

例文からの話し言葉用文法の半自動修正法

2R-4

大谷耕嗣 山本幹雄 中川聖一

豊橋技術科学大学

1 はじめに

自然言語理解システムの研究の1つの目的は、人間とコンピューターの使用言語の相違に基づくセマンティック・ギャップを解消し、人間とコンピューターとの間に新しいインターフェースを確立しようとするものである。我々の研究室での自然言語理解の研究もこの方針に基づき、富士山の観光案内をタスクとする対話システムを開発している。

対話システムは、ユーザーが発話するであろう文を文法として持っており、これによりユーザーが発話した文を認識する。しかし、システムに登録されていない文法を使った文をユーザーが発話した場合には、当然ながらシステムはそれを認識できない。

そこで、システムの文法で受理できない文をユーザーが発話した場合に、それを受理できるように文法を新たに登録することにより、システムが認識できる文の範囲を増やし（カバー率の向上）、その結果として認識率の向上を目指すことが本稿の目的である。

登録方法を未登録単語の登録と生成規則の登録の2種類に分けて考え、システムへのインプリメントを行わない評価・検討を行なった。

2 例文からの登録アルゴリズム

例文からの文法登録アルゴリズムを説明するが、登録は例文の解析の失敗の原因が単語の未登録にある場合と生成規則の未登録にある場合について考える。

2.1 未登録単語の登録

未登録単語を登録するアルゴリズムを図1に示す。未登録単語登録アルゴリズムは入力文の解析が単語の未登録が原因で失敗した時に、その未登録単語を登録するためのものである。解析にはトップダウンのパーザを用いる。これは、トップダウンのパーザが一つ先の単語（単語クラス）を予測しながら解析を進める性質を利用するためである。解析に際して未登録単語の単語クラスは任意と仮定しておく。解析が未登録単語まで来た時、未登録単語の単語クラスは任意であるので

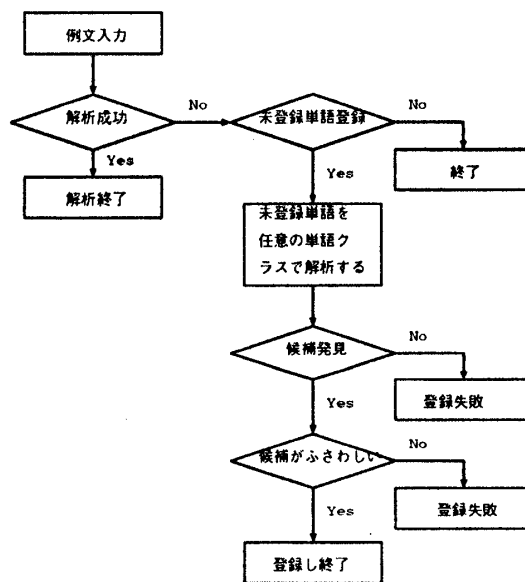


図1: 単語登録アルゴリズム

その時に予測されている単語クラスであるとして解析を進める。その結果、解析が成功した単語クラスがこの未登録単語の単語クラスの候補になる。候補の中でふさわしいものを判断してもらうために登録候補の単語クラスの単語の一部の提示と、未登録単語をその単語クラスに仮に登録してみて、未登録単語を含む例文を生成・提示することによりその登録が適切かどうかユーザに判断してもらう。ふさわしい単語クラスが複数見つかった時には、それら全てに登録する。

2.2 生成規則の登録

生成規則を登録するアルゴリズムを図2に示す。また図2中で使われている3つの登録方法についての例を図3に示す。生成規則アルゴリズムは例文の解析が生成規則の未登録が原因で失敗した時に、入力文が解析可能になる新たな生成規則を登録するためのものである。解析にはボトムアップのパーザを用いる。これはボトムアップのパーザで解析した時に作られる部分解析木を利用するためである。入力文の解析の結果できた部分解析木の組合せで木の数が最小になるものを選び、適度の一般化を与えるために3つの登録方法を使い分けて登録を行なう。以下、3つの登録方法について説明する。

Semi-Automatically Correction Method of Grammar Rules for Spontaneous Speech by Examples  
Kouji Ohtani, Mikio Yamamoto, Seiichi Nakagawa  
Toyohashi University of Technology

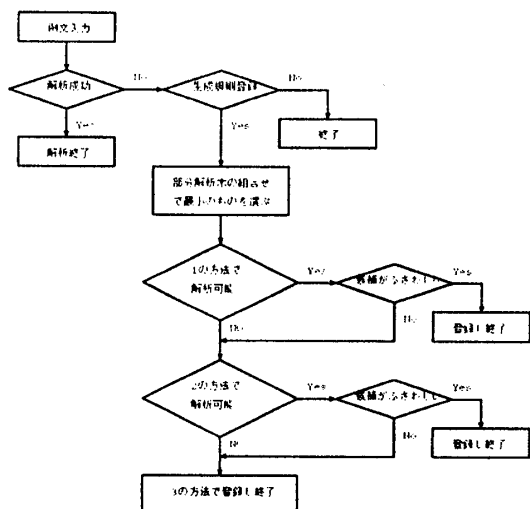


図2: 生成規則登録アルゴリズム

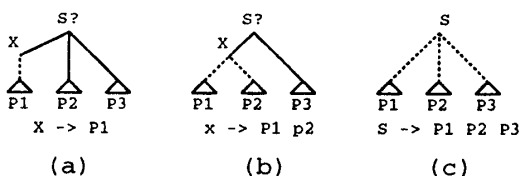


図3: 生成規則登録の際の3つの方法

△: 部分解析木    —: 存在する規則    .....: 登録する規則

1. ある部分解析木の非終端記号を一つ任意に変えてみて解析が成功するか調べる。例えば図3(a)のことで、 $X \rightarrow P1$  という規則を新たに作ってみて解析可能かどうか調べる。
2. 2つの部分解析木を接合する生成規則を作ってみて解析が成功するの調べる。例えば図3(b)のことで、 $X \rightarrow P1 P2$  という規則を新たに作ってみて解析可能かどうか調べる。
3. 最小の部分解析木を開始記号に直接つなげる。例えば図3(c)のことで、 $S \rightarrow P1 P2 P3$  という開始記号に直接つなげる規則を作り登録する。

これらの操作をふさわしい規則ができるまで登録方法の(a)から(c)まで順次実行する。登録方法の(a)と(b)ではふさわしい規則かどうかの判断は仮に登録をしてみても例文を生成・提示することによりユーザーが行ない、ふさわしい規則が見つかったら登録して終了する。ふさわしい規則がある登録方法で複数見つかった時にはそれらを全て登録する。

### 3 評価

今回使用した文法は、本研究室で作成している富山の観光案内をタスクとする対話システムで使用して

いるもので単語数241、生成規則数391、パープレキシティは77.6である。学習に使うための例文の内訳を表1に示すが、これは文献[1]で集められた10人の話者のデータをもとに作成した。表2に文法登録を行なうことにより新たに解析可能になった文の内訳と新たに登録した規則数を示す。表3に学習した文法を使って評価セットを解析してみた結果を示す。

表1: 学習セット(106文)の内訳

	解析可能文	解析不可能文
セット1	39	67
セット2	51	55

表2: 文法登録後、新たに解析可能になった文の内訳

	単語登録	生成規則登録	合計
セット1	0 (0)	35 (38)	35(38)
セット2	5 (6)	32 (30)	37(36)

( ) 内は新たに登録した規則数

表3: 評価結果(解析可能文)

	評価セット	
学習セット	セット1	セット2
セット1	39 → 74	39 → 52
セット2	51 → 58	51 → 88

文法を学習することにより解析可能文が増加しているがまだ評価文の半分ぐらしかカバーできていない。また、学習の結果、パープレキシティはセット1では94.0に、セット2では109.5と1.5倍程度になった。

### 4 むすび

システムには登録されていないがユーザーが発話するであろう文法を新たに登録することにより、システムが認識できる文の種類を増やし、その結果として認識率の向上を目指した。その結果、パープレキシティの増加の割にシステムが解析できる文の種類が増えなかった。そこで現在、パープレキシティの増加を抑えるためにCFGとバイグラムを組み合わせた文法を使って登録できるようにシステムを改良中である。

### 参考文献

- [1] 伊藤, 大谷, 肥田野, 山本, 中川: 「事前説明によるシステムへの入力発話の変化と認識結果の人間による復元」, 情報処理学会, 音声言語情報処理研究会, 94-SLP-47 (1994.12)