

7E-2

日本語対話文解析文法の開発環境

松尾秀彦 谷田泰郎 永田昌明*

* (株) 東洋情報システム

** (株) ATR自動翻訳電話研究所

1 はじめに

ATR自動翻訳電話研究所で開発された音声言語日英翻訳実験システムSL-TRANS¹⁾の日本語対話文解析部は、解析過程の制御が自由なアクティブ・チャート解析法²⁾と単一化に基づいた語彙-統辞論的な文法的枠組みであるHPSG³⁾を採用している。

この日本語対話文解析部では、文法の中に統語的情報だけでなく、意味的・運用論的情報を取り込んでいる。素性構造の単一化に基づく制約の伝搬の枠組みを用いながら、統語論、意味論、運用論といった異なる部門に由来する情報を統合することができる。そして、文法を中心とする語彙規則と生成規則を用いて、構文/意味解析を行ない、日本語意味構造を得る。

文法開発者は日本語対話文解析部の性能 (Coverage, Performance) となる文法に対して、言語学的な知識を記述するために、個々の言語現象の位置づけを明らかにするような努力をしている。その文法は様々な操作や膨大な情報が相互に密接に関係しあい非常に複雑多様な構造になっている。文法開発者は文法記述の正当性を種々の着想や方策により試行し、立証するために、頻繁に拡張や修正を加える。

このような試行錯誤を複雑な構造に対して繰り返してゆくためには、文法開発者が解析処理中の文法規則の適用状況や文法情報を適切な位置でモニタリングでき、容易に文法に対する拡張・修正を行なえるような柔軟な開発環境が必要である。

これまで、日本語対話文解析部の中では、パーザの動きを監視するデバッグ環境は存在していたが、文法を開発するための支援ツールは整備されていなかった。文法開発において文法の不整合や誤りを発見する場合、文法開発者の経験や知識などに頼っている状況であった。そこで、日本語対話文解析部の文法開発支援ツールを構築したので、この内容について我々は報告する。

2 支援概要

文法記述言語 (文脈自由文法の範囲で生成規則を書け、その規則に対しての素性構造、経路方程式及び注釈を記述することができる) とチャートパーザの中に文法開発者にとって有益な文法情報の分析が行なえるようなインターフェースを持つデバッグ環境を付加する。支援環境としては、

- チャート法による解析をトレースし、文法中の不整合を究明、発見するために、チャートパー

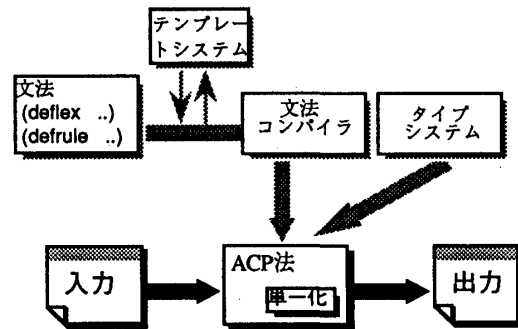


Fig. 1 日本語対話文解析部システム図

ザ/文法記述言語の拡張を行った。

- 文法中に現われるオブジェクト名を自動管理する機能を用意した。
- システムとチャートパーザのユーザーインターフェースとして、文法情報のインタプリタを作成した。

などである。

チャートパーザの拡張

宣言的な枠組みでは明示的な処理手順の指定が存在せず、規則が適用される動作環境を参照して、文法の不適合を捕捉することが困難であった。解析動作のトレースや中断 (ブレイク) ができ、その際エッジの選択状況や規則の適用状況などが参照できる。

オブジェクト名の自動管理

文法が大規模化されると、語彙の増加に伴い、文法中の素性名表、タイプ名表、生成規則名、語彙項目名、及びテンプレート名のオブジェクトの中に、定義されていないオブジェクトが出現する可能性がある。それらのオブジェクトの相互関係を検査し自動的に管理するツールを作成した。また、このツールは文法情報の検索システムとしても操作できる。

文法記述言語の拡張

アクティブチャートパーザでは、適用する規則が自由に動的に選択されるために、処理中のどのような場面で文法を適用しているかに着目することが困難であった。そこで、文法記述中に中断条件を記述できるようにした。

インターフェース

ACP法で日本語会話文を処理中、ブレイクされた後に使用できるコマンド群がある。文法開発者が各種操作によって価値のある情報を参照したり、抽出

したりできる。

3 支援内容

チャートパーザの拡張

ACP法では完全な部分木を表す不活性弧と不完全な部分木を表す活性弧を結合しながら解析処理を進め、待機弧リストを使用して次に弧の結合を行なうための弧の選択方法の指示をすることで解析処理の自由度を高くしている。このような処理方式では文法規則の枠組みや文法記述の妥当性を推し量るために、規則、チャート構造、弧などをどの状況で視覚化するかが問題になる。

ここでは、下記の状況で情報のアクセスを行なえるようにした。

- ・待機弧から弧を選択する前
- ・弧をチャートに加えようとした時
- ・弧を待機弧に加えようとした時

この状況下で処理の中断やトレース情報の作成を行なう。また、中断時に使用できる機能は、以下の通りである。

- ・表示機能
ある状態で、選択された弧、チャート、適用される文法規則、待機弧等の表示を行なえる。
- ・ステップ機能
チャートパーザの基本的なサイクルの中で、各種操作を行ない、チャートに弧を取り込む状況を監視できる。
- ・インスペクタ機能
言語的な知識が、内部的にはLISPのデータ構造を使って表現されている。その内部構造を文法開発者に印字機能を使って表示し、チャートや弧などがどのような文法から形成されているかを点検できる。
- ・ASSIGN機能
任意の変数を設定したり、その変数を取り出すことができる。
- ・モード設定機能
簡易言語を利用して各機能をインプリメントしている。文法開発者はこの簡易言語を使って、容易に機能をカスタマイズできる。

これらの機能によって、文法開発者はLISP中の動作方式やデータ構造を意識せずに、チャート法と単一化の基本的考えさえ把握しておけば、解析中、文法に関する情報を参照、検査できる。

オブジェクト名の自動管理

文法記述中に出現するオブジェクトを自動的に管理する。これを実現するためにLISPのドキュメンテーション機能に類似したものを用意した。部分文字列からオブジェクト名を検索するAPROPOS機能やオブジェクト名がどのファイルに存在しているかを検索するWHERE-IS機能など、その他にもONLINE-HELP、DOCUMENTATION、DESCRIBE、PRINT及びCOMPLETIONがある。また、素性名の階層関係を視覚化する機能も用意した。

文法記述言語の拡張

文法の定義を行なうための特殊形式である生成規

則の定義、語彙規則の定義の中にデバッグ用の印字構文やリスプ構文を埋め込めるようにした。それによって、解析処理中に、文法の適用位置の把握や文法情報の参照ができ、文法の修正/拡張が容易になった。

インターフェース

コマンドを文法開発者とシステムとのインターフェースとして用意している。解析中断後、文法の組み上げ状態やチャートのデータ構造がどのような状態であるかを、表示、ステップ、ASSIGN及びインスペクタを利用して会話型で点検してゆくことができる。

4 今後の課題

対話型日本語解析文法の開発支援環境として、チャートパーザの実行中に種々の点検作業が行なえる機能を作成した。現在の文法支援環境では、文法を修正してから実験するまでのターンアラウンドタイムが長く、効率よく文法の生産ができない。これらを時間短縮できるような仕組みが必要である。具体的には、文法適用が失敗した場合に文法記述のどの部分に不適格な記述があるかを半自動的に警告したり、文法コンパイル時の各種チェック、文法記述を簡素化するために用いられているテンプレートの自動修正機能などが重要課題である。また、現状の支援ツールのインターフェースはTTYイメージであるが、操作性の容易さを実現するためにはウィンドウベースのインターフェースなどを採用することが考えられる。

謝辞

本研究の機会をくださるとともに適切な助言を頂いたATR自動翻訳電話研究所 森元データ処理研究室長に感謝します。議論に参加して下さいたデータ処理研究室に所属する研究員の方々に感謝します。

参考文献

- [1] 森元、小倉、相沢、樽松：「音声言語日英翻訳実験システム(SL-TRANS)の概要」、情処学会第39回大会、1989。
- [2] Martin Kay: "Algorithm Schemata and Data Structures in Syntactic Processing", Nobel Symposium on Text Processing, 1990.
- [3] Pollard, C. and Sag, I.: "Information-Based Syntax and Semantics - Vol. 1", CSLI Lecture Notes No. 13.
- [4] 小暮：「解析過程の制御を考慮した句構造文法解析機構の検討」、ATRテクニカルレポート、1988
- [5] 永田、久米：「SL-TRANSにおける日本語文法の概要」、ATRテクニカルレポート、1990