並列構造の検出に基づく長い日本語文の構文解析

黒橋　禎夫[†]　長尾　眞[†]

従来の構文解析法は十分な精度の解析結果を得ることができず，とくに長い文の解析が困難であった．このことは従来の方式が局所的な解析を基本としていたことに原因があり，これを解決するためには文内のできるだけ広い範囲を同時的に調べることが必要である．我々は，すでに，このような考え方に基づき，長い文の中に多く存在する並列構造が文節列同士の類似性を発見するという手法でうまく検出できることを示した．本論文では，そのようにして検出した並列構造の情報を利用して構文解析を行なう手法を示す．長い日本語文の場合は 1 文内に複数の並列構造が存在することも多い． そこでまず，文内の並列構造相互間の位置関係を調べ，それらの入れ子構造などを整理する．多くの場合，並列構造の情報を整理した形で利用できれば，文を簡単化した形でとらえることができる．そこで，簡単化した各部分に対して単純な係り受け解析を行ない，その結果を組み合わせることによって文全体の依存構造を求めることが可能となる． 各部分の係り受け解析としては，基本的に，係り受け関係の非交差条件を満たした上で各文節が係り得る最も近い文節に係るという優先規則によって決定論的に動作する処理を考えた． 150 文に対して実験を行なったところ， 96% の文節について正しい係り先を求めることができた．

キーワード: 構文解析，係り受け解析，並列構造，文の簡単化

A Syntactic Analysis Method of Long Japanese Sentences based on Coordinate Structures' Detection

Sadao Kurohashi [†] and Makoto Nagao[†]

Conventional parsing methods can not analyze long sentences precisely, since these methods consider no information in a wide-range series of words in a sentence. We have succeeded in developing a method of detecting coordinate structures by a similarity measure of two arbitrary series of words. This paper describes a method of parsing a sentence by using the information of coordinate structures. A long sentence can be reduced into a shorter form by recognizing coordinate structures in it. Consequently the total dependency structure of a sentence can be obtained by relatively simple modifier/modifiee rules. We report the results of analyzing 150 Japanese sentences to illustrate the effectiveness of this method.

KeyWords: *Parsing, Dependency structure analysis, Coordination, Reduced sentence form*

† 京都大学工学部 電気工学第二教室, Department of Electrical Engineering, Kyoto University