

## 構文エラーの自動リカバリシステムの評価

4D-6

伊貝 耕\* 柴田淳司\*\* 中田育男\*

\*筑波大学 \*\* (株)日立製作所

## 1 はじめに

手書きまたはコンパイラ生成系によるコンパイラにおける構文エラーリカバリは、一般にエラーパターンを予想しそれぞれのエラーパターンについてどのようなリカバリを行うかを決定しておく方法で行われている。しかしながら、この方法では実用的なコンパイラを作成する場合において言語設計者に多くの負担をかけ、しかもリカバリ対象範囲は限定される。

そこで、我々は [1] や [2] などで提案された自動エラーリカバリアルゴリズムにリカバリ選択アルゴリズムを加えた自動エラーリカバリシステムを作成し、そのシステムを組み込んだ構文解析器生成系を実現した。さらに、実際に構文解析器生成系を用いて生成した Pascal 構文解析器にエラーのある Pascal プログラムを入力し、リカバリの適用率が極めて高いものであることを確認した。

しかし、このシステムのリカバリは機械的なものであるため、必ずしも適切なものではない。リカバリをより適切なものに改良するために、全てのリカバリ結果の適切/不適切を人間の目で確認、集計した。本稿では、自動エラーリカバリアルゴリズムの簡単な説明と集計結果、およびその考察を述べる。

## 2 自動エラーリカバリアルゴリズム

[1] や [2] で提案されたアルゴリズムは入力トークンに対して何らかの修正を行い、その修正した結果を実際に解析することを繰り返して、妥当な修正を行うという方法である。我々の自動エラーリカバリシステムでは、この方法は以下の3種類に分類される。

## ○ 第一段階リカバリ

エラートークンのおよびその直前のトークンに対する、削除、別のトークンへの置換、直前への別のトークンの挿入。

## ○ 第二段階リカバリ

エラートークンの前後にまたがる部分列の削除。

## ○ スコープリカバリ

閉じていないスコープを閉じるようなトークンの挿入。

実際にはこれらのリカバリが組み合わせられて用いられることもある。また、リカバリの候補が複数ある時には、最小限の修正を行うようなリカバリを選択する。

## 3 集計

評価に用いたエラーを含む Pascal プログラムは 772 本である。これらを全て生成した構文解析器に入力した結果、4429 個のエラーが発見された。この内、リカバリが成功したものは 4304 個で全体の 97.2%。リカバリが失敗したものは、125 個で全体の 2.8% である。

得られたリカバリ結果を全て確認し、適切なもの、不適切なもの、不適切なリカバリによって生じた副次的エラーに対するものに分類、集計した。適切不適切の判断は、そのリカバリによって副次的な構文エラーあるいは意味エラーが起きるかどうかでを行った。その結果を表 1 に示す。また、各リカバリについて、適切なリカバリ、不適切なリカバリの割合を集計した。その結果を表 2 に示す。

次に、不適切なリカバリについて、副次的に意味エラーを起こすものと、構文エラーを起こすものに分類を行った。集計結果を表 3 に示す。

構文エラーを起こすリカバリについては、現在のリカバリアルゴリズムの適用を改良することで改善可能なもの、現在のリカバリアルゴリズムでは適切なリカバリが不可能なもの、人間の目で見ても正しいリカバリが不明なものに分類を行った。集計結果を表 4 に示す。

The Valuation of an Automatic Syntax Error Recovery System

Koh Ika\*

Atsushi Shibata\*\*

Ikuo Nakata\*

\*University of Tsukuba

\*\*Hitachi,Ltd

## 4 考察

表1において特徴的なこととして、副次的構文エラーの数が非常に多いことが挙げられる。実際に、自動エラーリカバリシステム自身が副次的構文エラーを作り込んでしまう、あるいはエラー部分を完全に排除しきれず直後で再びエラーが発生することがあった。

また、表2より、リカバリの種類によって、不適切なりカバリが行われる割合に大きな偏りがあることがわかった。特に、トークンの置き換えと第2段階リカバリについては不適切なりカバリ割合が高い。このことは、これらのリカバリが他のリカバリより比較的複雑な操作を行っているためとも考えられる。しかし、改善すべき点であるということもまた確かである。

表3によると、副次的エラーを起こすリカバリのうち、構文エラーを起こすものは394個あり、これは副次的エラーのリカバリを除いた本来のエラーのリカバリ総数1728個の16.7%と比較的少数である。しかし、副次的構文エラーを引き起こす不適切なりカバリ1個に対して、副次的構文エラーは実に平均6.9個発生している。そのため、このような不適切なりカバリを防ぐような改善を施すことは重要であると考えられる。

表4によると、現在のリカバリシステムのアルゴリズムでも、改良を加えることにより不適切なりカバ리를半減させられる可能性がある。現在のリカバリアルゴリズムでは改善不可能なものについては、複数トークンの挿入/置換といった新たなリカバリアルゴリズムを考える必要がある。また、人間の目で見てもどうすれば適切なりカバリが行えるのか不明といったリカバリ困難な場合も合わせて、そういった場合にリカバ리를いかに安全側に倒すかという方向も考えられる。

## 5 おわりに

エラーは千差万別であり、それら全てについて適切なリカバリを行うのは非常に難しい。しかしながら、リカバリシステムを実用的なものにするためには、現状のシステムよりさらに高い性能が要求される。それには、与えられたソースプログラムからより多くのリカバリに有用な情報を引出し、利用する必要がある。たとえば、意味情報、予約語情報、テキスト位置情報といったものが挙げられる。

今後は、汎用性を損なわない範囲での、不適切なりカバリの減少を目標としてシステムの改善を行う予定であり、すでにテキスト位置情報を考慮した改善を試みている。

	適切	不適切	副次的	計
リカバリ数	1103 個 24.9%	625 個 14.1%	2701 個 61.0%	4429 個 100.0%

表 1: リカバリの適切性

リカバリ種類	適切	不適切
トークンの削除	96.4%	3.6%
トークンの挿入	73.8%	26.2%
トークンの置換	28.5%	71.5%
第二段階リカバリ	24.9%	75.1%
スコープリカバリ	73.2%	26.8%

表 2: リカバリ種類別適切率

	意味エラー	構文エラー	計
リカバリ数	231 個 37.0%	394 個 63.0%	625 個 100.0%

表 3: 副次的エラー

	可能	不可能	不明	計
リカバリ数	180 個 45.7%	183 個 46.4%	31 個 7.9%	394 個 100%

表 4: 改善可能性

## 参考文献

- [1] Burke and Fisher: A Practical Method for LR and LL Syntactic Error Diagnosis and Recovery, ACM Transactions on Programming Languages and Systems, Vol.9, No.2, April 1987, Pages 164-197
- [2] Philippe Charles: A Practical method for Constructing Efficient LALR(k) Parsers with Automatic Error Recovery, Ph.D dissertation, New York University, May 1991