

## 抽出パターンの階層的照合に基づく日本語テキストからの内容抽出法

松尾 比呂志<sup>†</sup> 木本 晴夫<sup>†</sup>

テキストから情報を抽出する処理として、辞書的情報に頼らない未知語に強い内容抽出処理を提案する。具体的には、任意の文字列と一致するワイルドカード項を含む複数種類の項から構成される抽出パターン（文脈自由型文法に属する）と入力文字列との階層的な照合を行うことによって抽出する。ワイルドカード項に対応する文字列は、前後の項の制約で決まる。これを利用し、未知語が予想される部分に対してはワイルドカード項で表現する。製品紹介記事約 300 文を訓練データとして用い、抽出パターンの作成および改良を行い、約 98% の抽出成功率を得た。この抽出パターンを用いて、非訓練データ 109 文に対し抽出実験を行った結果、約 86% の抽出成功率が得られた。256 単語相当の辞書的情報と 178 個のパターンで、このような高い抽出成功率が得られ、本方法が未知語に強く実用的にも有効な方法であることを明らかにした。

### A Content Extraction Method from Japanese Texts Based on Pattern Matching Using Extraction Patterns

HIROSHI MATSUO<sup>†</sup> and HARUO KIMOTO<sup>†</sup>

We propose a new information-extraction method that is capable of handling unknown words and does not rely on lexical information. The system using this method extracts contents of sentences by hierarchical matching between each sentence and extraction patterns. Extraction patterns take the form of context-free rules where the symbols include wild-cards. The strings matching the wild-cards are determined by the previous and following terms. Therefore, the part that may include unknown words can be represented by a wild-card. Extraction patterns are made and modified based on test data of about 300 sentences taken from news which introduce new products. In test data, the experiment shows 98% of extraction accuracy. In raw data of 109 sentences, the experiment using the same patterns shows 86% of extraction accuracy. High extraction accuracy is achieved using 178 patterns and a minimum amount of lexical information which corresponds to 256 words. The result shows that the method is suitable for handling unknown words and therefore practically effective.

#### 1. はじめに

テキスト中から特定の内容を抽出する処理は、データベース自動構築や要約生成など応用範囲の広い処理であり、文書の構造解析を用いた方法<sup>1)</sup>、見出し情報を利用する方法<sup>2)</sup>、事象解析による方法<sup>3)</sup>などが研究されている。また、MUC<sup>4)</sup>において、内容抽出処理（ここでは、キーワード抽出等と区別するため「情報抽出」の代わりに「内容抽出」という言葉を用いる）のコンテストも実施されている。

これらいずれの内容抽出処理も、テキスト中の語のほとんどが辞書に登録されていることを前提としてい

る。一般に、形態素解析および構文解析を行って、解析結果の格構造を参照しながら、抽出条件を指定したフレームとの照合を行って抽出する。これらの各処理においては、単語の辞書情報（品詞や意味カテゴリ）を利用するため、辞書中になく単語（未知語）がある場合には、解析の失敗で、抽出に失敗する可能性がある。形態素解析や構文解析で正しい解析結果を得るために、単語辞書、格構造パターンなどを充実させておく必要がある。

例えば、製品紹介記事の場合、新製品の製品名はほとんど未知語であり、未知語を含むことを前提とした処理が必要となる。構文解析処理等では、未知語を考慮した処理の研究がなされている<sup>5)</sup>。内容抽出を目的とする場合には、未知語に対しても、「発売対象の製品名」を意味するというような意味的解釈を与える必要

<sup>†</sup> NTT 情報通信研究所

NTT Information and Communication Systems Laboratories

があり、構文解析等の未知語の処理をそのまま使用することはできない。

そこで、本論文では、定型的内容が書かれている文章を対象として、構文上のキーとなる語（助詞、述語等）を手がかりに未知語の意味的解釈を与えることが可能な内容抽出法を提案する。この有効性を確認するため、できる限り辞書的情報を持たせずに、高精度に抽出を行うことを試みる。

具体的には、任意の文字列と一致するワイルドカード項を含む幾種類かの項の組み合わせで記述する抽出パターンと入力文字列との照合を行うことによって抽出する。ワイルドカード項では、前後の項の制約により抽出が行われ、ワイルドカード項に対応する文字列は、未知語を含んでよい。提案する方法は、多数の未知語を含むという前提の基で、パターンに埋め込む語（登録語）を少数の語に限定して、ワイルドカード項を積極的に利用する。

パターン照合に基づく情報抽出処理として、有限オートマトンを用いた FASTUS<sup>9)</sup>があるが、このシステムでは、辞書的情報を用いて、句を同定したあと、その結果を入力としてパターン照合を行っている。これ

に対し、提案する方法は、文字列そのものを入力とするパターン照合であり、少数の既知の語のみをパターン中に記述し、いわゆる辞書は用いていない。

## 2. 内容抽出処理方法

### 2.1 抽出パターン記述言語 EPL

一般に、自然言語処理では、単語辞書を用いて形態素解析を行い、形態素情報を頼りに、CFG<sup>7)</sup>等の文法を用いて構文的な解析を行っており、未知語は、例外的なものとして扱われているのが現状である。

一方、対話システムとして有名な ELIZA<sup>8)</sup>では、〈ワイルドカード〉によるパターン照合を利用している。また、文字列サーチ等で用いられる正規パターンも、〈ワイルドカード〉という概念を利用する。〈ワイルドカード〉は、緩やかな制約での照合が可能であるが、これまでのパターンでは、再帰的なパターンを記述できないなど、表現できるパターンに限界がある。

そこで、〈ワイルドカード〉を使えて、しかも CFG に属するパターンの記述が可能な抽出パターン記述言語 EPL (Extraction Pattern Language) を開発した。

表 1 に、EPL の仕様を示す。EPL では、5 種類の項

表 1 抽出パターンの仕様  
Table 1 Specification extraction pattern.

<p>〈抽出パターン〉 : 〈定義文〉の集合</p>
<p>〈定義文〉 : (〈S 定義文〉 〈W 定義文〉)</p>
<p>〈S 定義文〉 :                  〈サブパターン項〉あるいは〈開始記号〉に対し、〈規定パターン〉および〈例外パターン〉を定義する。〈例外パターン〉は、省略可能。〈規定パターン〉および〈例外パターン〉は、〈パターン記述〉で表現する。</p>
<p>〈W 定義文〉 :                  〈ワイルドカード項〉に対し、〈詳細パターン〉および〈例外パターン〉を定義する。いずれか一方は、省略可能。〈詳細パターン〉および〈例外パターン〉は、〈パターン記述〉で表現する。</p>
<p>〈パターン記述〉 := 〈項〉〈項〉…〈項〉</p>
<p>〈項〉 := ( 〈ワイルドカード項〉   〈サブパターン項〉   〈一般項〉   〈選言項〉   〈省略項〉 )</p>
<p>〈ワイルドカード項〉 :                  [一致条件] 例外を除き、任意の文字列と一致する。(本文参照)                  [書式] @〈文字列〉</p>
<p>〈サブパターン項〉 :                  [一致条件] 〈規定パターン〉と一致し、〈例外パターン〉と一致しない。                  [書式] %〈文字列〉</p>
<p>〈一般項〉 :                  [一致条件] 指定した文字列そのものと一致する。                  [書式] &lt;文字列&gt;</p>
<p>〈選言項〉 :                  [一致条件] 〈選言子〉のいずれかと一致する。                  [書式] (〈選言子〉 … 〈選言子〉)                  〈選言子〉は、〈パターン記述〉で表現する。</p>
<p>〈省略項〉 :                  [一致条件] 〈パターン記述〉と一致するか、空文字と一致する。                  [書式] [〈パターン記述〉]</p>

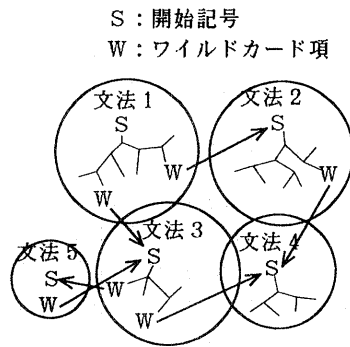


図1 複数の文法のワイルドカード項による結合  
Fig.1 Combined grammars by wild-card terms.

を使用できるが、ここでは、特に、ワイルドカード項について説明を行い、他の項について（表1と2.2節を参照してもらいたい）は、詳細な説明は省略する。ワイルドカード項は、抽出処理において以下の2つの重要な役割を果たす。

1) 終端記号列としての役割

ワイルドカード項は、以下の2つの条件で制約される以外は、任意の文字列と一致可能であり、前後のパターンの制約で対応する文字列が決まる。これにより、未知語を含んでいても抽出可能である。

①括弧対を分断しない。例えば、ワイルドカード項に対応する文字列として、『ab「c」』が抽出されることはなく、必ず、括弧対の関係は保たれる。

②例外パターンが定義されている場合には、例外パターンと一致しない。

2) 開始記号としての役割

詳細パターンが定義されたワイルドカード項の場合には、ワイルドカード項に対応する文字列が決定されると、その文字列を新たな入力文字列として、詳細パターンとの照合処理を行う。これは、抽出パターン全体を、図1のように、<ワイルドカード項>を媒介として結合された複数の文法で構成できることを意味する。これにより、階層的に照合を行え、照合対象の組み合わせ数が膨らむのを防ぐことができる。

2.2 抽出パターン記述例

図2に、抽出パターン記述例を示す。最初の定義文は、開始記号<文>の規定パターンを記述しており、10個の項の連系である。1番目の項<@販売元>と3番目の項<@会社情報>は、ワイルドカード項であり、任意の文字列が一致し得る。2, 4, 5, 8, 10番目の項は、一般項であり、各々、“(”, “)”, “は”, “を”, “。”の文字と一致することが要求される。6番目の項<%月日>, 7番目の項<%製品表現>, 9番目の項<%売

文	= <@販売元> <@会社情報> は<%月日> <%製品表現>を<%売り>。
@会社情報	= <社長><@社長名>氏
%製品表現	= <@製品>「@製品名」[ <@価格> ]
%価格	= <@値段>円
%売り	= (<販売> <発売>)<した>
%月日	= <%暦月><%日にち>

図2 抽出パターン記述例  
Fig.2 An example of extraction patterns.

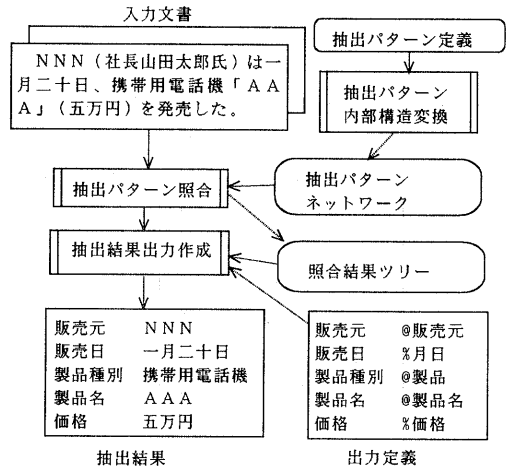


図3 内容抽出処理のフロー  
Fig.3 An information extraction process.

り>は、サブパターン項であり、各々、3番目から6番目の定義文で規定パターンが記述され、その規定パターンと一致することが要求される。

2番目の定義文では、ワイルドカード項<@会社情報>の詳細パターンを記述しており、<文>との照合で決定された<@会社情報>に対応する文字列が、このパターンと照合される。

例えば、入力文『NNN(社長山田太郎氏)は一月二十日、携帯電話機「AAA」(五万円)を発売した。』は開始記号<文>と照合成功する。このとき、<@会社情報>に対して、文字列『社長山田太郎氏』が得られるが、<@会社情報>には、2番目の定義文で<詳細パターン>が定義されており、これと照合が行われる。その結果、<@社長名>として、『山田太郎』が得られる。なお、ここで、サブパターン<%月日>の詳細な定義は省略しているが、「四月一日」のような「X月Y日」を表すパターンである。

2.3 抽出処理

抽出処理は、図3に示すように、抽出パターン内部構造変換、抽出パターン照合、抽出結果出力作成から

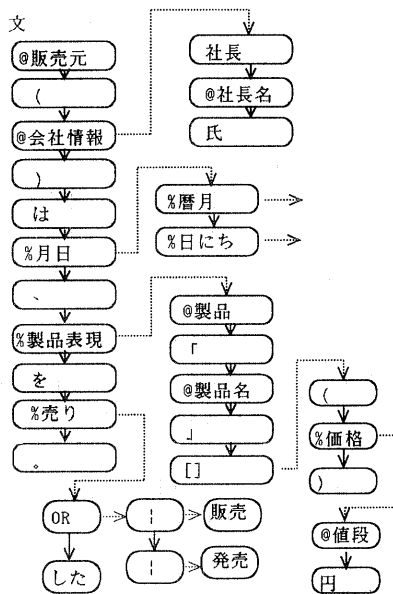


図4 抽出パターンネットワーク  
Fig. 4 An extraction pattern network.

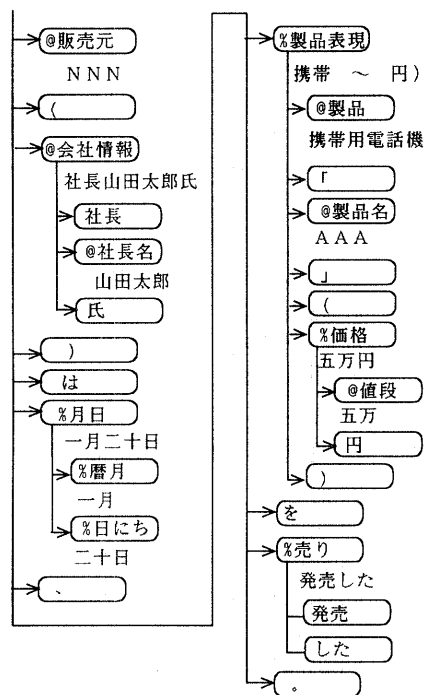


図5 照合結果ツリー  
Fig. 5 A matching result tree.

なる。

(1) 抽出パターン内部構造変換

抽出パターンを定義したファイル (図2) を読み込み、各項をノードとして連糸関係や選言関係等で接続された図4のようなネットワーク構造を作成する。

(2) 抽出パターン照合

入力文を読み込み、抽出パターンの<文>を開始記号として、トップダウン的に処理を行い、図5のような照合結果ツリーを作成する。

アルゴリズムの詳細な説明は省略するが、パターンを先頭の項と残りのパターンにわけ、各項の種類に応じた照合と残りのパターンの再帰的な照合処理により、実現している。

ワイルドカード項、省略項、選言項がある場合には、複数解が存在し得るが、ここでは、多義は考慮せず、最初に照合成功した解のみを出力としており、以下の照合順序を意識して抽出パターンを記述する。

ワイルドカード項：入力文字列を現在位置から右へ順次シフトさせながら、次の項以降のパターンの照合を行う。省略項：省略しない場合を先に照合し、失敗した場合には、省略項のパターンをとばして、次の項以降の照合を行う。選言項：左に記述された選言子から順に照合する。

(3) 抽出結果出力作成

出力定義で指定された項を、照合結果ツリーから探し、その内容を出力する。図2の場合には、<販売元>

に対してワイルドカード項<@販売元>、<発売日>に対してサブパターン項<%月日>、のように定義されており、<販売元>として「NNN」、<販売日>として「一月二十日」が出力される。

3. 製品紹介記事を対象とした内容抽出

製品紹介記事は、1)ほとんどの記事において、販売元、製品名、製品種別、販売日、価格の情報が記述されており、大部分は、本文の1文目に記述されている、2)複数製品を1文で記述することも多い、3)製品の特徴が連体修飾で記述され、複雑な埋め込みも多い、という特徴を持つ。

本論文では、ワイルドカード項を含む抽出パターンによる内容抽出処理の有効性を検証するため、製品紹介記事の本文の1文目を処理の対象とする。抽出する項目としては、販売対象製品の<販売元>、<製品特徴>、<製品種別>、<製品名>、<発売日>の5項目を抽出することを目標とする。これらの項目の内容を抽出するための考え方を以下で述べる。

3.1 抽出項目と抽出パターン

(1) 基本パターン

製品紹介文では、「~は~を発売する[と発表した]」(P1)と「~が発売するのは~」(P2)の2つの基本パ

ターンを考え、以下のように記述する。

P0: 文 = (<%文1> | <%文2>)

P1: %文1 = [<@主部>は[,]] (<%発表日表現> <@販売文> <%発表表現> | <%販売文>)

P2: %文2 = <@主部>が[ <%販売日表現> <%売り>]の<@製品表現>

ワイルドカード項<@販売文>の詳細パターンとサブパターン項<%販売文>の規定パターンは、同じパターンであり、「～を発売する」に相当する。<%発表表現> (「と発表した」等) を述語とする文の場合に、ワイルドカード項を用いたのは、<%発表表現>の述語の存在を先に判定し、無駄な照合を削減するためである。

#### (2) 販売元の抽出

<販売元>は、<@主部>の部分に存在し、一般には固有名詞である会社名が記述されている。すべての固有名詞を登録するのは困難である。たとえ、現在知られているすべての固有名詞を登録したとしても、新設の会社名などは未知語となる。このため、固有名詞に関しては未知語の可能性を前提としなければならない。そこで、固有名詞に相当する部分をワイルドカード項で記述する。

<販売元>に関しては、さらに以下の点を考慮する。

①<@主部>の部分には、<販売元>である<会社名>の他に、その会社に関する情報が括弧 ( ) でくくられて記述されることがある。例えば、<社長名>、<電話番号>、<住所>等が記述される。

②販売元が1社でなく複数社の場合もあり、「A, B など2社」のように、会社名そのものが主語とならず「2社」のような表現が主語となる場合もある。

#### (3) 販売日の抽出

<販売日>の表現は、他の項目に比べ、様々な位置に挿入されるが、表現を限定しやすい。そこで、<販売日>のパターンは、ワイルドカード項を用いずに、制約の強いパターンで記述する。<販売日>の表現のほとんどは、「X月Y日」といった具体的な月日で表現したものがほとんどであるが、その他に、年を表す表現や「上旬」のような日にちの範囲を表す表現等を考慮した。

#### (4) 製品表現の抽出

「～を販売する」の目的語 (<製品表現>と呼ぶ) には、<製品特徴>、<製品種別>、<製品名>が含まれる。これらには、様々な語が含まれ、これらの語をすべて登録するのは不可能である。これらの項目を抽出するために、ワイルドカード項を使用するとともに、

以下の表現上の性質を制約として利用する。

①<製品特徴>は、連体修飾句として表れる。

②連体修飾句の語尾は、特定の文字列 (「した」、「の」など) で制約できる。

③<製品名>は「」で囲まれて表現される。

④<製品種別>は、<製品特徴>、<製品名>以外の部分になる。

⑤一般に、<製品特徴>で修飾された複合名詞が、<製品種別>となる。

### 3.2 並列表現に対する対処

複数の製品が1つの文で紹介される場合も多く、並列表現を考慮する必要がある。例えば、2つの製品を紹介する場合、

P4: <@製品表現>と<@製品表現>を発売した。のように、2つの[製品表現]が「と」のような並列接続表現で結ばれる。

並列接続表現として、「と」、「、」、「、」を考える。「と」、「、」は、主文の述語の格に対する並列表現に用いられ、修飾句中で用いられることはまれである。したがって、<製品表現>を「<@製品表現>[と、<@製品表現>]」のように、2つのワイルドカード項を「と」、「、」で結んでも不都合はない。それに対して、「と」、「、」は、様々な表現で用いられるため、2つのワイルドカードを「と」や「、」で結んだパターンでは、不都合を生じる。例えば、

P5: <@製品表現>[と<@製品表現>]

は、1つの製品を紹介している場合と、「と」で結ばれた2つの製品を紹介している場合の2通りでマッチすることを想定している。構文的には、<@製品表現>は名詞句であることを想定しており、これらには製品の特徴を記述した表現が含まれる。しかし、

「重さ百五十gと業界最軽量の携帯用電話機」

は、1つの製品を紹介しているにもかかわらず、以下のように2つの製品として抽出されてしまう。

<@製品表現>: 重さ百五十g

<@製品表現>: 業界最軽量の携帯用電話機

並列表現の同定法には、様々な方法<sup>9)~14)</sup>が報告されているようにむずかしい問題である。単語の意味を基に行う方法では、未知語を前提とする場合には使えない。表層上の特徴による解析は未知語を前提とする処理でも使えるが、並列要素候補間の比較は、2章で述べた抽出パターンでは実現できず、並列表現のための新たな付加的な処理を必要とする。

そこで、付加的な処理を必要とせず、抽出パターンのみで対応できる方法として、並列要素の表層の特徴(3.1節(4))を利用した以下の方法を用いた。

- ①<製品名>の後に来る「と」や「、」の前までを並列要素となり得る1つの要素とする。
- ②<製品名>が記述されていない場合には、<製品特徴>の後に来る「と」や「、」の前までを並列要素となり得る1つの要素とする。先の例では、「重さ百五十gと業界最軽量の」が、<製品特徴>の表現として解釈され、1つの要素しかないと判断される。
- ③並列接続表現以外の「と」と結合する語(例:「接続できる」)を例外パターンに記述し、並列接続表現でない「と」が並列接続表現と解釈されるのを防止する。

#### 4. 内容抽出実験

電話機に関する製品紹介記事を対象に、本方法の有効性を確認する抽出実験を行った。

##### 4.1 実験対象記事の収集

工業関係の新聞の記事データベースから、キーワード「デンワキ」で検索した記事で、部品関係や開発中の製品の記事を除いた。さらに、新聞記事本文の1文目を抽出対象として取り出し、抽出パターン作成に用いる訓練データを約300文、抽出能力の評価に用いる非訓練データを109文用意した。

##### 4.2 評価尺度

<販売元>、<製品特徴>、<製品種別>、<製品名>、<価格>、<販売日>のすべての項目において、実験対象の文に記述された内容が過不足無く抽出された場合、その文を抽出成功(対象文に記述されていない項目に関しては、抽出されない場合を成功とする)とし、以下の式で表される抽出成功率を評価尺度とした。

$$\text{抽出成功率} = \frac{\text{抽出成功した文数}}{\text{実験対象の文数}}$$

##### 4.3 抽出パターン作成

訓練データを用いて、抽出パターンを作成し、抽出

に失敗した記事があれば、追加修正を行った。原則として、すべての記事が抽出成功するよう試みたが、この方法の原理的境界と思われる場合は、失敗のまま残した。このようにして、訓練データに対して、約98%の抽出成功率が得られる抽出パターンを作成した。

作成した抽出パターンは、パターン定義文の個数が178個、項の個数が約1700個、そのうち、単語(一般項をカウントし、読点等の記号も含む。)の異なり数は、265個であった。このように、きわめて少ない辞書的情報(256個の単語相当)で抽出パターンを記述できる。また、構文的な関係は他の分野にも共通に適用でき、用言とその格との関係において、各々の格が、抽出すべき内容のどの項目に相当するかが、対象分野によって異なるだけである。したがって、定型的内容を持つ記事であれば、その内容を表現する表現形態を整理すれば、比較的容易に他の分野へ適用できるであろう。

##### 4.4 実験結果

4.3節で作成した抽出パターンを用いて、非訓練データ109件に対して抽出実験を行った。その結果、約86%の抽出成功率が得られた。図6に示す成功例のように、複数の製品が記述され、さらに特徴が連体修飾の形で埋め込まれている文も正しく抽出できている。

失敗した文は、109文中15文である。失敗した文の詳細は、以下の通りである。1文は、「発売」を意味する述語がなく、「<製品>を発表した」のパターンであったため、いずれの項目も抽出されなかった。4文は、<販売元>のみ抽出されたが、他の項目は、抽出されなかった。これは、「発売」を意味する語は述語として用いられているが、抽出パターンにない文が挿入されていたり、「輸入販売する」のような登録してない複合動詞が用いられたのが原因である。3文は、<製品種別>のみの誤りであり、抽出された文字列には、<製品種別>を含んでいるが、余分の文字列をも含んでいた

入力文

NNNは十八日、同社のAAA(BBB通信網)サービス「CCC」に対応した、デジタル電話機「DDD」とG4デジタルファクシミリ「EEE=写真」、「FFF」のAAA端末三種を十九日からGGG地域で販売を始めると発表した。

抽出結果

販売元	NNN
製品特徴	同社のAAA(BBB通信網)サービス「CCC」に対応した
製品種別	デジタル電話機
製品名	DDD
製品種別	G4デジタルファクシミリ
製品名	EEE
製品名	FFF
製品種別	AAA端末
販売日	十九日

図6 内容抽出成功例

Fig. 6 An example of successful information extraction.

ものである。例えば、「<製品種別>として」がその例であり、パターンを追加登録すれば、解決できる。残りの失敗は、種々雑多であるが、本方法では困難なものとしては、1文で複数の製品を紹介した文で、しかも、複雑な連体修飾で表現された<製品特徴>が各製品に記述された文で、その結果、製品と製品との区切りの判別が困難となっている文である。

以上の実験では、本文の1文目のみを対象としているとはいえ、本論文で提案した方法を用いれば、比較的複雑な構文をもち、未知語も多く含んでいる文であっても、高い成功率で抽出が可能であると言える。

## 5. む す び

ワイルドカード項、省略項、サブパターン項、選言項で記述される抽出パターンを用いて、定型的な内容が記述された記事から、未知語を許容して情報を抽出する内容抽出法を提案した。電話機に関する製品紹介記事を用いて抽出実験を行い、訓練データに対して約98%、非訓練データに対して約86%の抽出成功率が得られ、提案した内容抽出法の有効性を確認した。

今後は、本文の1文目だけでなく、2文目以降や見出し文に対しても適用し、どの製品がどの特徴にあたるかなどの照応処理等の文脈処理を加えた、記事全体からの抽出法の実現が、課題となる。

## 参 考 文 献

- 1) 小松英二, 加藤安彦, 安原 宏, 椎野 努: 要約支援システム COGITO (文書の構造解析), 情報処理学会研究会資料, NL 64-11 (1987).
- 2) 高松 忍, 西田富士夫: 見出し情報を用いたテキスト解析と情報抽出, 情報処理学会論文誌, Vol. 29, No. 8, pp. 760-769 (1988).
- 3) 稲垣博人: 事象解析による要約情報の抽出, 情報処理学会研究会資料, NL84-3 (1991).
- 4) *Proc. Fourth Message Understanding Conference (MUC-4)*, Defence Advanced Research Projects Agency (1992).
- 5) 元吉文男, 大場健司, 石崎 俊, 井佐原均, 横山昌一: 未定義語を含む文の多段階構文解析法, 信学論 (D-2), Vol. J72-D-2, No. 10, pp. 1672-1679 (1989).
- 6) Hobbs, J. R., Appelt, D., Tyson, M., Bear, J. and Israel, D.: SRI International: Description of the FASTUS System Used for MUC-4, *Proc. Fourth Message Understanding Conf.*, pp. 268-275 (1992).
- 7) 嵩 忠雄, 都倉信樹, 谷口健一: 形式言語理論, p. 290, 電子情報通信学会 (1988).
- 8) Weizenbaum, J.: ELIZA-A Computer Program for the Study of Natural Language Communications between Man and Machine, *Comm. ACM*, Vol. 9, pp. 36-45 (1966).
- 9) 長尾 真, 辻井潤一, 田中伸佳, 石川雅彦: 科学技術論文における並列句とその解析, 情報処理学会研究会資料, NL 36-4 (1983).
- 10) 黒橋禎夫, 長尾 真: 長い日本語文における並列構造の推定, 情報処理学会研究会資料, NL 86-2 (1991).
- 11) 田村直良, 田中穂積: 意味解析に基づく並列名詞句の構造解析, 情報処理学会研究会資料, NL 59-2 (1987).
- 12) 稲垣博人, 壁谷喜義, 小橋史彦: 意味連結パターンを用いた係り受け解析, 情報処理学会研究会資料, NL 67-5 (1988).
- 13) 押金章悟: パターンを用いた日本語の名詞並列句解析, 第36回情報処理学会全国大会論文集, 3 T-2 (1988).
- 14) 首藤公昭, 吉村賢治, 津田建蔵: 日本語技術文における並列構造, 情報処理学会論文誌, Vol. 27, No. 2, pp. 183-190 (1986).  
(平成6年12月13日受付)  
(平成7年4月14日採録)

### 松尾比呂志 (正会員)



1955年生。1978年九州大学工学部電気工学科卒業。1980年同大学院修士課程修了。同年日本電信電話公社(現, NTT)入社。以来, 大規模集積回路設計技術, 文字認識処理技術, 自然言語処理応用技術(情報検索, 情報抽出), マルチメディアデータベース応用技術の研究に従事。現在, NTT 情報通信研究所主任研究員。電子情報通信学会会員。

### 木本 晴夫 (正会員)



1949年生。1973年大阪大学基礎工学部制御工学科卒業。1975年同大学院基礎工学研究科物理系専攻修士課程修了。工学博士。同年日本電信電話公社(現 NTT)入社。以来, 電気通信設備の計画・設計業務, 光ファイバーケーブル接続技術, 自動索引技術, 情報検索技術, マルチメディア DB 検索技術の研究開発に従事。現在, NTT 情報通信研究所主幹研究員, 東海大学大学院非常勤講師, 情報処理学会データベースシステム研究会情報検索システム評価用ベンチマーク DB 構築 WG グループリーダー, ACM/SIGIR 95, 96 プログラム委員, 47th FID 国際会議プログラム委員。電子情報通信学会, 人工知能学会各会員。