

# 未定義語を含む文の多段階構文解析

5E-5

大場健司\* 元吉文男\*\* 井佐原均\*\* 横山晶一\*\* 石崎俊\*\* 板橋秀一\*  
\*筑波大学 \*\*電子技術総合研究所

## 1. はじめに

自然言語を計算機で構文解析しようとするとき、辞書に登録していない語が現れることがある。このため、未定義語を含む場合でも、解析が正常に終了する方法を考える必要がある。従来の未定義語を扱う研究では、縦型の解析をしているものがほとんどであり、すべての未定義語が得られない場合がある。この対策として、本論文では横型の構文解析システムに未定義語処理を加えた。これは、助詞や用言などを頼りに、名詞(またはサ変動詞語幹)の未定義語を見つける。このとき用言句と名詞句それぞれの文法を用意し、多段階に解析を行う。これにより、解析木の数を抑え、効率よく解析できることが分かった。

## 2. 解析方法

システムは日本語用完全横型探索の構文解析法<sup>[1]</sup>を基本にして構築した。これは、左から1文字ずつ読み込んでボトムアップに解析を行うが、無駄な解析木が伸びないように、次のカテゴリーを予測するトップダウンの手法も採り入れている。まずこれに未定義語処理を加えた。対象とする未定義語は、名詞とサ変動詞の語幹に限定している。未定義語のほとんどはこれらに含まれるため、これにより解析の負担を低減する。未定義語の生成は、次の文法規則によって行う。

未定義語→文字

未定義語→未定義語+文字

ここで問題となるのが、解析に要する処理量の増大である。一般に未定義語を扱っていると、解析過程で文法構造が多く生じてしまい、特に横型では急速に増加する。そこで処理量を減らす目的で、多段階の構文解析を行う。そのために、文法規則を用言句を解析するものと、名詞句を解析するものに分けて、2組用意する。文法を2つに分けると、片方の文法では解析できない部分が生じるので、これは未定義句として残しておき、後でもう一方の文法で解析する。この

方法を発展させると、多段階の解析ができる。例えば、初めに入力文を用言句の文法を用いて解析する。その中に名詞句が含まれていれば、それが未定義句として残るので、次に名詞句の文法を用いて解析する。さらにその中に用言句の未定義句が残れば、再び用言句の文法を用いるというような操作を繰り返す。すると、同じ文字列を2回、3回と解析することもあるが、用言句を解析したいときには名詞句を解析するための文法を使わずに済み、また名詞句を解析したいときにも同様に適用する文法規則を減らせるので、それほど負担にはならないことが予想される。

多段階の構文解析によって効率が上がるのは、一部の木構造が異なるために、解析木の数が増えてしまう場合である。いま仮に、横型の構文解析を多段階にしないで行い、ある文を入力したとき、文頭の名詞句が2通りの解釈ができたとする。横型なので、その後の部分が全く同じ構造であっても、各々の木を成長させなくてはならない。このため、最初の名詞句以外では、同じことを繰り返すことになり、無駄が生じる。多段階解析にすると、まず用言句の文法を使う

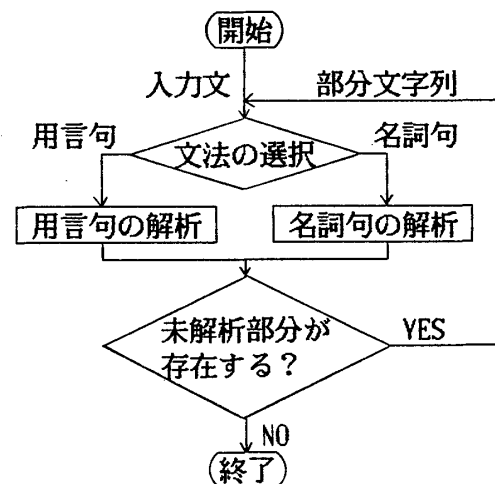


図1 解析アルゴリズム

Multistage parsing for Japanese Sentence Containing Undefined Words

Kenji OBA\*, Fumio MOTOYOSHI\*\*, Hitoshi ISAHARA\*\*, Shoichi YOKOYAMA\*\*, Shun ISHIZAKI\*\*, Shuichi ITAHASHI\*

\*University of Tsukuba \*\*Electrotechnical Laboratory

として、最初の名詞句は1つの未定義句になり、木構造を増やさない。次に名詞句の文法を用いて、未定義句の部分だけを解析して、2通りの解釈を得る。

多段階構文解析のアルゴリズムは、図1のようになる。このとき、用言句の解析の直後に再び用言句の解析を呼ぶことはあるが、名詞句ではそのようなことはない。

3. 解析例

試験的に、文法規則数40程度、辞書語彙数100余りでいくつかの文を入力した。「うみ」、「うみがめ」、「め」、「まえ」という名詞が辞書にある場合に、「うみがめのまえにある。」という文を入力したときの結果を図2に示す。1段目で2つの解釈ができ、2段目で、「うみがめのまえ」と「めのまえ」の各々に対して2つの解釈ができる。3段目では、解釈の数は増えない。この例では、結局4つの解析木が得られることが分かる。また、未定義語として取り出されたのは、「うみがめのまえ」と「めのま

え」である。

4. おわりに

横型構文解析システムに未定義語処理を加え、いくつかの文について解析を行ったところ、未定義語を効率良く取り出せることが分かった。

現在の構文解析システムで、多段階解析を行う際に同じ文字列を複数回走査しないように改良中である。また、意味処理を考えていないため、ふさわしくない解析木が現れているので、今後は意味解析と結び付けて、構文上の曖昧性を解消することを考えている。

参考文献

- [1]元吉・井佐原・石崎：日本語用完全横型探索構文解析法，情報処理学会第32回全国大会4s-3,1986
- [2]高橋・奥村・伊藤・小川：構文解析における未定義語の抽出，情報処理学会第33回全国大会3k-7,1986

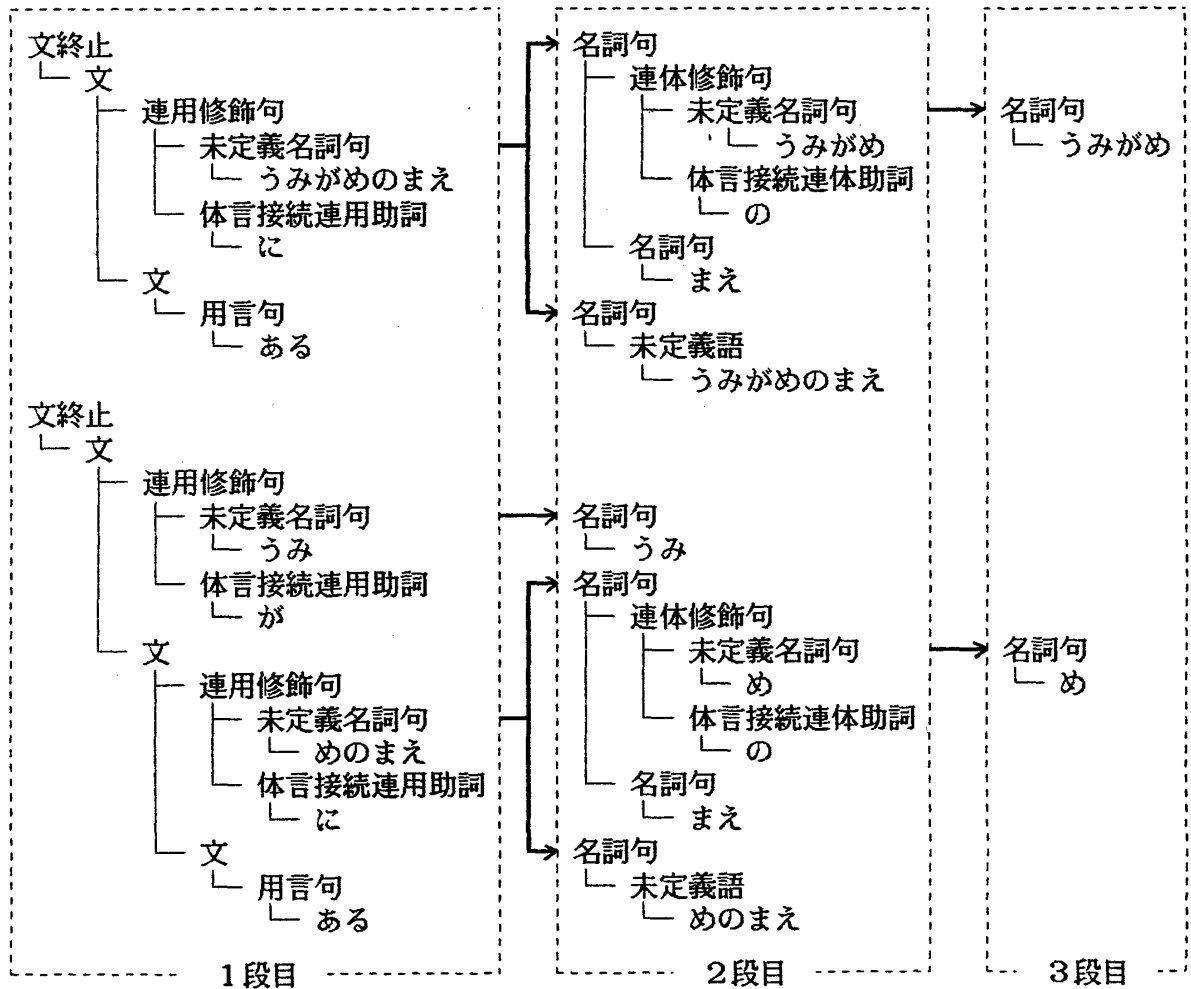


図2 構文解析木の例