

3C-10

知的情報検索システム IRIS における 構文意味解析部のデバッグ環境

玉田 郁子, 秋山 幸司, 鈴木 香緒里, 川崎 正博
(富士通株式会社)

1. はじめに

知的情報検索実験システム IRIS (an Intelligent information Retrieval and Information Selective system) は、簡単な日本語で入力された検索要求について、その回答となる意味内容を持つテキスト群をテキストベースから検索することをめざすシステムである〔杉山 '86〕。現在、IRIS ではシステムの枠組みの妥当性の検証を目的としたプロトタイプを試作を終了し、実用性の検証を目的として、大規模テキストベースに対応可能な拡張システムの試作に取り組んでいる〔伊吹 '87, 秋山 '88〕。

IRIS では、質問文の意味に合ったテキストを取り出すために、質問文と検索されたテキストを構文意味解析して意味の照合を行っている。ところが、新規のテキストを扱う場合、システムが持つ従来の知識(意味モデル・解析ルールなど)が不足し、解析に失敗する例が現れてくる。従来の IRIS の性能を維持するには、意味モデル・解析ルールを拡張/改良しなければならない。しかし、テキストベースの規模が大きくなるほど、システム開発における拡張/改良にかかる工数は増加する。これをいかに効率良く進めるかが、システムの実用性を左右するとも言える。

そこで、拡張システムでは、システムの知識の拡張を支援する環境の一環として構文意味解析部のデバッグ環境を開発した。本稿では、構文意味解析部のデバッグにおける問題点とその具体的な解決方法について報告する。

2. 構文意味解析部のデバッグの問題点と解決方法

構文意味解析部のデバッグのプロセスは、(1)解析結果のチェックと(2)失敗した解析の原因究明から成る。

(1) 解析結果のチェック

IRIS における構文意味解析部の解析結果は網構造(通常はアサイクリックグラフ)である。デバッグする者は単語間の格関係が正しいか否かで解析結果の妥当性を判断する。

解析結果としての網構造は、表示の重複を許せば、図1のような単語毎にインデントを付けた形式で

```
Debugging Environment of syntactic and semantic
analysis in IRIS
```

[kuko TAMADA, Kohji AKIYAMA, Kaori SUZUKI and
Masahiro KAWASAKI (FUJITSU Limited)

```
object : $company class : #company
name : 0
lexicons :
class : #company desc : A社 name : 0
- < agent-case>
object : $actionPut class : #actionPut
name : 0
lexicons :
class : #actionPut desc : 販売 name : 0
< object-case>
object : $product class : #product
name : 0
lexicons :
class : #product desc : ソフトウェア name : 0
< agent-case>
object : $company class : #company
name : 0
lexicons :
class : #company desc : A社 name : 0
```

図1 テキスト文「ソフトウェアを販売するA社」
に対する構文意味解析の解析結果の表示例

表示することができる。プロトタイプでは解析結果の表示にこの形式を用いた。

デバッグの手順を考えると、まず格関係をみるのが先であり、単語に関する詳細情報は必要な時に得ることができればよい。図1の表示形式では、格関係と単語に関する詳細情報が同時に表示されていて、格関係が一目で分かりにくい。解析結果は格関係が分かり易いように図形的に表現し、単語に関する情報はデバッグする者との対話形式で表示する必要がある。

(2) 失敗した解析の原因究明

解析結果だけでは原因を推測できない場合、構文意味解析部をトレースして失敗の原因であるルールを捜し出さなければならない。しかし、トレーサだけでは大量テキストの処理に限界がある。そこで、構文意味解析を行うと同時に単語に適用された解析ルール・意味関数の履歴(解析ログ)をとり、これをデバッグする者が参照し易い形式で提供する環境が必要になる。

以上の検討から、拡張システムが提供すべきデバッグ環境の要点として以下の二点を絞った。

課題① 解析結果の表示方法

課題② 解析ログの参照方法

3. 解析結果の表示方法

IRISのシステムの開発マシンはPSI、記述言語はESPである。解析結果の表示ツールは、図形表示や、単語に関する情報をデバッグする者との対話形式で表示することを考慮して、ウィンドウ・システムのグラフィック機能やメニュー・選択機能を利用している。

図2は、IRISの構文意味解析部の解析結果表示ツールによる網構造の表示例である。網構造における単語をボックスで、格関係を矢印で表している。矢印の傍らの略号は単語間の格関係を示している。

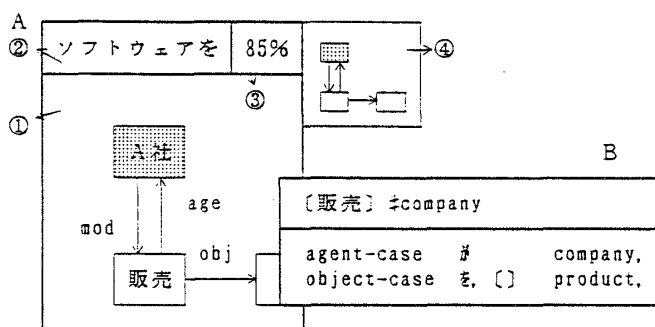


図2 解析結果の図形的表示例

図2 Aのウィンドウは3種類の部分ウィンドウで構成されている。それぞれは、①網構造表示、②入力文表示、③表示領域の割合表示用のウィンドウである。③の数値は①で表示している網構造が網構造全体の何割にあたるかを示している。また、③は網構造全体の略図(図2 ④)を表示する機能を持っている。網構造の単語の数が多くて表示しきれない場合は、表示させたい部分を略図の上で指定すればよい。①にその部分にあたる網構造が表示される。

表示における煩雑さを避けるために、図2 ①のウィンドウでは単語しか表示していないが、デバッグの際に利用する単語の情報には単語が持つ格マーカ・格フィルターなどがある。これらを参照する時は、単語情報表示の指示を図2 ①のウィンドウに与える。図2 Bは単語「販売」に関する情報の表示例である。

4. 解析ログの参照方法

解析ログとしては以下が考えられる。

- i) 解析の対象となった単語
- ii) 解析の対象となった単語のインスタンス名
- iii) 適用した解析ルール名、あるいは、意味関数名
- iv) 格関係名
- v) 結果(成功/失敗)

解析ルール、意味関数は適用条件部と実行部に分かれて

いる。i) ~ v) は適用条件部の実行の前後で機械的に抽出する。取り出した解析ログは、検索と形式化による取扱の容易性を考慮してDB化した。検索にはシステム用に開発した機械的DB検索システムを活用している。

デバッグ者は、解析に失敗した単語や格関係をキーワードとして検索し、これに適用したルールから失敗の原因であるルールを特定する。図3は単語「販売」の主体格(agent-case)に関係するルールの検索例である。図3 Cは検索条件入力ウィンドウ、Dは検索結果の表示ウィンドウである。「A社」を「販売」の主体格としてまとめあげるルール advp-op1 に失敗したことがわかる。

検索 解析ログ DB						表示 検索 And Or
<pre>>search(単語 1 = "販売"). - ir-Set0 8件 >search(格=agent-case) . - ir-Set1 4件 >and(Set0, Set1). - ir-Set2 3件</pre>						
単語 1	単語 2	ルール名	関数名	格	結果	
販売	ソフトウェア	advp-op1	Acheck	agent	失敗	
販売	A社	advp-op1	Acheck	agent	失敗	

図3 解析ログの検索例

5. 今後の課題

意味構造の図形的表示による部分網構造の表示では、機械的に部分網構造を展開しているため、うめこみ文などの表示では煩わしさを感じる。人間の考える図形により近いものを表示する方法を検討する必要がある。

謝辞 : 本研究は第五世代コンピュータプロジェクトの一環として行われた。御指導頂いた ICOT の方々に深謝致します。

【参考文献】

- 〔杉山 '86〕杉山ほか「自然言語理解に基づく情報検索システム IRIS」情報処理学会自然言語処理研究会資料58-8, 1986.
- 〔伊吹 '87〕伊吹ほか「自然言語インターフェースとしての IRIS」情報処理学会第34回全国大会予稿集 pp. 1325-6, 1987.
- 〔秋山 '88〕秋山ほか「テキスト情報の知的検索における諸問題」情報処理学会データベース研究会資料64-3, 1988.