ことにより、当該分野が実際に成長していくうえで必要な労働力が確保されてきた。

例1は、「就職する」を「職に就く」と言い換えた結果、「就く」の対象を表す用法である、「職に」と「成長分野に」という格が重複し、不自然な文になっている。しかし、同様に「就職する」を「職に就く」に言い換えた例でも、元の文に「就く」の対象を表す二格が存在しなければ不自然な文とはならず、「何に」就いたのかを明確にする言い換えとなった例もあった。

こうした場合、意味内容が同じか、もしくは一方が他方を詳細に述べているのならば、詳細な方をとることで重複は解消できる。しかし、違う意味を持っている場合、情報を欠落させないためには、どちらかの格を別な格に振り替えるなどの処理が必要となる。

6.1.2 格と修飾語

次の例では、動詞相当句に含まれる格と、元々から

文に含まれている修飾語が合わず、不自然な文になっている。

例2 日本の雇用慣行の下で、仕事をいったん途中でやめる（中断する）と、就業条件が大きく悪化するという事情なども反映していると思われる。

例2は「中断する」を「途中でやめる」に言い換えた結果、「いったん途中でやめる」となった。この表現は間違いではないが、「途中でいったんやめる」という語順の方が自然である。これは「いったん」という副詞が、「途中で」という格より動詞のそばにあるのが自然に感じられるからである。

副詞に係る「サ変名詞+する」を、格を含む動詞相当句で言い換える場合には、語順について考慮する必要があるだろう。

6.2 動詞相当句のコロケーション

意味的に「サ変名詞+する」に対応した動詞相当句であるが、文の他の要素と合わず、不自然になる例も見られた。

6.2.1 動詞相当句と格

次の例は、動詞相当句が元々の文の格と合わず、不自然な言い換えになっている。

例3 これを受けて ARIB では、検討に始まっている（着手している）。

「検討に着手する」の「着手する」は自動詞であり、「何かを新しく始める」の意である。そこで自動詞「始まる」に言い換えた結果、例3に示したように「検討に始まる」という不自然な文となった。これは、「着手する」が二格をとる自動詞であるのに対し、「始まる」は二格をとらない自動詞であるからである。

また、「着手する」が自動詞であるために、「始まる」という自動詞に言い換えたが、意味を考えるならば「始める」という他動詞に言い換える方が自然であるようにも考えられる。しかし、そうすると「検討に始めている」となり、これも不自然である。「検討」は「始める」の対象であり、「検討を始めている」と言い換えるのが最も自然である。

対応する動詞相当句の決定には、文中の格とのコロケーションも考慮する必要がある。さらに、格を変更することも含めた言い換えを考えるならば、自動詞から他動詞、またはその逆の言い換えもありうる事が分かる。

6.2.2 動詞相当句と修飾語

次の例は、動詞相当句が元々の文の副詞と合わず、不自然な言い換えとなっている。

例4 日本の雇用慣行の下で、仕事をいったん中断

すると、就業条件が大きく悪くなる（悪化する）という事情なども反映していると思われる。

例4は、「悪化する」を「悪くなる」に言い換えている。しかし、「大きく悪化する」を「大きく悪くなる」に言い換えると座りが悪い。「非常に悪くなる」「とても悪くなる」などの方が適当であるだろう。これは副詞と動詞相当句とのコロケーションの問題であり、「サ変名詞+する」の言い換えだけでは解決できない問題である。

7. ま と め

本稿では、「サ変名詞+する」からより平易な動詞相当句への言い換えを機械的に実現する方法を提案した。この方法は、5つの言い換え規則と4つの規則適用アルゴリズムから構成される。規則適用アルゴリズムは、テイル形、使役表現、受動表現、可能表現のそれぞれの場合に、言い換え規則を適用するか、適用するならばどの規則にするかを決定するために用いられる。これらのアルゴリズムは、実行のために、サ変名詞-動詞相当句対応辞書の他に、自動詞-他動詞対応辞書、有情/非情名詞辞書、活用辞書の合計4つの辞書を使用する。本稿ではこれらの規則とアルゴリズムに基づいた言い換えシステムを作成し、その有効性を確認した。ただし、辞書は実験のために作成したものであり、語義も人手で一意に決定している。そのため、辞書作成の際の個人の揺らぎや多義性の解消という問題は残されている。

本研究は、主に科学的興味、すなわち、言い換えを実現するための機構とそれに必要な知識を明らかにすることを通して、人間の言語を扱う能力の一端を明らかにしようという立場に立って進めたが、言い換えの研究を行うことは、工学的にも意味がある⁸⁾。本研究で取り上げた「サ変名詞+する」の言い換えは、文章を平易化するシステムにおいて、その手法の1つとして用いることができる。文章の平易化システムは、たとえば、日本語を母国語としない人々が比較的難しい表現の文章を読む場合の、ある種の支援システムとして有用である。

参 考 文 献

- 1) Dras, M.: Reluctant Paraphrase : Textual restructuring under an optimization model, *Proc. Pacling97*, pp.48-55 (1997).
- 2) Dras, M.: Representing Paraphrases Using Synchronous Tree Adjoining Grammars, *Proc. 1997 Australasian NLP Summer Workshop*, pp.17-24 (1997).

- 3) 益岡隆志, 田窪行則: 基礎日本語文法(改定版), くろしお出版(1994).
- 4) 乾 裕子, 元吉文男, 井佐原均: 語構成論に基づいたサ変動詞の分類, 情報処理学会研究会報告, 95-NL-110, pp.103-110 (1995).
- 5) 経済企画庁: 平成8年国民生活白書(1996).
- 6) 経済企画庁: 平成9年国民生活白書(1997).
- 7) 国立国語研究所: 分類語彙表, 秀英出版(1964).
- 8) 佐藤理史: 論文表題を言い換える, 情報処理学会研究会報告, 98-NL-127, pp.187-194 (1998).
- 9) 山崎邦子, 三上 真, 増山 繁, 中川聖一: 聴覚障害者用字幕生成のための言い替えによるニュース文要約, 言語処理学会第4回年次大会発表論文集, pp.646-649 (1998).
- 10) 松本裕治, 黒橋禎夫, 山地 治, 妙木 裕, 長尾 真: 日本語形態素解析システム JUMAN 使用説明書 version3.3 (1997).
- 11) 情報処理振興事業協力技術センター: 計算機用日本語基本動詞辞書 IPAL マニュアル(1996).
- 12) 情報処理振興事業協力技術センター: 計算機用日本語基本名詞辞書 IPAL マニュアル(1996).
- 13) 須賀一好, 早津恵美子: 動詞の自他, ひつじ書房(1995).
- 14) 西尾 実, 岩淵悦太郎, 水谷静夫: 岩波国語辞典, 岩波書店(1994).
- 15) 大野 晋, 浜西正人: 角川類語新辞典, 角川書店(1981).
- 16) 日本電子化辞書研究所: EDR 概念体系辞書マニュアル(1996).
- 17) 日本電子化辞書研究所: EDR 日本語単語辞書マニュアル(1996).
- 18) 法務総合研究所: 平成8年犯罪白書(1996).
- 19) 法務総合研究所: 平成9年犯罪白書(1997).
- 20) 郵政省: 平成8年通信に関する現状報告(1996).
- 21) 郵政省: 平成9年通信に関する現状報告(1997).

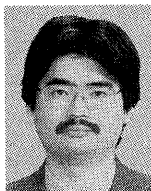
(平成10年10月19日受付)

(平成11年9月2日採録)



近藤 恵子 (学生会員)

1994年東京工芸大学工学部電子工学科卒業。1996年同大学院工学研究科修了。同年より北陸先端科学技術大学院大学に在学。



佐藤 理史 (正会員)

1983年京都大学工学部電気工学科第二学科卒業。1988年同大学院博士課程研究指導認定退学。京都大学工学部助手を経て、1992年より北陸先端科学技術大学院大学情報科学研究科助教授。1997年より科学技術振興事業団研究員を兼任。1998年より京都大学大学院情報学研究科社会情報学専攻助教授を併任。京都大学博士(工学)。自然言語処理, 機械学習, 情報の自動編集等の研究に従事。言語処理学会, 日本認知科学会, AAAI, ACL各会員。著書:「自然言語処理」(共著, 岩波書店, 1996), 「アナロジーによる機械翻訳」(共立出版, 1997), 「言語情報処理」(共著, 岩波書店, 1998)等。



奥村 学 (正会員)

1962年生。1984年東京工業大学工学部情報工学科卒業。1989年同大学院博士課程修了。同年, 東京工業大学工学部情報工学科助手。1992年北陸先端科学技術大学院大学情報科学研究科助教授, 現在に至る。工学博士。自然言語処理, 知的情報提示技術, 語学学習支援, 語彙知識獲得に関する研究に従事。人工知能学会, AAAI, 言語処理学会, ACL, 認知科学会, 計量国語学会各会員。