

的確な誤り指摘を行なうための並列構文解析法

7G-6

茨木康之 武田正之

東京理科大学理工学部

1. 序論

我々は特定の言語に対する知識を持ち、編集時に誤り診断、自動回復などの編集支援を対話的に行う機能を提供するエディタ(言語指向エディタ [1])を開発中であり、その構文解析機構に並列構文解析システム PAX[2]を導入し、複数の構文の可能性に対する並列な誤り探索において有効性を検討している。しかし PAX も含めた一般的な構文解析法は本質的に「誤りを含まない文書」の解析を目的としており、言語指向エディタで扱われるような「誤りを前提とする文書」に対する誤り指摘という観点では的確さの欠如が否めず、それを補うには対象とする言語固有の処理を付加する必要がある。そこで上昇型チャート構文解析 [3]を拡張し、「誤りを前提とする文書」に対する的確な誤り指摘を行うための汎用な言語解析法を提案し、本稿にて概要を報告する。

2. 上昇型チャート構文解析法における誤り指摘

チャート法は文脈自由文法における構文解析アルゴリズムの枠組であり、以下のデータ構造と、その生成手続きにより構文解析アルゴリズムを表す。

[項] 任意の非終端記号を根とする構文木を表す。木構造が決定した場合、その部分木を表す項を内包し、未定の場合、空所と呼ばれる特別な項で表す。また項に含まれる最左の空所を最左空所と呼ぶ。

[接点] 解析対象となる字句間に与えられた一意な番号で、2接点の組により字句の範囲を表す。このとき字句の範囲の左端を始点、右端を終点と呼ぶ。

[弧] 構文木および、それが含む字句列の範囲を項と2接点の組により表す。特に木構造が完成している項をもつ弧を不活性弧、それ以外を活性弧と呼ぶ。

PAX を包含する上昇型チャート構文解析法は、全ての終端記号に対応する不活性弧が存在する初期状態から、以下の手続き 1,2 を、新たな弧が生成できなくなる終了状態まで繰り返す。以下では非終端記号および終端記号を文法記号と呼ぶ。

[手続き 1] 任意の生成規則の右辺での最左の文法記号に対応する不活性弧に対し、右辺の文法記号に対応する空所を内包し左辺の非終端記号に対応する項の最左空所を不活性弧の項で置き換えた項をもち、不活性弧と同じ範囲をもつ弧を生成する。

[手続き 2] それぞれ終点と始点を共有する任意の活性弧と不活性弧に対し、活性弧の最左空所と不活性弧の持つ項の文法記号が一致するとき、最左空所を不活性弧の項で置き換えた項をもち、活性弧の始点と不活性弧の終点を範囲としてもつ弧を生成する。

終了状態において、開始記号に対応する項をもち全ての字句を範囲にもつ不活性弧が存在すれば解析は成功であり誤りは含まれない。PAX では、この不活性弧の部分木になり得ない弧に対し手続 1 を適用しないことにより無駄な弧の生成を制限し、その制限に以下の予測と呼ぶ概念を使用している。

[予測] 字句列の先頭の接点は開始記号を予測する。任意の活性弧が始点に予測される文法記号を根とする構文木の部分木となり得るとき、その終点は項の最左空所に対応する文法記号を予測する。

解析が失敗したとき接点が予測する任意の文法記号に対し、その文法記号を根とする構文木の部分木になり得る弧が存在しないとき、その接点において誤りがあると考え、「予測される文法記号の欠落」、「直後の終端記号の余剰」が誤りとして指摘できる。

しかし PAX では上記のどちらの指摘が的確かは判断できない(問題点 1)。また構文木の最左の葉となるべき終端記号が存在しない場合、終端記号自体の欠落を指摘できない場合がある(問題点 2)。

3. 的確な誤り指摘のための拡張

本稿で提案する解析法は上記の問題点1,2に対応するために上昇型チャート法に以下の拡張を加えたものである。ただし弧には構文誤りを表す情報を付加できるものとする。

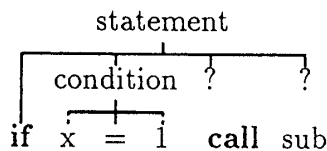
[拡張1] 手続き2において文法記号が一致しない不活性弧を読み飛ばす手続き3と空所を読み飛ばし欠落したものととして木構造を決定する手続き4を追加する。読み飛ばした空所および不活性弧は生成された弧に情報として付加される。

[拡張2] 手続き1における「右辺における最左の文法記号に対応する不活性弧」という制限を外し、「右辺における文法記号に対応する不活性弧」に対し弧を生成することとし、同時に手続き2,3,4に対し、それぞれ終点と始点を共有する不活性弧と活性弧に対する新たな弧の生成も可能とする。

[拡張3] 弧に付加された情報から誤りの個数を算出し、誤りの個数が少ない弧に対し優先的に各手続きを適応する。

拡張3での誤りの個数とは弧に付加された情報から一意に定められる値で、どのような状態を1つの誤りと見なすかという判断基準を数値化したものである。例えば、「高々1つの文法記号の欠落および余剰」を1つの誤りとするならば、手続き3,4において生成された弧の誤りの個数は、そのもととなった弧の誤りの個数を1増加させたものである。

例えば上記の誤りの個数算出方法を適用しPL/0 [4]の文法においてstatementを予測する時点で字句列“if x=1 call sub”を解析するとき、PAXでは以下のような活性弧が張られた段階で解析が停止する。

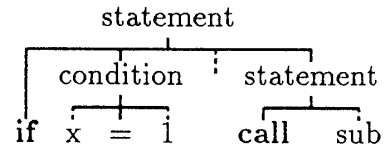


この活性弧は以下の生成規則により木構造が決定し、

statement \rightarrow if condition then statement

その終点(“call”の直前)は“then”を予測しており、「then」の欠落」または、「call」以降の字句の余剰」が誤りの候補として指摘可能だが、どちらの

指摘が的確かは判断できない。しかし本解析法においては以下のような誤りの個数が1の不活性弧が得られる。



このとき「“call”以降の字句の余剰」を示唆する不活性弧の誤りの個数は最低でも3となり、「then」の欠落」が的確な誤り指摘であると判断可能となる。

また同じく statement を予測する時点で ident である字句列 “sub” を解析する場合以下の生成規則

statement \rightarrow ident := expression

statement \rightarrow call ident

に関する誤りと考えられる。しかしPAXを用いた場合、解析のきっかけとなる“call”の自体の欠落は指摘できない。しかし本解析法の場合 ident をきっかけとして statement に対応し後者の生成規則による不活性弧が生成され「call」の欠落」も指摘可能となる。

4. 結論

本稿では誤りを含む文書に対する的確な誤り指摘を行なうための並列構文解析法を提案した。また本解析法を並列論理型言語 KL1 により実現し、言語指向エディタへの組み込みを通じ実験を行ない、誤り指摘を行なうための十分な情報を生成する点において既存の構文解析法に比べ適当であるとの評価を得た。しかしながら解析効率の面や、意味解析機構との併合が行なわれていない点など、未解決の問題が残されており、今後の研究課題としたい。

参考文献

- [1] 武田 正之: 知識ベースに基づく Language oriented Editor, 情報処理学会論文誌, 28, 11 (1987).
- [2] 渕 一博: 並列論理型言語 GHC とその応用, 第7章, pp.131-146 (1987).
- [3] 田中 瑞穂: 自然言語理解, オーム社, 第3章, pp. 59-100 (1988).
- [4] Niklaus Wirth: Algorithm + Data Structures = Program, Prentic-Hall, Inc. (1976).