連用修飾表現の省略可能性に関する知識の獲得

酒井　浩之[†]　篠原　直嗣[†]
増山　繁[†]　山本　和英[††]

文内要約の一要素技術として，連用修飾表現の省略可能性に関する知識を獲得する手法を提案する．具体的には，省略できる可能性のある連用修飾表現を含む節に対して，同一の動詞をもち，かつ，格助詞出現の差異が認められる節をコーパスから検索し，検索された節対から省略可能な連用修飾表現を認定する．また，連用修飾表現の内容および前後の文脈を考慮して，重要な情報が多く含まれている連用修飾表現に対しては省略可能と認定できる可能性を低く，逆に，認定対象としている連用修飾表現に，それより以前の文に存在する情報が含まれている場合に対しては，省略可能と認定できる可能性が高くなるような工夫を施した．本手法によって省略可能と認定された連用修飾表現を評価したところ，適合率 78.0%，再現率 67.9%との結果を得た．また，本手法を，格フレーム辞書によって動詞に対する任意格として記述される格要素を，省略可能な連用修飾表現として認定する手法と比較した．その結果，適合率，再現率ともに比較手法より良好な結果を得ることができ，提案手法の有効性を確認した．

キーワード:　連用修飾表現の省略，類似文，コーパス，テキスト自動要約

Knowledge Acquisition about the Abbreviation Possibility of Verb Phrases

Hiroyuki Sakai[†] , Naotsugu Shinohara[†] , Shigeru Masuyama[†]
and Kazuhide Yamamoto[††]

This paper proposes a method of acquiring knowledge about the abbreviation possibility of verb phrases. In a certain clause containing a verb and including verb phrases, the proposed method extracts some clauses which contain the same verb and have different case postpositional particles from a large corpus. Then, our method recognizes verb phrases possible to be abbreviated by comparing the verb phrases with the verb phrases contained in the extracted clauses. In our method, the verb phrases containing important piece of information is hard to recognize as being possible to be abbreviated, and the verb phrases containing information which appear in previous sentences is easy to recognize as being possible to be abbreviated. The evaluation of our method by experiments shows that the precision is 78.0% and the recall is 67.9%. We compare our method with the method which recognizes verb phrases possible to be abbreviated by recognizing optional case elements described in a case frame dictionary as being possible to be abbreviated. By the evaluation results, we conclude that our method outperforms the method which recognizes verb phrases possible to be abbreviated by using a case frame dictionary.

KeyWords:　*Abbreviation of verb phrases, Similar sentence, Corpus, Summarization*

1　はじめに

近年，テキスト自動要約の必要性が高まってきており，自動要約に関する研究が盛んに行なわれてきている (奥村 難波 1999)．要約とは， 人間がテキストの内容の理解，取捨選択をより容易にできるようにするために，元のテキストを短く表し直したものをいう．

これまでの研究で提案されてきた要約手法は，主に次の 3 つに分類される．

- 文書を対象とした，重要文抽出による要約
- 文を対象とした，不要個所削除 (重要個所抽出) による要約
- 文を対象とした，語句の言い換えによる要約

どのような使用目的の要約でも作成できる万能な要約手法は存在しないため，要約の使用目的に応じた手法を選択し，時には複数の手法を併用して要約を作成することが必要となる (山本, 増山, 内藤 1995)．

要約技術の応用はいくつか考えられている．例えば,「WWW 上の検索エンジンの検索結果を一覧するための要約」を作成する場合には，元の文書にアクセスするかどうかを判断するための手掛りとしての役割から，ユーザに読むことの負担を与えないために，簡潔で自然な文が必要となる．したがって，重要文抽出によって作成した要約結果に対し，必要に応じて不要個所削除と語句の言い換えによる要約手法を用いるという方法が適切であると考えられる．また「ニュース番組の字幕生成，及び文字放送のための要約」を作成する場合には，重要文抽出による要約では文書の自然さが損なわれやすいことと情報の欠落が大きすぎること，そしてテキストをそれほど短くする必要がないことなどから，不要個所削除と語句の言い換えによる要約手法を用いることが適切だと考えられる．

このように，要約の使用目的に応じて，それに適した要約手法を用いることで，より効果の高い要約を作成することができる．また，テキストの種類に応じて適切な要約手法もあると考えられる．将来，テキストの種類を自動判別し，ユーザの要求に応じられる要約手法を選択し，テキストを要約するといった要約システムを実現するためには，様々な要約手法が利用可能であることが望まれる．

本論文で提案するのは，不要個所削除による要約を実現するための要素技術である，文中の省略可能な連用修飾表現を認定するために必要な知識を獲得する手法である．不要個所省略による要約手法として，山本ら (山本他 1995) は，一文ごとの要約ヒューリスティックスに基づいた連体修飾節などの削除を提案している．この手法は，重要文抽出による要約結果をさらに要約するという位置付けで提案されているが，単独で用いることも可能である．若尾ら (若尾, 江原, 白井 1997) や山崎ら (山崎, 三上, 増山, 中川 1998) は，人手で作成された字幕とその元と

† 豊橋技術科学大学 知識情報工学系, Department of Knowledge-based Information Engineering, Toyohashi University of Technology.
†† ATR 音声言語コミュニケーション研究所, ATR Spoken Language Translation Research Laboratories.

なったニュース原稿とを人手で比較し，それによって作成した言い換え規則を用いた要約手法を提案している．また，加藤ら (加藤 浦谷 1999) は記事ごとに対応のとれたニュース原稿と字幕放送の原稿を用いて，言い換えに関する要約知識を自動獲得する研究を行なっている．

ところが，これらの手法には次のような問題点がある．まず，不要箇所の削除や言い換えに関する規則を人手で作成するには多大な労力が掛かり，網羅性などの問題も残ることが挙げられる．また，加藤らが使用したような原文と要約文との対応がとれたコーパスは要約のための言語知識を得る対象として有用であるのは明らかであるが，一般には存在しておらず，入手するのが困難である．また，そのようなコーパスを人手で作成するには多大な作業量が必要であると予想される．

このような理由から，本論文では，原文と要約文との対応がとれていない一般のコーパスから，不要個所省略による要約において利用できる言語知識を自動獲得し，獲得した言語知識を用いて要約を行なう手法を提案する．ここで不要箇所の単位として連用修飾表現に注目する．

連用修飾表現の中には，いわゆる格要素が含まれている．格要素の省略は日本語の文に頻出する言語現象である．格要素が省略される現象には次の 2 つの原因がある．

(1)　格要素の必須性・任意性

(2)　文脈の影響

(1)：動詞と共起する格要素には，その動詞と共起することが不可欠である必須格と，そうではない任意格があるとされている (情報処理振興事業協会技術センター（IPA） 1987)．必須格は，主格，目的格，間接目的格など，動詞が表現する事象の内部構造を記述するものであり，任意格は，手段や理由，時間，場所などを記述するものである場合が多い．必須格がないことは読み手に文が不自然であると感じさせる．ただし，必須格でも文脈によって省略可能となる場合があり，任意格についても動詞と共起するのが任意的であるというだけで，文中の任意格が必ず省略可能となるとは限らない．

(2)：本論文における文脈とは，読み手が当該文を読む直前までに得ている情報のことを指す．文脈の影響により省略可能となるのは，読み手にとって新しい情報を与えない格要素，または文脈から読み手が補完するのが容易な格要素である．なお，文脈から省略可能となるのは格要素だけに限らず，格助詞を持たない連用修飾表現においても，文脈から省略可能となる可能性がある．したがって，上で述べたように必須格の格要素でも，それが読み手にとって旧情報であれば省略可能となる場合があり，任意格の格要素でも，読み手にとって新情報であれば，省略することは重要な情報の欠落につながる場合がある．

格要素の必須性・任意性を求めることで，省略可能な格要素を認定する手法として，格フレーム辞書を用いた手法を挙げることができる．現在，利用できる格フレーム辞書としては，IPAL の基本動詞辞書 (情報処理振興事業協会技術センター（IPA） 1987) や日本語語彙大系 (池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林 1997) の構文意味辞書といった人手により収集された

表 1　動詞「進める」の格フレーム

No.	格フレーム	文例
1	N1 ガ N2 ヲ （N3 ニ／ヘ）	彼は 船を 沖へ 進めた。
2	N1 ガ N2 ヲ N3 ニ	彼は 娘を 大学に 進めた。
3	N1 ガ N2 ヲ	彼は 会の準備を 進めている。
4	N1 ガ N2 ヲ	政府は 国の産業を 進めている。

ものがある．また，格フレームの自動獲得に関する研究も数多く行なわれてきている．例えば，用言とその直前の格要素の組を単位として，コーパスから用例を収集し，それらのクラスタリングを行なうことによって，格フレーム辞書を自動的に構築する手法 (河原 黒橋 2002) がある．この手法は，用言と格要素の組合せをコーパスから取得し，頻度情報などを用いて格フレームを生成する．その他には，対訳コーパスからの動詞の格フレーム獲得 (宇津呂, 松本, 長尾 1993) 等がある．本論文で提案する手法は，格要素も含めた省略可能な連用修飾表現を認定する手法であり，その点が格フレーム生成の研究とは異なる．だが，これらの研究で提案されている手法により獲得した格フレームを用いても，省略可能な格要素の認定が実現可能であると考えられる．しかし，格フレームを用いた格要素の省略には次のような問題点がある．

(1)　格要素以外の省略可能な連用修飾表現に対応できない．例えば，節「そのために必要な措置として二百八十二の指令・規則案を定めた。」の動詞「定めた」に対する連用修飾表現「そのために必要な措置として」は文脈から省略可能だが，格要素ではないので格フレーム辞書では対応できない．

　特に，我々の調査の結果，格要素ではない連用修飾表現で省略可能な表現は多数（後述の実験では，省略可能な連用修飾表現のうち，約 55%が格要素ではない連用修飾表現であった）存在する．

(2)　格フレーム辞書に記載されていない動詞に関しては，省略可能な格要素が認定できない．

(3)　動詞の必須格，任意格は，その格の格成分によって変化する．例えば，IPAL 基本動詞辞書において，動詞「進める」の格フレームに関する記述は表 1 のようになっている．この情報から N3 が「大学」である場合のみニ格が必須格になる．このように，たとえ大規模な辞書が構築できたとしても，用例によっては任意格が必須格に変化する場合があり，辞書のような静的な情報では対応できない場合がある．

(4)　格要素を省略可能と認定する場合，読み手が当該文を読む直前までに得ている情報から，省略可能と認定できる場合がある．しかし，格フレーム辞書では静的であるため，文脈を考慮した省略可能な格要素の認定ができない．

(5)　認定対象としている連用修飾表現に重要な情報が含まれていれば，任意格であっても，そのような連用修飾表現を省略してしまえば情報欠落が大きくなる．しかし，格

フレーム辞書では情報の重要度を考慮して認定することができない．

そこで，本論文では，対応する要約文，もしくは格フレーム等を用いない省略可能な連用修飾表現の認定を行なう教師なしの手法を提案する．具体的には，省略できる可能性のある連用修飾表現を含む節に対して同一の動詞をもち，かつ，格助詞出現の差異が認められる節をコーパスから検索し，検索された節対から省略可能な連用修飾表現を認定する．そのため，格フレームでは対処できない格要素以外の連用修飾表現に対しても省略可能かどうかの判定が可能である．また，ある連用修飾表現が省略可能かどうかの判定の際に，その内容および前後の文脈を考慮して，その連用修飾表現に含まれている情報が以前の文にも含まれている情報である場合には，省略可能と認定されやすくなる．逆に，その情報が以降の文に含まれている場合や，重要な情報が含まれている場合には省略可能と認定されにくくなるような工夫を行なっている．

本手法によって抽出された省略可能と認定された連用修飾表現は，その内容および前後の文脈を考慮している上に，格要素以外の連用修飾表現も含まれている．これらは現状の格フレーム辞書にはない知識であり，要約のみならず換言や文生成にも有用であると考える．

本研究でコーパスとして想定するのは，形態素情報などの付与されていない一般のコーパスである．したがって CD-ROM などで提供されている新聞記事のバックナンバーや電子辞書，WWW 上で公開されている文書などを利用することができ，コーパスの大規模化も比較的容易に実現可能である．

以下，第 2 章では，本論文で提案する手法を説明する．第 3 章では，手法を実装して，それによって省略可能と認定される連用修飾表現を示す．第 4 章では，本手法の性能を評価し，評価結果の考察を示す．第 5 章では，格フレーム辞書を用いた手法と本手法によって省略可能と認定された連用修飾表現を比較した実験について述べ，実験結果について考察する．

2　提案手法

本手法では，省略できる可能性のある連用修飾表現を含む節に対して同一の動詞をもち，かつ格助詞出現の差異が認められる節をコーパスから検索し，検索された節対から省略可能な連用修飾表現を認定する．これは「省略できる可能性のある連用修飾表現を含む節に対して，同一の動詞と類似した名詞を含み，かつ，格助詞出現の差異が認められる節が存在し，一方にしか出現しない格助詞を含む連用修飾表現は省略可能である」という仮定に基づいている．ここで，本論文における節とは，ひとつの動詞と，それに対する連用修飾表現をもつ文の構成要素である．なお，本論文では主節も従属節も節とする．以後，省略できる可能性のある連用修飾表現を含有する節を含有節とする．また，含有節と同一の動詞をもち，かつ格助詞出現の差異が認められる節を差異節とする．

2.1　含有節の取得

本手法における含有節は，ある節における動詞に対して 2 つ以上，連用修飾表現が係っている節であるとする．例えば,「市場統合はＥＣが八五年から進めてきた計画.」における動詞「進める」には「ＥＣが」と「八五年から」という 2 つの連用修飾表現が係っている．そのため，節「ＥＣが八五年から進めてきた」は含有節となる．ただし,「する」「ある」「なる」の 3 つの動詞に関して，これらの動詞に係る連用修飾表現を省略すると情報欠落が大きい場合が多いという理由から，精度向上のため，これらの動詞を含む節は含有節としない．

本手法では，精度の向上のため，連用修飾表現が 1 つしか係っていない動詞に対する連用修飾表現を認定対象としない．もし，そのような連用修飾表現を省略可能としてしまうと，その節は連用修飾表現が係っていない動詞になってしまい，情報欠落が大きいものと考える．日本語には，文脈や文末表現によっては連用修飾表現がかかっていない動詞も存在している．しかし，それは一般的に何かの連用修飾表現が省略されている場合であり，文脈から補完ができる情報である．そこまで精度のよい完全な文脈解析ができないため，本手法では精度の向上のため，2 つ以上の連用修飾表現が係っている動詞に対する連用修飾表現を対象とする．すなわち，その動詞に 2 つ以上の連用修飾表現が係っている場合，最も重要な連用修飾表現以外は省略できる可能性がある．よって，本手法を適用した動詞には，少なくとも 1 つの連用修飾表現が必ず係ることになる．

2.2　差異節の検索

取得した含有節に対する差異節をコーパスから検索する．ここで差異節とは，先に定義したように，含有節と同一の動詞をもち，かつ，格助詞出現の差異が認められる節である．すなわち上記の例文「市場統合はＥＣが八五年から進めてきた計画」における差異節は，動詞「進む」をもち，かつ,「ガ格」「カラ格」のうち,「ガ格」をもつ節，および,「カラ格」をもつ節である．ただし，動詞「進む」をもち「ガ格」「カラ格」を両方，含んでいる節は，節対における格助詞出現の差異が認められないため，本手法における差異節ではない．また，厳密には動詞「進む」をもち「ガ格」「カラ格」を両方持たない節も差異節ではあるが，そのような節は省略可能な連用修飾表現の認定に必要な知識の獲得に有用ではないので，本手法においては対象外とする．

例えば，含有節「ＥＣが八五年から進めてきた」に対して，節「市が整備を進めてきた」は，ガ格を含んでいるがカラ格を含んでいないので差異節である．

2.3　省略可能な連用修飾表現の認定

含有節と差異節とを比較して省略可能な連用修飾表現を認定する．ここで，ある含有節に対して k 個の差異節が検索されたとし，以下のように，記号を定義する．

$SP(V)$:　動詞 V をもつ含有節,

$DP_i(V)$:　$SP(V)$ の差異節．なお，$i = 1, 2, \ldots, k$

$C(x)$:　節 x の動詞 V に係る連用修飾表現において，動詞 V に対する格助詞の集合．ただし，$x = SP(V)$ *or* $x = DP_i(V)$,

$E(c_j, x)$:　節 x において格助詞 c_j を含む連用修飾表現．ただし，$c_j \in C(x)$,

例えば，$E(c_1, SP(V))$ と $E(c_1, DP_1(V))$ は，含有節 $SP(V)$ において格助詞 c_1 をともなう連用修飾表現 $E(c_1, SP(V))$ と，差異節 DP_1 において同じ格助詞 c_1 をともなう連用修飾表現 $E(c_1, DP_1(V))$ という関係がある．なお，含有節の動詞 V に対する格助詞を伴わない連用修飾表現の c_j は，その動詞 V に係る品詞によって決定される．品詞は，動詞，名詞，副詞，形容詞，助詞，接続詞，指示詞の 7 つに分類した[1]．例えば,「ＥＣは市場統合に続いて通貨・政治統合を推進する」という節で,「ＥＣは市場統合に続いて」が「推進する」に係っているが，この連用修飾表現は格助詞を伴っていないうえ,「続いて」が動詞なので $c_j =$ “動詞” と定義する．また,「参入分野としては、大蔵省が当初から認める全分野を想定している。」という節で,「参入分野としては」が「想定している」に係っているが，この連用修飾表現は，格助詞を伴っていないうえ，最後に副助詞「は」が含まれているので，$c_j =$ “助詞” と定義する．本論文では格助詞を伴わない連用修飾表現も，格助詞のかわりに他の品詞を伴っている連用修飾表現であるものと定義している．以下，省略可能な連用修飾表現の認定手法を説明する．

Step 1　含有節 $SP(V)$ の各連用修飾表現 $E(c_j, SP(V))$ に対して，含有節における各連用修飾表現の重み $W(E(c_j, SP(V)))$ を計算する．重みが高い連用修飾表現ほど，その含有節において重要であり，省略できない連用修飾表現になる．

$$
\begin{aligned}
W(E(c_j, SP(V))) &= \max_{i=1,2,\ldots,k} (1 + SIM(n_{E(c_j,SP(V))}, n_{E(c_j,DP_i(V))})) \\
&\times \frac{f(V, c_j)}{f(V)} \times (1 + \sum_{n \in N(c_j, SP(V))} B(n, SP(V))) \qquad (1) \\
B(n, SP(V)) &= \frac{1 + after(n, SP(V))}{2(1 + before(n, SP(V)))} \times \log \frac{P}{df(n)} \qquad (2)
\end{aligned}
$$

但し,

$n_{E(c_j,x)}$:　節 x において格助詞 c_j をともなう連用修飾表現 $E(c_j, x)$ において，格助詞 c_j の前に出現する名詞,

例えば,「欧州共同体の市場統合が十二月三十一日で完成する」という含有節 $SP(V)$ における連用修飾表現 $E(c_j, SP(V))$「欧州共同体の市場統合が」では，ガ格 c_j の前に出現する名詞「統合」が $n_{E(c_j,SP(V))}$ である．なお，格助詞を含まない連用修飾表現では，その連用修飾表現の最後に出現する名詞とする．例えば「そのために必要な措置として」では，最後に出現する名詞「措置」が $n_{E(c_j,SP(V))}$ である．

1 品詞情報は形態素解析器として採用した JUMAN version 3.5 に準拠する．

$f(V)$:　含有節 $SP(V)$ に含まれている動詞 V の全コーパスにおける出現頻度,

$f(V, c_j)$:　全コーパスにおいて，動詞 V が格助詞 c_j と共起した頻度,

$N(c_j, SP(V))$:　連用修飾表現 $E(c_j, SP(V))$ に含まれる名詞の集合．ただし，複合名詞の場合は，分解せずに複合名詞で 1 つの名詞として扱う．
例えば，連用修飾表現「欧州共同体の市場統合が」では，{ 欧州共同体，市場統合 } が $N(c_j, SP(V))$ である．

$after(n, SP(V))$:　含有節 $SP(V)$ が存在する文 S より後の文に名詞 n が出現する頻度．ただし $n \in N(c_j, SP(V))$,

$before(n, SP(V))$:　含有節 $SP(V)$ が存在する文 S より前の文に名詞 n が出現する頻度．ただし $n \in N(c_j, SP(V))$,

$df(n)$:　対象とした全コーパスにおいて，名詞 n を含んでいる文書の頻度,

P:　対象とした全コーパスにおける全文書数,

ここで，式 (1) は，先に示した「含有節に対して，類似した名詞を含んだ差異節が存在し，一方にしか出現しない格助詞を含む連用修飾表現は省略可能である」という仮定に基づいて考案した．

式 (1) の第 1 項 $SIM(n_{E(c_j,SP(V))}, n_{E(c_j,DP_i(V))})$ は，含有節 $SP(V)$ の格助詞 c_j を伴う連用修飾表現 $E(c_j, SP(V))$ における名詞 $n_{E(c_j,SP(V))}$ と，差異節 $DP_i(V)$ の格助詞 c_j を伴う連用修飾表現 $E(c_j, DP_i(V))$ における名詞 $n_{E(c_j,DP_i(V))}$ の類似度である．名詞間類似度の計算については後述する．$n_{E(c_j,DP_i(V))}$ は，最大で差異節の数だけ存在するが，類似度は，その中の最大値を採用する．

例えば，含有節の連用修飾表現 $E(c_j, SP(V))$ の名詞 $n_{E(c_j,SP(V))}$ と，差異節の連用修飾表現 $E(c_j, DP_i(V))$ の名詞 $n_{E(c_j,DP_i(V))}$ の類似度が最も高かったとする．その場合，コーパスには連用修飾表現 $E(c_j, SP(V))$ に対して同一の動詞に係り，同一の格助詞，類似した名詞をもつ連用修飾表現が存在したことになる．その最も高い類似度を乗算することで重みを大きくし，そのような連用修飾表現 $E(c_j, SP(V))$ を省略されにくくする．しかし，その他の連用修飾表現（例えば，$E(c_i, SP(V))$）は，$E(c_j, SP(V))$ の重みが大きくなることで，Step 2 の処理によって相対的に重みが小さくなるため省略されやすくなる．

式 (1) の第 2 項 $\frac{f(V,c_j)}{f(V)}$ は，コーパス全体でその動詞 V に対して格助詞 c_j を含んだ連用修飾表現が係る割合であり，動詞 V に対して多く係る格助詞を含む連用修飾表現ほど省略されにくくなる．

式 (1) の第 3 項 $(1 + \sum_{n \in N(c_j,SP(V))} B(n, SP(V)))$ は本手法における文脈補正項と定義し，詳細は後述する．

Step 2　$W(E(c_j, SP(V)))$ を，含有節 $SP(V)$ の動詞 V に係っているいくつかの連用修飾表現の重みの最大値で正規化する．ここで，含有節 $SP(V)$ の動詞 V には m 個の連用修飾表現が係っているものとする．

$$Ws(E(c_j, SP(V))) = \frac{W(E(c_j, SP(V)))}{\max_{i=1,2,\ldots,m} W(E(c_i, SP(V)))} \tag{3}$$

Step 3　$Ws(E(c_j, SP(V)))$ が，ある閾値以下の連用修飾表現を省略可能と認定する．最大値で正規化しているので，複数ある $Ws(E(c_i, SP(V)))$ のどれか 1 つは値が 1 になっており，必ず省略不可能と認定される．ただし，連用修飾表現が提題，ガ格，ヲ格であった場合は無条件に省略不可能と認定する．ここで提題とは，文における主格となる表現である．

2.4　名詞間類似度

名詞間類似度は，コーパス内の動詞と名詞の出現に関する相互情報量からその類似度を検出するヒンドル法 (Hindle 1990)(平岡 松本 1994) を採用した．以下，ヒンドル法について述べる．

Step 1　ある格助詞 c_j において，動詞 v_i と名詞 n_k の出現に関する相互情報量 $MI(c_j, v_i, n_k)$ を求める．

$$MI(c_j, v_i, n_k) = \log \frac{\frac{f(c_j, v_i, n_k)}{N}}{\frac{f(v_i)}{N} \times \frac{f(n_k)}{N}} \tag{4}$$

N**:**　全コーパス中の文の総数,

$f(n_k)$**:**　名詞 n_k の出現頻度,

$f(v_i)$**:**　動詞 v_i の出現頻度,

$f(c_j, v_i, n_k)$**:**　n_k が格助詞 c_j を伴って v_i と共起した頻度,

Step 2　格助詞 c_j と動詞 v_i からみた名詞 n_k, n_l の類似度 $RSIM(c_j, v_i, n_k, n_l)$ を求める．

$MI(c_j, v_i, n_k) > 0$ かつ $MI(c_j, v_i, n_l) > 0$ のとき

$$RSIM(c_j, v_i, n_k, n_l) = min(MI(c_j, v_i, n_k), MI(c_j, v_i, n_l))$$

$MI(c_j, v_i, n_k) < 0$ かつ $MI(c_j, v_i, n_l) < 0$ のとき

$$RSIM(c_j, v_i, n_k, n_l) = |max(MI(c_j, v_i, n_k), MI(c_j, v_i, n_l))|$$

上記以外 のとき

$$RSIM(c_j, v_i, n_k, n_l) = 0$$

Step 3　n_k と n_l の名詞間類似度を次式で求める．

$$SIM(n_k, n_l) = \sum_i \sum_j RSIM(c_j, v_i, n_k, n_l) \tag{5}$$

なお，名詞を一つも含まない連用修飾表現（例えば,「しかし」といった接続詞,「さらに」といった副詞）は，名詞間類似度が算出できない．そのため，そのような連用修飾節の $SIM(n_{E(c_j,SP(V))}, n_{E(c_j,DP_i(V))})$ は 0 になる．

2.5　数値表現の省略

連用修飾表現 $E(c_j, SP(V))$ で，その $n_{E(c_j,SP(V))}$ が数値表現であった場合は，省略可能性の認定手法が異なる．例えば,「欧州共同体の市場統合が十二月三十一日で完成する」という節における連用修飾表現「十二月三十一日で」で，デ格の前に出現する名詞は「三十一」である．この三十一が数値表現なので，この連用修飾表現の省略可能性の認定は以下で述べる手法を用いる．

数値表現の省略は，数値情報の重要性が読み手によって異なるので，省略すべきかどうかの判断が難しい．本手法では，コーパス中で数値情報が頻繁に係る動詞における数値表現は省略しないとする．これを反映させるため，数値表現が含まれている連用修飾表現については，以下のような手法をとる．

Step 1　まず，数値表現は全て “Number” という表現に置き換える．

Step 2　含有節 $SP(V)$ の各連用修飾表現において，その $n_{E(c_j,SP(V))}$ が “Number” であった連用修飾表現 $E(c_j, SP(V))$ の重み $W(E(c_j, SP(V)))$ を以下の式で計算する．

$$Ws(E(c_j, SP(V))) \quad = \quad f('Number', c_j, V) \times \frac{f(V, c_j)}{f(V)} \tag{6}$$

$f('Number', c_j, V)$**:**　動詞 V における格助詞 c_j を伴う連用修飾表現で，格助詞 c_j の前に出現する名詞が “Number” である頻度,

$f(V)$**:**　動詞 V の全コーパスにおける出現頻度,

$f(V, c_j)$**:**　全コーパスにおいて，動詞 V が格助詞 c_j と共起した頻度,

Step 3　$Ws(E(c_j, SP(V)))$ がある閾値以下の連用修飾表現を省略可能と認定する．

2.6　文脈補正項について

式 (1) における第 3 項，$(1 + \sum_{n \in N(c_j,SP(V))} B(n, SP(V)))$ を，本手法の文脈補正項と定義する．$B(n, SP(V))$ は $tf \cdot idf$(Salton 1988) を改良した計算式である．従来の $tf \cdot idf$ は，名詞のある文書における重要度を算出する計算式であるが，本手法で提案する改良した計算式は，文書中における名詞の出現位置によって重要度が変化する点が従来の $tf \cdot idf$ と異なる．例えば，名詞 $n \in N(c_j, SP(V))$ が以前の文に出現している位置で $B(n, SP(V))$ を計算すると，$before(n, SP(V))$ の値が大きくなるので値は小さくなる．そのため，省略可能であるかどうかの認定対象となっている連用修飾表現に，それより以前の文に出現した名詞が含まれている場合は値が小さくなる．よって，そのような連用修飾表現は省略可能と認定されやすくなる．し

かし，以降の文に出現する名詞が含まれている場合は値が大きくなり，省略可能と認定されにくくなる．また，重要な名詞の多い，長い連用修飾表現には情報が多く，それを省略することで情報欠落が大きくなる危険がある．しかし，重要な名詞は一般にコーパスにおける頻度が小さく，$df(n)$ が小さい．よって，$B(n, SP(V))$ は $\log \frac{P}{df(n)}$ によって大きくなり，そのような名詞を多く含む連用修飾表現の文脈補正項 $(1 + \sum_{n \in N(c_j, SP(V))} B(n, SP(V)))$ は大きくなる．そのため，重要な名詞の多い，長い連用修飾表現は省略可能と認定されにくくなる．

3　手法の実装

本手法を実装して，文書の要約システムを作成した．コーパスは 1993 年の日本経済新聞記事 1 月 1 日から 3 月 31 日までの，32729 記事，278628 文を採用した．この中から含有節と，それに対する差異節を検索する．なお，名詞間類似度を求めるための情報も，この範囲内で抽出し獲得する．形態素解析器として JUMAN version 3.5 を，構文解析器として KNP version2.0b6 を採用した．

本手法によって省略可能と認定できる例をいくつか以下に示す．下線で示された部分が省略可能と認定された連用修飾表現である．

- 欧州共同体の市場統合が 十二月三十一日で 完成し、十二カ国、人口三億四千万人の世界最大の単一市場が一日発足する。
- ＥＣ域内の自由な経済活動を妨げてきた国境規制や基準の違いから生じる障壁を取り除こうというもので、そのために必要な措置として 二百八十二の指令・規則案を定めた。
- 通貨・政治統合は九二年夏のデンマーク国民投票のマーストリヒト条約批准否決以来揺れているが、ＥＣは市場統合の完成をステップに 実現を急ぐ考えだ。

実際には，本手法によって省略可能と認定された連用修飾表現を文から削除することによって，削除型の文内要約を実現することができる．

4　評価実験

4.1　実験方法および結果

実装したシステムを評価した．実験における対象記事は，1993 年の日本経済新聞記事 1 月 1 日から 3 月 31 日までの 32729 記事の中から，無作為に 11 記事を選択した．選択した 11 記事には全 183 文が存在し，この中から含有節は 196 節（1 文で 2 つ以上の含有節を有する文もある）存在した．つまり，この 196 節が認定対象となる．ただし，省略可能な連用修飾表現認定に必要となる差異節や，名詞間類似度を求めるための情報等は，日経新聞 93 年の 1 月 1 日から 3 月 31 日までの 32729 記事から取得した．196 個の含有節に対する差異節は合計 76314 個存在し，1 含有節あたりの平均差異節は 389 個であった．本手法によって含有節の 196 節から省略可能

表 2　評価実験結果

閾値	再現率 (%)	適合率 (%)	F値	省略認定数
0.04	55.5	79.5	65.3	83
0.05	59.7	80.7	68.6	88
0.06	61.3	79.3	69.2	92
0.07	67.2	80.8	73.4	99
0.08	69.7	79.0	74.1	105
0.09	71.4	78.7	74.9	108
0.1	71.4	77.3	74.2	110
0.11	71.4	76.6	73.9	111
0.12	73.1	76.3	74.7	114
0.13	73.1	75.7	74.4	115
0.14	73.1	74.4	73.7	117
平均	67.9	78.0	72.4	103.8

な連用修飾表現を認定した．評価方法は，対象記事群から省略可能な連用修飾表現の正解データを作成し，適合率，再現率で性能を評価する．ここで正解データは，対象記事群における全ての含有節から，省略しても妥当な連用修飾表現を人手で抽出し，作成した．再現率，適合率の定義を示す．

$$\text{再現率} = \frac{\text{本手法による結果と正解データで一致する省略可能な連用修飾表現の数}}{\text{正解データの省略可能な連用修飾表現の数}}$$

$$\text{適合率} = \frac{\text{本手法による結果と正解データで一致する省略可能な連用修飾表現の数}}{\text{本手法によって省略可能と判定された連用修飾表現の数}}$$

実験結果を表 2 に示す．表 2には本手法の $Ws(E(c_j, SP(V)))$ における閾値を 0.04 から 0.14 まで変化した場合の適合率，再現率，F値，省略認定数と，おのおのの平均値を示す．なお，$n_{E(c_j,SP(V))}$ が数値表現である場合は閾値を 10 倍した値を使用する．本手法を評価するための対象記事では，閾値が 0.09 のときが最もF値が高い結果となった．しかし，対象記事によっては多少，変化するものと考える．そこで，最大のF値のときの閾値 0.09 を中心に，閾値を 0.04 から 0.14 まで変化させたときの再現率，適合率，F値，省略認定数の平均を本手法の評価結果として採用する．よって，本実験によって再現率 67.9%，適合率 78.0%を得た．

4.2　文脈補正項を導入した場合と導入しない場合との比較

本手法では，含有節において認定対象の連用修飾表現の内容および前後の文脈を考慮した文脈補正項を導入している．これによって，省略可能であるかどうかの認定対象となっている連用修飾表現に，それより以前の文に出現した名詞が含まれている場合は省略可能と認定されやすくなる．また，認定対象となっている連用修飾表現に重要な名詞が多数含まれていれば，省略可能と認定されにくくなる．この文脈補正項が，どの程度，性能に影響しているのかを調べ

表 3　文脈補正項を導入しない場合の性能

閾値	再現率 (%)	適合率 (%)	F値	省略認定数
0.04	52.9	77.8	63.0	81
0.05	56.3	77.9	65.4	86
0.06	61.3	78.5	68.9	93
0.07	63.0	77.3	69.4	97
0.08	66.4	76.7	71.2	103
0.09	67.2	76.2	71.4	105
0.1	67.2	76.2	71.4	105
0.11	67.2	75.5	71.1	106
0.12	68.9	75.9	72.2	108
0.13	68.9	75.2	71.9	109
0.14	70.6	74.3	72.4	113
平均	64.6	76.5	69.9	100.5

る．そのために，計算式に文脈補正項を導入した場合としない場合との性能を比較する．表 3 に文脈補正項を導入しない場合の実験結果を示す．平均を採用した場合，再現率 64.6%，適合率 76.5%，F 値 69.9 を得た．

4.3　考察

評価の結果，提案手法は再現率 67.9%，適合率 78.0%の結果を得た．再現率 67.9%であることから，対象記事には，まだ省略できる連用修飾表現が残されているといえるので，手法には改良の余地がある．しかし，再現率よりも適合率のほうが実際に手法を適用する場合に重要であると考える．なぜなら，省略箇所の網羅性が少なくても読み手に悪影響を与えないが，間違った省略による要約は読み手に情報の欠落した情報を提供するからである．適合率は 78.0%であり，本手法によって認定された省略箇所は，概ね妥当であると考える．

提案手法で，閾値を 0.1 としたときに省略可能と認定された連用修飾表現を，表層格の種類によって分類した．表 4に，各表層格において，省略が妥当と判定された連用修飾表現の数，妥当ではないと判定された連用修飾表現の数を挙げる．また，対象記事とした日本経済新聞 1993 年の 1 月 1 日から 3 月 31 日までの記事において，含有節における各表層格の出現の割合を示す．省略が妥当と判定された連用修飾表現の中で最も数が多かったのは，含有節の動詞に対する格助詞を持たない連用修飾表現であった．本論文では，そのような連用修飾表現を，動詞に係っている品詞情報を利用して「名詞」「接続詞」「副詞」「助詞」「指示詞」「形容詞」「動詞」の 7 つに分類した．これらの出現割合の合計は約 26.4%であり，格要素以外にも省略可能な連用修飾表現が多数，存在していることが分かる．このような連用修飾表現には「比較的順調に」といった形容動詞が変化したもの,「そのために必要な措置として」といった動詞「する」が変化したものが見られた．これらの連用修飾表現の省略可能性の判定は格フレームでは対応できず,

表 4　各表層格の判定の傾向と含有節における出現傾向

表層格の種類	妥当であった数	妥当ではない数	正解データの数	出現割合 (%)
ニ	9	5	19	14.16
ヨリ	1	0	1	0.20
デ	6	1	13	7.32
ヘ	1	1	2	0.14
ト	4	0	5	3.73
カラ	8	1	9	2.51
マデ	4	3	4	0.67
名詞	4	4	9	7.82
接続詞	3	0	3	1.38
副詞	14	0	15	4.09
助詞	8	6	12	7.44
指示詞	1	1	1	0.31
形容詞	11	1	13	2.43
動詞	11	1	12	2.90

本手法の有効性を示すと考える．

しかし，省略が妥当ではないと判定された連用修飾表現の中で最も数が多かったのも，含有節の動詞に対する格助詞を持たない連用修飾表現であった．妥当ではないと判定されたものには,「イデオロギーや安全保障上の根本的な対立がない以上」のような，重要な情報を含む連用修飾表現があった．これらの連用修飾表現を省略することは，情報欠落が大きくなってしまう原因になる．本手法では，重要な名詞を多く含む連用修飾表現は省略可能と認定されにくくなるような補正項を導入している．しかし，それでも省略可能と認定されてしまう重要な情報を含む連用修飾表現が存在する．

このような格助詞を持たない連用修飾表現を省略可能とすることでシステムの性能が低下してしまうとすると，格助詞を持たない連用修飾表現に対して省略可能かどうかを認定することが有害になってしまう．そこで，表層格の種類ごとに適合率，再現率を算出し，どの連用修飾表現の省略が有効であったか調べる．表 5に結果を示す．表 5によると，格助詞を持たない連用修飾表現の中では,「動詞」「副詞」「形容詞」が比較的，頻度が大きいにもかかわらず，適合率，再現率が高い．頻度が大きい格要素の中では，ニ格での適合率，再現率の低さが目立つ．その理由は以下のとおりである．ニ格は動詞に対する目的格となるものが多く，妥当ではないと判定されたニ格は，目的格を省略したため情報の欠落が大きい連用修飾表現が多かった．また，場所を表すニ格の連用修飾表現もあり，それを省略可能と認定した場合も情報欠落が大きい．そのため，妥当ではないと判定された場合が多かった．逆に，妥当であると判定されたニ格の連用修飾表現には「今月中に」「秋口に」といった時期を表す句,「絶対に」といった程度を表す連用修飾表現が多く，このような連用修飾表現は省略しても情報欠落が少なく，妥当であると判定された．つまり，ニ格に関しては，それが目的格であるかどうかが判別できれば，目的格は

表 5　各表層格の適合率，再現率

表層格の種類	再現率 (%)	適合率 (%)
ニ	47.4	64.3
ヨリ	100.0	100.0
デ	46.2	85.7
ヘ	50.0	50.0
ト	80.0	100.0
カラ	88.9	88.9
マデ	100.0	57.1
名詞	44.4	50.0
接続詞	100.0	100.0
副詞	93.3	100.0
助詞	66.7	57.1
指示詞	100.0	50.0
形容詞	84.6	91.7
動詞	91.7	91.7

省略可能と認定しないという条件をつけて，適合率をより上げることができると考える．

文脈補正項については，実験結果より，対象の含有節において文脈補正項によって再現率，適合率が上がり，性能への効果があったことがわかる．文脈補正項によって新たに省略可能となった連用修飾表現の数は 11 個であり，この中で省略が妥当な連用修飾表現の数は 7 個であった．一方，文脈補正項で省略不可となったのは 6 個であり，この中で省略が妥当な連用修飾表現の数は 2 個であった．文脈補正項を導入したことによって，全体の認定数の中で省略が妥当ではない連用修飾表現の数を減らし，省略が妥当な連用修飾表現の数を増やすことができた．よって，この文脈補正項は妥当であると考える．文脈補正項によって省略不可になった連用修飾表現には,「新たな財投資金の調達多様化策として」,「各省庁の許認可や審査事務の迅速化への取り組みについて」といった，重要な情報が多く含まれている表現が多かった．これは，文脈補正項の構成要素 $B(n, SP(V)) = \frac{1+after(n,SP(V))}{2(1+before(n,SP(V)))} \times \log \frac{P}{df(n)}$ の $\log \frac{P}{df(n)}$ が，高い値をとったため，省略不可と認定されたと考える．例えば，連用修飾表現「新たな財投資金の調達多様化策として」の「財投資金」や「調達多様化策」といった複合名詞はコーパスにおける頻度が少なく，$df(n)$ が小さくなる．よって，このような名詞の $\log \frac{P}{df(n)}$ は高くなる．そのため，文脈補正項の値は高くなり，省略不可と認定されるようになった．実際,「新たな財投資金の調達多様化策として」の重み $Ws(E(c_j, SP(V)))$ は，文脈補正項導入前で 0.038 であったが，文脈補正項導入後で 0.212 となった．しかし，既に示したように，情報が多く含まれている連用修飾表現でも省略可能と認定される場合があるので，文脈補正項は改良の余地があると考える．

文脈については，該当の連用修飾表現に，以前の文に出現した名詞が含まれていた場合，重みが下がることで省略可能と認定されやすくなる．例えば,「また、単一市場の誕生で北欧や東欧を巻き込んだ欧州経済圏の結び付きが強まるのは確実で、世界の経済体制の行方にも影響し

そうだ」という文において，連用修飾表現「単一市場の誕生で」は，文脈から省略可能となる．それは，以前の文に「欧州共同体（ＥＣ）の市場統合が完成し、世界最大の単一市場が一日発足する」という内容の文が存在しているからである．よって，連用修飾表現「単一市場の誕生で」の文脈補正項を計算する場合,「単一市場」の $B(n, SP(V)) = \frac{1+after(n,SP(V))}{2(1+before(n,S(P)))} \times \log \frac{P}{df(n)}$ における $before(n, S(P))$ は，以前の文に「単一市場」が存在するため値が大きくなる．そのため，文脈補正項が小さくなり省略可能と認定されやすくなる．しかし，実際には「単一市場の誕生で」における重み $Ws(E(c_j, SP(V)))$ は，文脈補正項導入前で 0.0586 であったが，文脈補正項導入後で 0.0865 となった．これは,「単一市場」という複合名詞の $\log \frac{P}{df(n)}$ が高いためであると考える．文脈から省略可能な連用修飾表現に対しては，より重みが小さくなるように，文脈補正項を改良する必要があると考える．

5　格フレームによる手法との比較実験

5.1　実験方法および結果

本手法は，省略可能な連用修飾表現をコーパスの情報から認定する手法であるが，省略可能な連用修飾表現の認定は格フレーム辞書を使用することでも認定できる．すなわち，格フレームに記述されている格要素を動詞に対する必須格として省略不可能，記述されていない格要素を任意格として省略可能と認定する．

実験では，格フレーム辞書を用いることによって省略可能な格要素の認定を行ない，提案手法との比較実験を行なった．認定対象，および対象記事は上記の実験と同じである．格フレーム辞書には日本語語彙大系 (池原他 1997) の構文意味辞書を使用した．使用した格フレーム辞書には意味素性による意味制約が記載されている．一つの動詞に複数の格フレームが記載されている場合は，格フレーム辞書に記載されている意味素性との照合を行ない，省略可能な連用修飾表現の認定を行なう．以下に格フレームを使用した場合の認定手法を示す．なお，格フレームとの照合は人手で行なった．

Step 1　含有節 $SP(V)$ の動詞 V に対する格フレームを格フレーム辞書から得る．

Step 2　連用修飾表現 $E(c_j, SP(V))$ の格助詞 c_j と，その前に出現する名詞 $n_{E(c_j,SP(V))}$ の意味素性を日本語語彙大系 (池原他 1997) の単語体系を利用して取得する．複数，意味素性がある場合は全て採用する．

Step 3　まず，格フレームに記載されていない格助詞を含む連用修飾表現を省略可能と認定する．例えば，含有節「欧州共同体の市場統合が十二月三十一日で完成する」には，動詞「完成する」にガ格の「欧州共同体の市場統合が」とデ格の「十二月三十一日で」が係っているが,「完成する」の格フレームは表 6のとおりである．格フレームにはデ格が記載されていない．よって，デ格を省略可能な格要素と認定する．

表 6　動詞「完成する」の格フレーム

No.	格フレーム	意味素性
1	N1 ガ	N1=＊
2	N1 ガ N2 ヲ	N1=3 主体 N2=＊

表 7　動詞「導入する」の格フレーム

No.	格フレーム	意味素性
1	N1 ガ N2 ヲ N3 ニ/ヘ	N1=3 主体 N2=＊ N3=362 組織 388 場所 760 人工物 1001 抽象物 1236 人間活動 2054 事象

表 8　格フレームを用いた手法との比較

手法	省略可能と認定された数	再現率 (%)	適合率 (%)
提案手法	103.8	67.9	78.0
格フレームによる手法	150	77.3	61.3

Step 4　格フレームに記載されている格助詞を含む連用修飾表現は，格フレーム辞書に記載されている意味素性と，Step 2 で取得した名詞 $n_{E(c_j,SP(V))}$ の意味素性との照合を行なう．その際，Step 2 で取得した名詞 $n_{E(c_j,SP(V))}$ の意味素性か，その上位概念が格フレームに記載されていれば，その連用修飾表現は省略不可と認定する．しかし，記載されていない場合は，その連用修飾表現は省略可能と認定する．例えば，連用修飾表現「銀行にＥＣ単一免許が導入される」の動詞「導入する」の格フレームは表 7のとおりである．ニ格の連用修飾表現「銀行に」の名詞「銀行」の意味素性は「374 企業 428 仕事場」である．「374 企業」の上位概念に「362 組織」があり，二格の格フレームと照合を行なうと，格フレームに記載されている意味素性であることがわかる．よって，連用修飾表現「銀行に」は省略不可である．ガ格の連用修飾表現「ＥＣ単一免許が」の名詞「免許」の意味素性は「1735 許可 1166 権利」である．この上位概念に「3 主体」は含まれていない．しかし，ガ格とヲ格の連用修飾表現を省略可能とすることは情報欠落が大きいので，ガ格，ヲ格に「3 主体」が記載されている場合は，たとえ上位概念に「3 主体」が含まれていなくても省略不可と認定する．

ただし，格フレーム辞書によって省略可能性を判定できるのは，動詞に対する格要素のみである．しかし，実際には格要素以外の省略可能な連用修飾表現が多数存在する．そのため，格フレームを使用した手法では，格助詞を伴わない連用修飾表現は全て省略可能と認定した．なお，格フレーム辞書に記載されていない動詞の格要素は，どれも省略不可能とした．また，連用修飾表現が提題である場合も省略不可能と認定した．実験結果を表 8に示す．提案手法の再現率，適合率は節 4.1 で得た結果である．

表 9　格要素のみを対象とした場合の比較

手法	省略可能と認定された数	再現率 (%)	適合率 (%)
提案手法	45	61.8	74.5
格フレームによる手法	49	51.9	57.1

表 10　格フレームを用いた手法における各表層格の適合率，再現率

表層格の種類	再現率 (%)	適合率 (%)
ニ	42.1	50.0
ヨリ	0.0	0.0
デ	92.3	70.6
ヘ	0.0	0.0
ト	20.0	100.0
カラ	44.4	66.7
マデ	75.0	60.7
名詞	100.0	60.0
接続詞	100.0	100.0
副詞	100.0	100.0
助詞	100.0	36.4
指示詞	100.0	50.0
形容詞	100.0	86.7
動詞	100.0	64.7

5.2　格要素のみを対象とした場合の比較

節 5.1 の比較実験では，格助詞をともなわない連用修飾表現に対しては全て省略可能と認定した．しかし，全て省略可能としては適合率の低下が予想できる．そこで，格助詞をともなわない連用修飾表現に対しては全て省略不可とする場合と提案手法とを比較した．提案手法も，格助詞をともなわない連用修飾表現を全て省略不可とした．すなわち，格要素のみを対象に，提案手法と格フレームによる手法とを比較したことになる．この場合，正解データも格要素のみを対象として，適合率，再現率を算出した．実験結果を表 9に示す．なお，提案手法の結果は，この実験では閾値を変化させると 1.2 のとき F 値が最も大きくなったので，閾値を 0.07 から 0.17 まで変化したときの適合率，再現率の平均を採用した．

5.3　考察

表 8によると，格フレームを用いた手法は再現率が提案手法に比べて上がっているが，適合率は下がっている．これは，格助詞をともなわない連用修飾表現を全て省略可能であると認定しているためであると考える．このことを確かめるため，格フレームを用いた手法でも，表層格の種類ごとに適合率，再現率を算出し，どの連用修飾表現の省略が有効であったかを調べる．表 10に結果を示す．格助詞を伴わない連用修飾表現の省略で，それぞれの再現率が 100%なの

は，そのような連用修飾表現は全て省略可能としたからである．また，適合率は「名詞」以外はそれぞれ提案手法のほうが同等，もしくは良い適合率を出している．この結果から，格助詞を伴わない連用修飾表現の省略は有効であるが，全てを省略可能とすることは性能の低下を招くことが分かる．

格要素のみを対象とした場合においても本手法が格フレームを用いた手法に比べて，再現率，適合率ともに上回った．この結果によって，格要素のみを対象とした場合でも従来の格フレームを用いた手法より本手法が優れていることが分かる．

格フレームを用いた手法において，再現率に関しては，含有節の動詞が格フレームに記述されていなかった場合，その動詞に係る格要素を全て省略不可としたので低下したと考える．なお，全 196 の含有節において，格フレームに記述されていなかった動詞は 41 個存在した．この 41 個の動詞に係る格要素は省略可能かどうかの判別ができないので，全て省略不可とするしかないが，これらの中には省略可能な格要素も含まれている．よって再現率が低下したと考える．採用した格フレーム辞書は日本語語彙大系の構文意味辞書であり，現在，一般に使用できる格フレーム辞書の中でも最大規模である．しかし，実に約 20%もの動詞が記述されていなかった．本手法は格フレームに記述されていないような動詞に係る格要素でも省略可能かどうかの認定ができるので，格フレームを用いた手法より優れた手法であると考える．

格フレームに記述されていない動詞の格要素は全て省略不可としたので，適合率に関しては格フレームの網羅量の影響を受けない．しかし，適合率に関しても格フレームを用いた手法より，本手法のほうがよい結果であった．これは，格フレームの質が適合率の結果に影響を与えるためである．つまり，ある動詞に関して，格フレームに記載されていない格要素を省略することで情報欠落が大きい場合があったので適合率が低下した．例えば,「当面は最大手の日住金への金利減免支援の強化を最優先する.」という文において，動詞「優先する」の必須格はガ格，ニ格と格フレームに記述してある．しかし，この文におけるヲ格「最大手の日住金への金利減免支援の強化を」を省略することは妥当ではない．他にも「大蔵省・日銀は住専支援策について、すでに一部関係金融機関に非公式に打診を始めた」という文において，動詞「始める」の必須格はガ格，ヲ格，カラ格と格フレームに記述してある．そのため，ヲ格「打診を」は省略不可であるが，ニ格「一部関係金融機関に」は省略可能となる．しかし，それは「どこに」に相当する部分を省略したことになり，省略が妥当とはいえない．この結果は，文は多様に変化し，格フレームのような静的な情報では対応に限界があることを示していると考える．

6　結び

本論文では，省略できる可能性のある連用修飾表現を含む節に対して，同一の動詞をもち，かつ，格助詞出現の差異が認められる節をコーパスから検索し，検索された節対から省略可能な連用修飾表現を認定する手法を提案した．省略可能な連用修飾表現を認定する手法として，

格フレームを用いる手法が考えられるが，格フレームでは格助詞を持たない連用修飾表現に対しては，省略可能かどうかの判定ができない．提案した手法は，そのような欠点を克服し，格助詞を持たない連用修飾表現でも省略可能かどうかの認定ができる．また，連用修飾表現の内容および前後の文脈を考慮して，重要な情報が多く含まれている連用修飾表現に対しては省略可能と認定できる可能性を低く，逆に，認定対象としている連用修飾表現に，それより以前の文に存在する情報が含まれている場合に対しては，省略可能と認定できる可能性が高くなるような工夫を施した．これにより，格フレームのような静的な情報ではなく，動的な情報で省略可能な連用修飾表現を認定できる．

評価実験によって，本手法による省略可能な連用修飾表現は再現率 67.9%，適合率 78.0%を示し，比較的，良好な結果であった．これは，格フレームを用いた手法より高い値であった．さらに，格要素のみを対象とした場合も本手法のほうが良い結果であった．これは，本手法によって，省略可能な格助詞を持たない連用修飾表現を高い精度で抽出できたからであった．加えて，格フレームを用いた手法では，多様に変化する文に対して，格フレームのような静的な情報では対処しきれず，重要な情報を含む連用修飾表現を省略可能と認定した場合が多かった．これらの理由によって，本手法は格フレームを用いた手法を上回る性能を示すことができたと考える．

しかし，本手法では，ニ格に関しては，他の格要素と比べて性能がよくなかった．ニ格は動詞に対して目的格となるものが多く，そのようなニ格の格要素は，多くが省略不可であった．しかし，本手法ではニ格の中で目的格であるものと，そうでないものとの判別ができず，目的格であっても省略可能と認定してしまうことがある．よって適合率が下がったと考える．本手法の適合率を上げるには，ニ格の格要素を目的格であるか，そうでないかを判別し，目的格である場合には省略不可とするような制限を加えることで，適合率を上げることができると考える．

謝辞

言語データとして，日本経済新聞 CD-ROM 版の使用を許可して頂いた日本経済新聞社に深謝する．また，日本語語彙大系から意味分類を取得するために用いた形態素解析システム ALT-JAWS ver.2.0. の使用を許可して頂いた日本電信電話 (株) に深謝する．

参考文献

Hindle, D. (1990). “Noun Classification From Predicate Argument Structures.” In *Proceedings of the 28th Annual Meeting of ACL*, pp. 1268–1275.

平岡冠二, 松本裕治 (1994). “コーパスからの動詞の格フレーム獲得と名詞のクラスタリング.” 情報処理学会研究報告, 94-NL-104, pp. 79–86.

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. 岩波書店.

加藤直人, 浦谷則好 (1999). “局所的要約知識の自動獲得手法.” 自然言語処理, **6** (7), pp. 73–92.

河原大輔, 黒橋禎夫 (2002). “用言の直前の格要素の組を単位とする格フレームの自動獲得.” 自然言語処理, **9** (1), pp. 3–19.

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** (5), pp. 1–25.

Salton, G. (1988). *Automatic Text Processing*. Addison Wesley.

宇津呂武仁, 松本裕治, 長尾眞 (1993). “二言語対訳コーパスからの動詞の格フレーム獲得.” 情報処理学会論文誌, **34** (5), pp. 913–924.

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, 97-NL-122, pp. 83–89.

山本和英, 増山繁, 内藤昭三 (1995). “文章内構造を複合的に利用した論説文要約システム GREEN.” 自然言語処理, **2** (1), pp. 39–55.

山崎邦子, 三上真, 増山繁, 中川聖一 (1998). “聴覚障害者用字幕生成のための言い換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.

情報処理振興事業協会技術センター（IPA） (1987). 計算機用日本語基本動詞辞書 IPAL （Basic Verbs） –解説編–. 情報処理振興事業協会.

略歴

酒井 浩之: 2002 年 豊橋技術科学大学大学院修士課程知識情報工学専攻修了. 現在，同大学院博士後期課程電子・情報工学専攻在学中. 自然言語処理，特に，検索，要約の研究に従事. e-mail: sakai@smlab.tutkie.tut.ac.jp

篠原 直嗣: 2001 年 豊橋技術科学大学大学院修士課程修了. 現在，(株) リコー勤務. 在学中は，自然言語処理，特に，テキスト自動要約の研究に従事.

増山 繁: 1977 年 京都大学工学部数理工学科卒業. 1982 年 同大学院博士後期課程単位取得退学. 1983 年 同修了 (工学博士). 1982 年 日本学術振興会奨励研究員. 1984 年 京都大学工学部数理工学科助手. 1989 年 豊橋技術科学大学知識情報工学系講師，1990 年 同助教授. 1997 年 同教授. アルゴリズム工学，特に，並列アルゴリズム等，及び，自然言語処理，特に，テキスト自動要約等の研究に従事. 言語処理学会，電子情報通信学会，情報処理学会等会員. e-mail: masuyama@tutkie.tut.ac.jp

山本 和英: 1996 年豊橋技術科学大学大学院博士後期課程システム情報工学専攻修了. 博士 (工学). 同年より (株) 国際電気通信基礎技術研究所 (ATR) に所属し, 現在は ATR 音声言語コミュニケーション研究所, 研究員. 1998 年 中国科学院自動化研究所, 国外訪問学者. 換言処理, 機械翻訳, 要約処理, 中国語及び韓国語処理の研究に従事. 1995 年 NLPRS'95 Best Paper Awards. 言語処理学会, 情報処理学会, ACL 各会員. e-mail: yamamoto@fw.ipsj.or.jp

(2001 年 11 月 6 日 受付)

(2002 年 2 月 8 日 再受付)
(2002 年 4 月 10 日 採録)