

「サ変名詞+する」の動詞への言い換え

近藤 恵子, 佐藤 理史, 奥村 学

e-mail: {k-kondo,sato,oku}@jaist.ac.jp

北陸先端科学技術大学院大学 情報科学研究科

【概要】

本稿では言い換えの一研究として、「サ変名詞+する」をより簡単な表現に言い換える方法について述べる。このような言い換えは、変換テーブルによる単純な置換で可能であると思われるがちである。確かに、多くの場合は、サ変名詞を対応する動詞に相当する句に言い換え、活用と接尾辞を写像することで言い換えは可能である。しかし、「サ変名詞+する」は接尾辞の付加によりヴォイスを表現する。ヴォイスは変換テーブルによる単純な置換と接尾辞のコピーでは写像できず、言い換えのために特別な方法を必要とする場合がある。本稿は言い換えの方法を分析、明らかにし、それに基づいたパラフレーズシステムを実装、実験を行ないその結果を示す。

Paraphrase of “sahen-noun + SURU”

KONDO Keiko, SATO Satoshi, OKUMURA Manabu

School of Information Science, Japan Advanced Institute of Science and Technology
(Tatsunokuchi Ishikawa 923-1292 Japan)

Abstract

This paper proposes a method of paraphrase of “sahen-noun + SURU” to a simpler expression. In most cases, this paraphrase can be achieved by replacing a “sahen-noun + SURU” into the corresponding verb. However, there are particular suffixes that require special handling when they are attached to “sahen-noun + SURU”; the usual replacing rule cannot be applied. This paper describes the details of both the usual replacing rule and special handling rules, an implementation of the paraphrase system, and an experimental result.

1 はじめに

言語表現には多様性があり、同じ内容でも様々な表現が考えられる。もし状況や相手との関係、内容に対する相手の理解度等が異なれば、伝える内容が同じであっても違った表現が用いられる。それらは、それぞれの条件により互いに言い換えられているとも言えるだろう。言い換えのメカニズムを明らかにすることは、多様な言語表現をコンピュータで扱っていく上で重要である。

本稿では、言い換えについての一研究として「サ変名詞+する」をより簡単な表現に言い換える方法を検討する。「サ変名詞+する」は、活用形および接尾辞の付加という形でテンスやヴォイスが表現される。そのため、変換テーブルを用いた単純な語の置換では、言い換えを実現することが出来ない。

本稿では、第2章において言い換えの具体例を検討し、それに基づき言い換えの方法を明らかにする。第3章でその方法に基づく言い換えシステムの実現方法について述べる。第4章ではその言い換えシステムに必要な各種情報を既存辞書から獲得する方法について述べる。第5章で実験と検討、第6章で議論を行ない、第7章でまとめを述べる。

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、言い換えの方法を明らかにする。“*”は正規表現に従い0個以上を意味する。

言い換え先は動詞一語の他に、「形容詞+する」「形容詞+なる」や「手に入れる」等の熟語、「だんだんに増える」等の動詞に係る句を含んだものとした。本稿ではこれらを動詞に相当する句という意味で、動詞一語も含め、動詞相当句と呼ぶ。なお、動詞相当句の活用型、活用形という時には、動詞相当句の末尾に含まれている動詞もしくは「なる」「する」の活用型と活用形を指すこととする。

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。

例1. 入学式に桜が開花する → 入学式に桜が咲く

例2. 入学式に桜が開花した → 入学式に桜が咲いた

例3. 入学式に桜が開花しない → 入学式に桜が咲かない

サ変名詞「開花」は「咲く」に言い換えられ、「する」の活用形は「咲く」の活用形にコピーされる。例3において接尾辞「ない」はそのままの形でコピーされている。「形容詞+する」や熟語へも、同様な方法で言い換えが可能である。このような単純な言い換えの方法は、図1のように表すことができる。

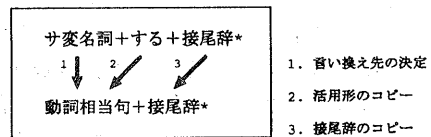


図1: 単純な言い換え

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、図1に示した方法では、自然な言い換えにならない場合がある。

2.2.1 テイル形

動作の継続を意味する表現に「動詞(テ形)+接尾辞“いる”」という表現がある。これは「テイル形」とも呼ばれる。テイル形は言い換える時にコピーできる場合とできない場合がある。

例4. 入学式には桜が開花しているだろう → 入学式には桜は咲いているだろう

例5. 文明が存在していた → 文明があった(×文明があっていた)

例4の言い換えはテイル形をコピーできるが、例5の言い換えはテイル形をコピーできない。例5が「あっていた」にならないのは、「ある」という動詞にはテイル形が存在しないからである。このような動詞として他に「いる」「おる」がある。「サ変名詞+して+いる+接尾辞*」から「ある」「いる」「おる」への言い換えの場合、テイル形はコピーされないことが分かる。

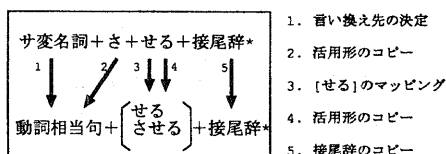
2.2.2 使役表現

「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。

例6. 遊んでいる生徒を練習に参加させる → 遊んでいる生徒を練習に加わらせる

例7. 嘘をついて信用させる → 嘘をついて信じさせる

使役を表現する接尾辞として「せる」と「させる」のどちらが付くかは、動詞相当句の活用型によって決定される。「サ変名詞+する」の「する」はサ変動詞であるため、「せる」が付随している。言い換え先の活用型がサ変動詞か子音動詞ならば「せる」に、それ以外ならば「させる」にマッピングされる。この言い換えは図2に示した方法で実現される。



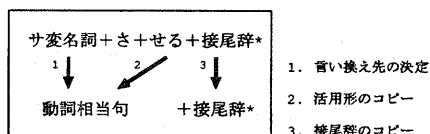
1. 言い換え先の決定
2. 活用形のコピー
3. 「せる」のマッピング
4. 活用形のコピー
5. 接尾辞のコピー

図2: 使役表現の言い換え:「せる」のマッピング

しかし、次のような言い換えも見られる。

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

例8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。この言い換えは図3のように表現することができる。



1. 言い換え先の決定
2. 活用形のコピー
3. 接尾辞のコピー

図3: 使役表現の言い換え:他動詞化

しかし、自動詞の使役表現が常に他動詞になるわけではない。例えば、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なわれていない。

図2の方法で言い換えるか、図3の方法で言い換えるかは、主にサ変名詞によって決定出来る場合が多いが、例外もある。

例9. 彼を前進させる → 彼を進ませる

例10. コマを前進させる → コマを進める

例9は図2の方法で言い換えられており、例10は図3の方法で言い換えられている。この「進ませる」と「進める」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[7]。

直接作用か間接作用かは、正確には文脈から読みとらねばならない。しかし、例9の「彼」は抱えられて進められたとは思えない、例10の「コマ」が促されて自分から進んだとも思えない。これは「彼」が意志を持ち自力で動けるものだからであり、それに対して「コマ」は意志を持たず自分から動かないものだからである。これを「有情」「非情」という[1]。

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、図2と図3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決定できる。

以上をまとめると次のようになる。

1. 必ず「せる」もしくは「させる」をマッピングするサ変名詞 → 図2
2. 必ず他動詞に言い換えるサ変名詞 → 図3
3. それ以外のサ変名詞
 - (a) 表層ヲ格の名詞が有情 → 図2
 - (b) 表層ヲ格の名詞が非情 → 図3

2.2.3 受身表現

「サ変名詞+する」の受身は「サ変名詞+さ+接尾辞“れる”」で表現される。

例11. 伝統が継承されてきた → 伝統が受け継がれてきた

例12. そう信用されても困る → そう信じられても困る

例11, 12の言い換えは図4のように表すことができる。「れる」と「られる」のどちらが付くかは、動詞相当句の活用型によって決定される。サ変動詞

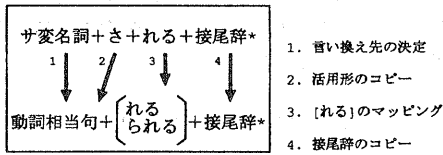


図 4: 受身表現の言い換え

か子音動詞ならば「れる」に、それ以外ならば「られる」にマッピングされる。

しかし、「形容詞+する」という動詞相当句に言い換える例では、図4とは違ったマッピングが行なわれる。その例をあげる。

例 13. 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

例 14. 組合の努力によって労働時間が短縮された → 組合の努力によって労働時間が短くなった

例 13, 14 の言い換えはそれぞれ図5と図6のように表すことができる。

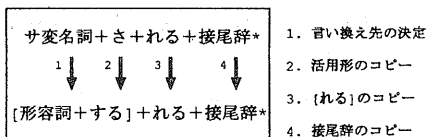


図 5: 受身表現の言い換え: 「形容詞+する」

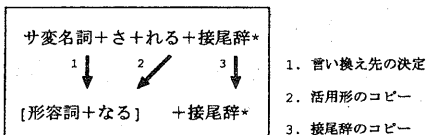


図 6: 受身表現の言い換え: 「形容詞+なる」

例 13 と例 14 の違いは主体の有情と非情にある。受身表現には被害のニュアンスが含まれ、それを残したい時には図5の方法で言い換えなければならない。しかし主体が非情である時には被害のニュアンスは存在しないため、図6のように言い換えられ

る。例 14 では「労働時間」は被害にあったわけではなく、図6によって言い換えられた。

以上をまとめると、次のようになる。

1. 言い換え先が形容詞を含まない動詞相当句 → 図4
2. 言い換え先が形容詞を含む動詞相当句
 - (a) 主体が有情 → 図5
 - (b) 主体が無情 → 図6

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。

例 15. 彼女は英語で会話できる → 彼女は英語で話せる

例 16. 彼を信用できますか → 彼を信じられますか

例 17. 発表時間を短縮できない → 発表時間を短くできない

これらの言い換えは次のように整理できる。

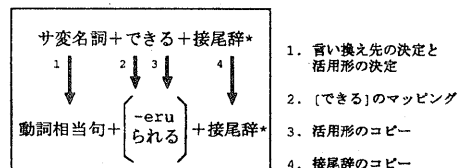


図 7: 可能表現の言い換え: 形容詞以外の動詞相当句

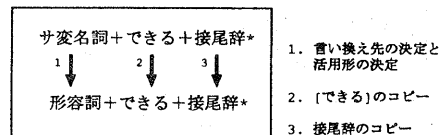


図 8: 可能表現の言い換え: 形容詞を含む動詞相当句

1. 言い換え先が形容詞を含む動詞相当句 → 図8

2. 言い換え先が形容詞を含まない動詞相当句 → 図7

図7の場合の動詞相当句の活用型と活用形、「できる」のマッピングは、動詞相当句の活用型に基づき、表1のようにまとめられる。

表1: 「できる」のマッピング

品詞	活用型	活用形	マッピング先
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

3 言い換えシステム

前章の方法を確認するために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9に示す。

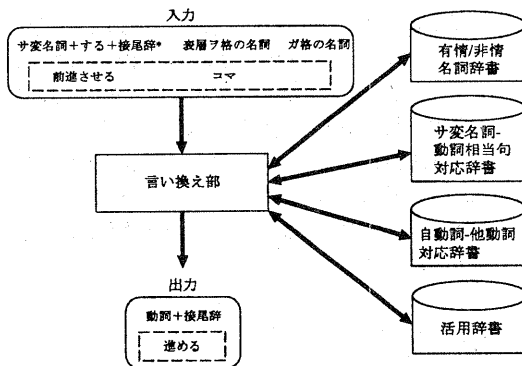


図9: 言い換えシステム構成図

入力は「サ変名詞+する+接尾辞*」と「表層ヲ格の名詞」「ガ格の名詞」である。言い換え部は2章で述べた方法に基づき、必要に応じて各辞書を参照し、入力を言い換える。

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2に示す。この辞書は、(1) サ変名詞、(2) 「サ変名詞+す

る」の時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) 使役の時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ている。

使役表現の言い換えの型は前章の分析から次の三つのいずれかで与える。

- 自動詞を「せる or させる」で活用
- 他動詞に言い換え
- 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3に示した。この辞書は、(1) 自動詞、(2) 自動詞と対応する他動詞、(3) 他動詞の活用型、の3つのフィールドから成っている。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 が選択された時に使用される。

表3: 自動詞-他動詞対応辞書

自動詞	他動詞	活用型
結び付く	結び付ける	母音動詞
進む	進める	母音動詞

3.4 活用辞書

形態素解析システム JUMAN に含まれる活用分類辞書を利用した（本研究の動詞の活用型と活用形は、JUMAN[5] に準拠している）。

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、使役表現と受身表現を言い換える際、名詞が有情か非情かを判断する必要が生じた時に、使用される。

表 2: サ変名詞-動詞相当句対応辞書のレコード

サ変名詞	自/他動詞	動詞相当句	活用型	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

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」「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。

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 [10] 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。これは難しいものを易しくする言い換えの一種であると考えられる。そこでサ変名詞を見出しとするレコードについて、見出しと日本語概念説明の末尾の動詞の組を作成した。作成の例を示す。

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 組を作成したが、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 [9]、分類語彙表 [4]、角川類語事典 [8] 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」には 361 エントリを登録した。

「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 有情/非情名詞辞書

名詞が有情か非情かを判断する名詞辞書は、IPAL 名詞辞書 [6] 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) 等 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、一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。

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、ヲ格にはなり得るが自発的に動作することは常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登録した。

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。これらの異なりを取り、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、「通信に関する現状報告」[13][14] を対象に実験を行なった。

5.1 実験システム

実験のためのシステムは、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決定する部分も、実験のために作成し組み込んだ。

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。

このアルゴリズムの精度を確認するために実験を行なった。生活白書 [2][3]、犯罪白書 [11][12] 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無いものを無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。

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 [2][3] と犯罪白書 [11][12] を対象に実験を行なった結果、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、漢語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。「通信に関する現状報告」は WWW 上で公開されているものを入手、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ったものを使用した。

5.5 結果

人手で与えた正解と比較した結果を表 4 に、その失敗の原因の内訳を表 5 に示す。

表 4: 実験結果

	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)

失敗の原因の項目は事前に行なった予備実験の際に見られた問題を元に定めたものであるため、結

表 5: 失敗の原因

原因	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
格の過不足	5	2
コロケーションの問題	0	0
自動詞が他の格と合わない	0	2
合計	53	53

果的に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両方発生したものについては、それぞれの原因の項目でカウントしているが、合計は一つの言い換えに対して失敗の原因が複数であっても一つとして計算している。

6 議論

失敗の 90% が、今回扱わなかった多義、自動詞他動詞の決定の問題で占められた。ここでは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コロケーションの問題について議論する。

6.1 格の過不足

句に言い換える場合、句の形によって問題が起きる場合と起きない場合がある。

- 例 18. 消費者自らも契約の条件や内容を理解し、その安全性に十分留意し、消費者被害に関する情報などを手に入れていく（入手していく）ことが大切である。

入手する → 手に入れる

- 例 19. 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上で必要な労働力が確保されてきた。

就職する → 職に就く

例 18 における「入手する」は二格を採らないため、「手に入れる」に言い換えても同じ二格が重複することはない。一方、例 19 の場合には「就職する」は二格を採ることができ、例文では「成長分野に」という格を二格として採っている。そのために

「職に就く」に言い換えた時に「職に」という二格が「成長分野に」という二格と重複してしまい不自然な文となる。

6.2 コロケーション

「サ変名詞＋する」と動詞相当句が意味的に正しく言い換えられたにも関わらず、言い換えた動詞相当句が前後の語と相性が悪いという問題が見られた。

- 例 20. 日本的な雇用慣行の下で、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。

悪化する → 悪くなる

「悪化する」は「悪くなる」に言い換えられる。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」等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。

6.3 自動詞の種類

自動詞の採り得る格を考慮しなかったために不自然な言い換えになったケースもあった。

- 例 21. これを受けてARIBでは、検討に始まっている(着手している)。

着手する → 始まる

「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21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。

動詞はヲ格を採る他動詞と採らない自動詞に分けられ、自動詞は更に二格を採る自動詞と採らない自動詞に分けられる。「着手する」は二格を採る自動詞であり、「始まる」は二格を取らない自動詞であるために、言い換えが不自然になったものと思われる。この問題は「サ変名詞＋する」に対応する動詞相当句を決定する際に、自動詞か他動詞かというだけでなく二格を採るか採らないかといったところまで合わせて選定する必要性を示している。

7 まとめ

「サ変名詞＋する」からより平易な表現である動詞相当句への言い換えを分析し、言い換えの機械的な実現方法を明らかにした。「サ変名詞＋する＋接尾辞*」の言い換えは、変換テーブルだけで行なえる容易なものではなく、特にヴォイスの写像は特別な方法を必要とする。それらの分析に基づいたシステムを作成し、実験を行なった結果、約70%の精度をあげた。失敗の90%以上は本研究で扱わなかった多義と自動詞/他動詞の決定の問題が占めていた。他の原因として「サ変名詞＋する」の枠組だけでは解決できない問題が見られた。

参考文献

- [1] 益岡隆志, 田窪行則. 基礎日本語文法 -改訂版-. くろしお出版, 1994.
- [2] 経済企画庁. 平成8年 国民生活白書, 1996.
- [3] 経済企画庁. 平成9年 国民生活白書, 1997.
- [4] 国立国語研究所. 分類語彙表. 秀英出版, 1964.
- [5] 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 version 3.3, 1997.
- [6] 情報処理振興事業協会技術センター. 計算機用日本語基本名詞辞書 IPAL マニュアル, 1996.
- [7] 須賀一好, 早津恵美子(編). 動詞の自他 日本語研究資料集. ひつじ書房, 1995.
- [8] 大野晋, 浜西正人. 角川類語新辞典. 角川書店, 1981.
- [9] 日本電子化辞書研究所. EDR 概念体系辞書マニュアル, 1996.
- [10] 日本電子化辞書研究所. EDR 日本語単語辞書マニュアル, 1996.
- [11] 法務総合研究所. 平成8年 犯罪白書, 1996.
- [12] 法務総合研究所. 平成9年 犯罪白書, 1997.
- [13] 郵政省. 平成8年 通信に関する現状報告, 1996.
- [14] 郵政省. 平成9年 通信に関する現状報告, 1997.