

## 曖昧性解消のための対話機構を備えた日本語解析系

亀田 雅之 石井 信 伊東 秀夫

(株)リコー・中央研究所

あらまし 自然言語文の曖昧さの解消には、解の組合せの爆発と適切な優先解釈の難しさがある。近年、機械翻訳システム等において、この曖昧さ解消を効果的に扱う手段として、ユーザの支援を受ける対話型機構を備えたアプローチが取り入れられつつある。この対話機構は、解析系の能力不足をユーザが補完するもので、ユーザの負担はあるが、それに応じた効果-解析結果の信頼性向上-が期待できる。

我々は、ユーザのさまざまな利用形態、品質要求に応える応用性の高い日本語解析系を目指し、日本語解析系単独での高い解析能力を求めつつ、システム側から効率的に問合せるシステム主導対話とユーザ側から解析結果を修正介入するユーザ主導対話という2種類の対話機構を備えることを目標にした。これに基づき、(1) 係り受け関係候補群により曖昧性を内在する表現の上で、2種類の対話の枠組みを検討し、さらに、(2) LFGを拡張した機能制約文法の枠組みの上で、構文意味レベルの有効な複数解を順位付きで、高速に得るべく試作した日本語解析系に、この対話機構を組み込んだ。また、対話的な文入力時のユーザの待ち時間の減少を狙った即時処理機能も合わせもたせた。

本稿では、まず、従来の曖昧性解消のための対話と関連する研究のレビュー及び我々の設定した要請を示した後、上記(1)、(2)について述べ、また、その解析実験の結果を示す。

## Interactive Disambiguation in a Japanese Analysis System

Masayuki KAMEDA Shin ISHII Hideo ITOH

R & D Center, RICOH COMPANY, LTD.

16-1, Shin'ei-cho, Kohoku-ku, Yokohama, Kanagawa, 223, Japan.

**Abstract** We have designed a practicable system for use in various situations requiring high-quality analyses of Japanese sentences, incorporating two styles of user interaction for disambiguation - system-directed and user-directed. The system-directed interaction is highly effective and the user-directed allows the user to correct analyses. 1) Their framework was first designed in a system which expresses ambiguities as a set of candidates representing possible dependency structure analyses, and then 2) they were incorporated into a system which produces ordered syntactic-semantic structures using a functional constraint grammar.

This paper first reviews related works and explains the design requirements of our interactive Japanese analysis system. Then, we detail the above two styles of user-interaction in the two approaches.

## 1. はじめに

近年、機械翻訳システムにおいて、従来の大量翻訳指向のバッチ型に対して、対話型のシステムが提案されるようになってきた [3,11,13,15,16]。ここでいう対話型とは、機械翻訳システムの前/後編集を対話的に行ういわゆる対話型エディタではなく、システムの能力をユーザが協調的に補完するための対話機構をいう。即ち、現在の自然言語処理技術で扱いの難しい解析での曖昧性解消や生成での訳語選択等の問題を、ユーザを含めた系で解決しようという試みである。このアプローチは、特に、原言語を母国語とした Export 型翻訳では、ユーザが文脈まで含めて原言語文の内容を十分に把握できる [16] ことから、解析の曖昧性解消には非常に有効となる。

日本語の解析技術は、日英機械翻訳システムだけでなく、DB 検索での日本語問合せ、日本語テキストの理解・内容検索あるいは要約、さらにはさまざまな大量の日本語テキスト DB に対する高度な対話/問合せ、あるいは辞書開発用分析ツール等、今後の高度な日本語処理のキーとなる重要な技術である。そこで、我々は、最も適用の進んだ機械翻訳に限らず、こうしたさまざまな応用系やユーザのさまざまな利用形態、品質要求に応じることが想定して、構文意味レベルで、高い解析能力、高速性、汎用性を備えた日本語解析系を目指している。

高い解析能力のためには、意味処理に加え、さらに文脈や一般知識、推論等も含めた高度な解析・理解技術が必要となるが、これらは、当面実現が難しい。そこで、我々は、解析系単独での高い解析能力を求めつつ、さらに、現状の解析技術の不十分さを補うために、人間(ユーザ)の支援を受けて、曖昧さの解消を行う対話機構を取り入れることで、全体としての解析能力の向上を目指した。この対話機構により、ユーザは、システムとの対話という労力を提供する代わりに、高い品質の解析結果を得ることができる。ただし、システムには、ユーザとの対話を必要最小限に止めることが求められる。

対話機構としては、システム側から効率的に問合せるシステム主導対話とユーザ側から解析結果を修正介入するユーザ主導対話という 2 種類の対話を備えることを目標にした。この目標に基づき、(1) 係り受け関係候補群により曖昧性を内在する表現の上で、システム主導対話とユーザ主導対話の枠組みを検討し、さらに、(2) LFG を拡張した機能制約文法の枠組みの上で、構文意味レベルの有効な複数解を順位付きで、高速に得るべく試作した日本語解析系に、このシステム主導対話とユーザ主導対話を組み込んだ。また、対話的な文入力時のユーザの待ち時間の減少を狙った即時処理機能も合わせもたせた。

本稿では、まず、従来の曖昧性解消のための対話と関連する研究のレビュー及び我々の設定した要請を示し、その上で、上記 (1), (2) について述べ、また、その解析実験の結果を示す。

## 2. 従来に関連するシステム・研究

### 従来型の機械翻訳システム

従来の日英機械翻訳の商品システムでは、処理効率や日本語の意味レベルの解析を含む曖昧さの扱いの難しさ等から、深さ優先の唯一解方式をとっているものが多い [1,2]。この方式では、解析途中のローカルな選択の積み重ねのため、最終的にグローバルには妥当な解に至らず、解析率を落としてしまう傾向にある。

この問題の対処としては、簡単には、直前の他の選択肢を順次選択し直すというバックトラック [1] が考えられるが、規則の適用順序に依存するため、次の解が妥当であるという場合は少ない。

### 全解探索パーサ

曖昧性は、本質的には、全解を求めるとともに、各候補に優先点を与えることができれば、その優先点の順に解を順位付けることができる。しかし、実際には、曖昧さの組合せ的爆発による処理の限界が問題になる。

この問題を縮退した解析木 (Shared-packed Forest) により避けながら、句構造規則による構文レベルでの全解探索を効率的に進める Tomita パーサ [3] や同様に高速に効率的に全解探索を行うパーサ [4,5] が提案され、実証されているが、いずれも構文レベルの英語パーサである。日本語解析での JPSG に基づいた構文意味レベルでの全解探索型のアプローチ [6] もあるが、一般的に意味レベルの解析が求められる日本語解析への適用は難しい。また、Tomita パーサを除き、これらの全解型構文パーサは、可能な複数解を得るに留まり、実際に優先解を選択する機構を備えるには至っていない。

### 最良解探索

計算コストの高い全解探索方式をとらず、曖昧さの爆発を避けながら、効率的に有効な解を得るために、ローカルに設定した加算性の優先点をキーにした、最良探索手法を適用したアプローチがある。

係り受け解析をベースに、優先点が付与された係り受け関係候補群の組合せ最適化問題ととらえるアプローチがわかりやすい。この問題では、曖昧性を内在した表現をとることで、解の組合せ爆発を避ける。実際には、探索手法として知られた分岐限定法 [7] やニューラルネットによる方法 [8,9] が報告されている。

規則の適用を最適に制御するアプローチとして、構文規則に付加した優先度から求めた探索費用に基づき、チャート法の解析処理を A\* アルゴリズムで制御して、効率的に最良解を得、さらに次解以降も順次求められる日本語構文解析系 [10] も、最近、報告された。

## 曖昧性解消のための対話

岡島らは、機械翻訳システムの校正機能の考察 [11] で、プレ/ポスト・エディットとともにインタ・エディットを取り上げ、構文解析木をユーザがコマンドで変更するユーザ主導型とパーサが処理中に不明な点が生じた場合にユーザに質問を発するパーサ主導型を挙げた。インタ・エディットは、ユーザの負担の大きいプレ・エディットと同等の効果があり、指定方法が限定されていることからユーザの負担が小さいが、ユーザ主導型では、解析木の画面表示の問題と多くの場合に再翻訳が必要であること、パーサ主導型では、システムの多義解消能力を高め、ユーザへの質問回数を極力減らす必要があることを指摘している。我々は、後に示すようにこの二つのタイプの対話を相補的な組合せとして展開するとともに、具体的に最適性の高い対話方式を提案する。

実際の日英機械翻訳システムへの実装としては、事前に求めた最優先解とそれに近い別解に基づき、曖昧な箇所を指摘し、前編集を促す前編集支援機能 [12] がある。また、解析木に対しユーザの修正指示を受け付け、その修正内容を制約として再解析する機能を付加したものも見られるが、いずれも、商品化された機械翻訳システムの後付け的な機能となっているため、解析過程を重ねる無駄をしている。尚、最近、形態素解析、構文解析、変換、訳文生成の各段階で、システムが絞りきれなかった部分の決定をユーザに委ねる本格的な対話型日英機械翻訳システムの研究が報告されている [13]。

丸山らは、前項に挙げた係り受け候補群の組合せ問題の枠組みで、候補に関する制約をユーザが対話的に適用していく日本語解析ワークベンチを示し [14]、さらに、日英機械翻訳システムでの解析結果のユーザによる対話修正の形態に進展させ、対話実験の結果を示した [15]。

Wood らは、英語モノリンガル向けの機械翻訳 [Import 型 Aidtran (日英) と Export 型 NTRAN (英日)] におけるユーザとのインタラクションの考察を示した [16]。NTRAN では、ユーザが原文を理解できることから、未登録語の範疇、複数解の非共通部分に関する選択、英語の注釈付きの訳語選択といった対話機能をもたせている。

Tomita は、個人用の手軽な少量翻訳の需要を指摘し、その条件として、対話型・高速、Export 型、安価等を挙げ、前記のパーサに、入力と同期したオンラインパーシングや膨大な意味処理のユーザへのバイパスとした曖昧性解消対話の機能を付加し、条件を実装して示した [3]。この対話は二種類の実装があり、ともにノード (句) 間の関係を示す Explanation を基にする。一つは、各展開解析木の Explanation のセット群を、共通 Explanation をキーに分けつつ質問を繰り返すもので、NTRAN での複数解の非共通部分に関する選択と同等の方法である。もう一方は、解析過程で Explanation を Shared ノードのマークを付けて生成し、パース終了後に同一 Shared ノードに関連した Explanation の組を基に質問を繰り返す。この際に、Packed-Shared Forest のまま、不要エッジを削除して解を絞る。特に、後者は、Pack された解析木を展開せずに、最終解析木を得ることができる。

## 3. 対話型日本語解析系への要請

ここでいう対話型日本語解析の主要な点は、現状の不十分な解析能力をユーザが協調的に補完し、解析結果の精度を高めるための曖昧性解消対話である。曖昧な点を全て質問するとか、可能な解を全て提示し選択を求めるという方法もあるが、本来、解析系に全て任せたいところを、それが不十分であるために行う対話である。従って、岡島らが指摘するように、対話は極力減らし、解析系単独での解析能力の向上も期すべきものである。

また、我々の日本語解析系は、さまざまな応用系への適用、多くの利用形態の提供も目指す。そのために、岡島らの示した機械翻訳でのパーサ主導型/ユーザ主導型インタ・エディットを、システム主導対話/ユーザ主導対話と呼び、その双方を備え、解析系単独モードと各対話モードをユーザが柔軟に指定、移行できることを求めた。ユーザが、対話の労力に拘らず、確実に正解を得たい、という要求に応じて起動する対話機構であるが、こうした構成により、ユーザは、利用目的に応じて、精度や労力、方法にある程度コントロールすることができる。

ここで、システム主導対話では、効率的な対話戦略をもち、対話回数や解析結果の予測品質を指定できること、また、ユーザが解析結果を監視 (モニタ) し、必要に応じて解析結果を修正介入できるユーザ主導対話では、ユーザが適切な判断ができるような支援が必要になる。

従来の Ntran や Tomita パーサ等で示された曖昧性解消のための対話は、システム主導対話に属するが、効率的な対話戦略の観点から、また、丸山らの対話型日本語パーサは、ユーザ主導対話に限られており、両主導対話が相補的、有機的な構成をなしているものはない。

尚、ここでいう曖昧性解消のための対話は、質問応答あるいは指令を与える対話であるが、上に述べた多くの利用形態の観点から、対話的に文を入力する形態も想定し、その際の解析系のリアルタイム性の向上が必要と考えた。Tomita は、英語解析で Left-to-Right 処理によるオンラインパーシング [3] を示したが、日本語解析では、形態素解析と構文解析の 2 パスがあり、難しい。

以上をまとめれば、次のようになる。

- ・解析系単独の優先解釈能力も充分高める
- ・システム/ユーザ両主導対話を相補的に組合せる
- ・システム主導対話には効率的な対話戦略をもたせる
- ・対話的な文入力時の即時性を高める

次に、上記の要請を反映した 2 つのアプローチの対話型日本語解析系を示す。前者は、検討を主としたもので、部分的な試作に留まったが、後者では、対話機構を含んだ日本語解析系の試作及び解析実験を行った。

#### 4. 対話型日本語解析系 1

本節では、係り受け関係候補群により曖昧性を内在した表現をベースにした日本語解析系 [7,14] でのシステム主導対話とユーザ主導対話の枠組みの一方を示す。

図 4.1 に、全体の構成を示す。左側の処理が終了した段階で、文中の各文節ノードとスコアの付きの文節ノード間の係り受け関係候補 (図 4.2) が保持される。スコアは、優先解釈のために、最良探索方式 [7] で各候補に与えられたのと同様な加算性の優先点の類である。

尚、本解析系は、PROLOG で部分的試作/実験を行った。

#### 制約適用と優先解釈

解析系単独では、制約適用と優先解釈のループを繰り返す、係り受け関係候補の尤もらしいセット (解) を一つに絞る。制約適用は、同一格排除、非交差等の制約により排除可能な係り受け関係候補を排除し、優先解釈は、曖昧さの残る文節ノードについて、それを係り側とする係り受け関係候補を一つに絞る。

優先解釈では、第 1 位と第 2 位のスコア差が最大の文節ノードを選び、その第 1 位の候補を選択する戦略をとる。これにより、確実性の高い候補から確定し、それに基づく制約適用により、不要な候補を早期に排除する。

#### ユーザ主導対話

ユーザ主導対話は、解が得られた時点で、指定に応じて起動される (図 4.3)。解析結果を解析木等でユーザに提示し、係り先や関係名等の修正指示を受け付ける。その修正指示を絶対的制約として、再び、制約適用、優先解釈を繰り返す、解を求め、最終的には、ユーザによる解の確認で終了する。解析木中の曖昧さをもつ部分にマークを付けたり、ユーザが指示した対象の候補一覧にスコア付けることで、ユーザの負担を減らすことができる。

#### システム主導対話

システム主導対話は、基本的には優先解釈部の代わりに起動され、ある文節ノードの係り受け関係候補をユーザに質問する (図 4.3)。しかし、実際には、全てを対話に頼る場合 (完全対話) を除き、曖昧さ解消について、優先解釈との役割分担や質問対象の順序の制御を行い、ユーザの負担を小さくする対話計画 (対話戦略) をもつ。

ここでは、a) 指定された対話回数または第 1 位と第 2 位のスコア差の最小が閾値以内のうちは、スコア差の小さい文節ノードから順にユーザに質問し、その後、b) スコア差の大きい文節ノードから順にシステムの優先解釈を繰り返す、という対話戦略を提案する。即ち、確実性の低い対象は、先にユーザにより確定した後、確実な対象をシステムが優先解釈する。対話を先に行うことで、早期に正解を落とす可能性を避ける (対話の健全性)。

システム主導対話は、対話回数が制限なしや予測品質が最高等の完全対話の指定をした場合、従来のシステム主導対話と同様に必要な対話を全て行うが、対話計画により確実性が低い曖昧さからユーザにより解決されていくので、途中で中断しても、正解性は高い。

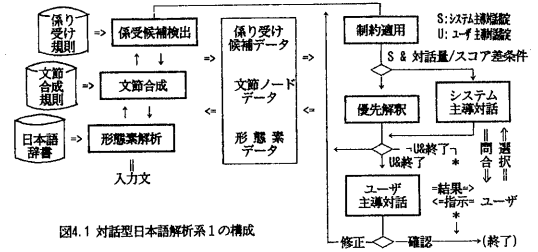


図 4.1 対話型日本語解析系 1 の構成

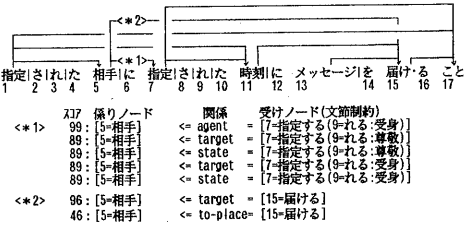
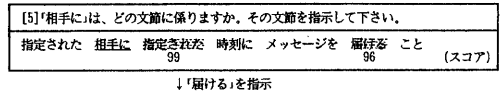


図 4.2 係り受け関係候補群の例

#### システム主導対話の例



#### ユーザ主導対話の例

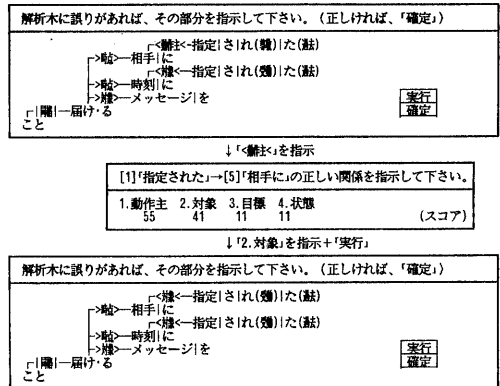


図 4.3 対話型日本語解析系 1 での対話のイメージ

尚、両対話モードを用意するだけでなく、相互間の移行を可能とすることで、さらに柔軟な対話の活用が期待できる (図 4.3)。

#### 対話的入力時の即時処理

図 4.1 の左側の処理では、文入力、形態素解析、文節合成、係り受け候補検出の各処理を Left-to-Right に疑似的に並行処理することで、文入力の時間を有効に利用し、文入力後の待ち時間の減少を図ることを想定した。

この日本語解析系では、差が有意性をもつ(加算性のある)スコアを設定する必要がある。スコアは、係り受けの条件である単語の意味や格パターンの出現確率を適切な重みで線形結合した値として、出現確率は統計的に、重みは実験により更新していくことを検討したが、実際には、係り受け規則中に人間の直感でスコアを決めた。

システム主導対話で、ユーザの指定する対話量の指標は、対話の回数(ユーザの負担が一定以内)あるいは解の予測品質(品質が一定以上)といったものになる。後者では、文の長さに応じた、スコア差の閾値と解の品質との対応付けが必要になる。品質が最高では、閾値が0、'無関心'では、閾値は $\infty$ となるが、その中間での対応は、かなりのシミュレーション実験が必要となろう。

ここでとった、解の構成要素の候補の形で曖昧さを保持しつつ、候補を排除していく方式は、曖昧さの組合せ爆発を回避しつつ、正解に辿りつくという点で巧妙であり、直感的にもわかりやすいモデルと考えられるが、複数解を同時に扱うことを避けているため、対話計画でグローバルな評価の観点から欠ける問題が逃れられない。

たとえば、システム主導対話の対話計画で示した a) では、第1位候補の確実性が高く、その選択に基づく非交差制約で曖昧性を大幅に減少させられる文節ノードを質問対象に選択するという戦略も有効な戦略として検討されるべきだが、全解が得られていないため難しい。

尚、本解析系からスコアとシステム主導対話を除くと、係り受け解析の枠組みでユーザ主導対話を備えた丸山らの対話型日本語パーサ [15] とほぼ同等なものになる。

## 5. 対話型日本語解析系 2

我々は、当初、上記の係り受け解析をベースとした方式の上で、システム主導対話とユーザ主導対話の枠組みの実現の方式を検討したが、より厳密な解析の扱いを目指し、句構造系文法により、複数解を順位付きで得ることができる全解型ベースの日本語解析系を試作し、その上にこれらの対話を組み込むことを試みた。特に、複数解を保持した場合の対話計画は、4で指摘したグローバルな評価の観点が期待できる。

従来、句構造規則だけでは、意味処理の扱いが難しく、また、全解型は、解析状態数の爆発といった問題があった。ここでは、前者に対しては、句構造規則に相補的に単一化による制約記述を導入した LFG [17,18] を、構文レベルと意味レベルを融合して扱うべく拡張した機能制約文法の枠組みを設定し、後者には、全解探索の構文解析法である CKEY法(三角行列法)[19]に SLEEP/AWAKE と呼ぶ機構を導入し、全解性を保持したビーム探索型方式をとることにより問題の解決を図った。

以下に、ベースとなる日本語解析系の概要、さらにその対話機構、対話入力処理、解析実験について述べる。

### 5.1 順位付き複数解出力機能を備えた日本語解析系

試作した日本語解析系の構成を図 5.1 に、また、日本語解析の出力例を図 5.2 に示す。尚、本解析系は、SUN3/60 上の UtilLisp で試作した。

入力は、べた書き日本語文であり、LFGのF構造形式に準じつつ、独自に規定した形式に従い、最尤解以下、順序付けされた複数の尤解を出力できる。

処理系は、ほぼ文節に準じた区間ごとに形態素列を出力する形態素解析部の結果を受けて、構文論レベルの解析を Left-to-Right に行う構文解析部と意味論レベル部分を担う機能解釈部からなる構文意味解析部、及び必要に応じて起動され、残る曖昧性を解消するためにユーザに支援を求める曖昧性解消対話部からなる。また、解が残らなかった場合に起動される AWAKE 処理部と最終的な構造の整形処理部をもつ。構文解析部以降、後処理部まで、対話部も含めて、ペナルティ付きの機能構造と呼ぶデータ構造を一貫して扱う。

言語系の知識としては、構文意味解析辞書、注釈として機能名が付加された句構造規則、機能名に関する制約が定義された機能制約規則、及び意味素性検査に参照される世界モデル(意味素性ネットワーク)がある。

5.2 及び 5.3 に示す曖昧性解消対話機構及び対話入力時の即時処理機構を除くこの日本語解析系の特徴を以下に示す。

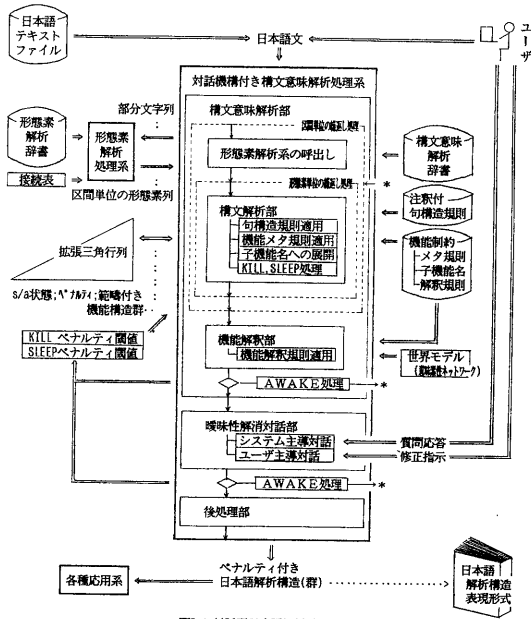


図5.1 対話型日本語解析系2の構成

## 機能制約文法

LFGにおける注釈による制御を拡張した機能制約文法の枠組みを設定する。この枠組みでは、句構造規則に注釈として指定された機能名に対して、構文及び意味的制約とそのペナルティ(加算性の不整合度)を局所的、宣言的に記述する。制約記述は、機能構造中の機能名を構文レベルから意味レベルに深化させることもできる。

## 拡張三角行列法 / SLEEP/AWAKE 機構

構文解析では、三角行列法の全探索方式を、ペナルティに基づく SLEEP/AWAKE 機構や KILL 処理を含む優先解釈により拡張し、全解性を保持したビーム探索を行う。SLEEP/AWAKE 機構は、解析途中状態のペナルティが高くなると、sleep 状態として凍結し、その後、解がなくなると、ペナルティの sleep 閾値を上げて AWAKE 処理を施し、途中状態から解析を再開する。これにより、解析処理の効率化と全解性、頑強性を両立させる。頑強性は、従来、排除されていた非文性の高い現象に高いペナルティを与えることで対応する。また、形態素解析の多義も扱うため、拡張三角行列の単位は、単語ではなく、形態素を連続した並びで表現可能な分割領域になる。

## 意味素性ネットワーク(世界モデル)

意味レベル素性検査のために、ペナルティ付きリンクが張られた意味素性ネットワークを備える。このネットワークでは、間接的な関係のある意味素性に対しても、ペナルティが与えられ、柔軟な意味レベルの解析に寄与する。意味素性ネットワークは、意味素性間の直感的な階層木や個別の関係の定義群に基づき生成される。

## 日本語解析構造表現形式

応用性を考慮した解析結果の出力形式として、機能構造に準じつつ、表層格に深層格を階層的に付加した関係名(機能名)や、コントロール現象や連体修飾節の痕跡を示すポインタ的表現を導入した形式を規定する(図 5.2)。

## 5.2 曖昧性解消のための対話機構

最終的に残った機能構造(解)のうち、ペナルティ最小を最尤解とするが、ペナルティで順序付けられた複数解に対して、曖昧性を解消する対話機構を導入した。

対話としては、前述の、対話計画に基づきシステム側からユーザに効率的に問合せを行うシステム主導対話とユーザ側から解析結果の修正介入を行うユーザ主導対話を用意し、ユーザの利用形態の選択に柔軟性をもたせた。

いずれの対話も、ペナルティ付きの機能構造の枠組みの中で扱い、ユーザの指定は、制約として扱い、その制約を満たさない機能構造に高いペナルティを加算し、必要に応じて KILL 処理を施す。対話の進行によって、解がなくなってしまう場合は、AWAKE 処理に入る。

## システム主導対話

システム主導対話の対話計画には、(平均)情報量最大の質問により、ユーザから最大の情報量を得る、即ち、知識の曖昧さを最も減らす、という最適化戦略を設定した[20]。

質問の対象は、各形態素ごとの修飾先、機能名、語義等(パラメータ)で、形態素  $W_i$  のパラメータ  $X$  [ $X$ : 修飾先、機能名、語義、 $\cdot$ ] についての情報量  $I(W_i, X)$  は、その各候補  $X_j$  の確率  $Prob(W_i, X_j)$  により、次式から得る。

$$I(W_i, X) = - \sum_j Prob(W_i, X_j) \cdot \log(Prob(W_i, X_j))$$

候補の確率  $Prob(W_i, X_j)$  は、あらかじめ各機能構造の確率(積算性)  $Prob$  をペナルティ(加算性)  $Pen$  の負のべき乗化  $K^{-Pen}$  によって変換しておいた上で、各候補を含む機能構造の確率の和と適当な正規化によって与える。最大の情報量  $I(W_i, X)$  を与える形態素  $W_i$  の種別  $X$  が、質問の対象となり、 $X_j$  が候補一覧を与える。

ユーザが質問の候補を選択すると、それを制約として、解が一つになるまでシステム主導対話を繰り返す(図 5.1)が、対話を中止したり、ユーザ主導対話に移ることも許す。その場合は、その段階で残った機能構造が、そのペナルティで順序付けられる。尚、候補一覧を確率順に提示することで、ユーザの選択が容易になる。

この方式は、係り受け方式で欠けていたその候補を含む解の確実性(確率)の考慮が加わっている。また、修飾先(係り先)、機能名(関係名)、語義なども情報量という一つの指針で同時に扱い、対話の有効性を高めている。

## ユーザ主導対話

ユーザ主導対話は、解析結果に対する制限された編集機能にあたる。編集対象は、システム主導対話と同様に、形態素ごとの修飾先や機能名等(種別)であり、ユーザが、形態素と種別及びその候補を指定すると、それを制約として、解を絞っていく。

尚、ペナルティ順に表示された機能構造一覧から解を確定する機能や、形態素と種別の指定により、その候補の一覧を確率順に示す支援機能をもつ。

## 5.3 対話入力時の即時処理機構

解析処理は、入力と同期して、区間ごとに形態素解析と構文解析の Left-to-Right 処理を逐次的に繰り返す疑似的な並行処理を行い、対話的な文入力形態でのユーザの待ち時間の減少を図った。これは、形態素解析系に、1文字ずつの入力と区間ごとの即時出力機能を加えた上で、構文解析側が、区間単位の結果を受けるまで、形態素解析系に対話入力文字を順次与えることで実現した。

また、形態素解析系は、当初の総当たり方式を、動的計画法により高速化を図り、区切り位置の相違まで含めた区間内の第N解までの多義出力機能も加えた。尚、区間の分割は字種情報による[21]。

私は昨日日本で買った小さな本を読みながら彼女の家へ向かった。

```

0 私 time 0.0sec
1 は time 0.0sec
  :
9 を time 0.5sec
10 読み time 0.8sec
  :
14 家 time 1.2sec
15 へ time 1.2sec
16 向かつ time 1.8sec
17 た time 2.6sec
** time 4.8sec

```

可能な解は 27 通りあります

文節「昨日」の係り先を教えてください

```

1 向かった
2 読みながら
3 買った
0 その他
どれですか？ 3

```

PENALTY = 25

```

[sem 向かった
@は#topic [sem 私 ]
@が#obj ./@は#topic
@mod#<時制>[sem 昨日 ]
@ながら#co-occur
[sem 読む
@を#obj [sem 本
@<連体形>#mod [sem 買う
@-を#obj ./
@で#place [sem 本屋 ]
tense
た#past ]
]
@#mod [sem 小さな ] ]
@へ#goal [sem 家
@の#mod [sem 彼女 ]
tense
た#past ]

```

図 5.2 対話型日本語解析系 2 の解析例

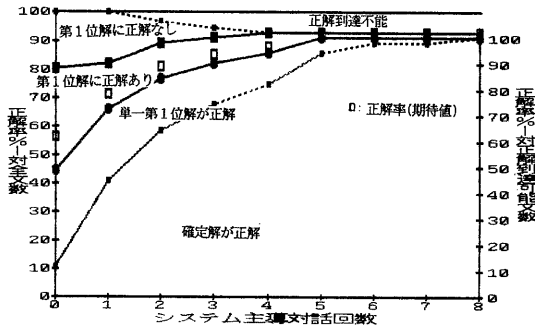


図 5.3 システム主導対話を含む解析実験

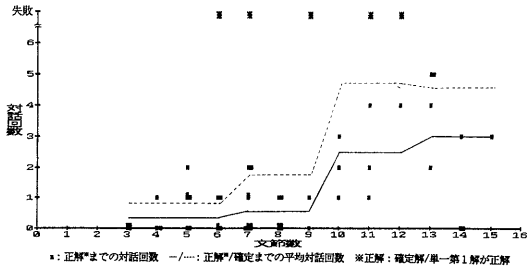


図 5.4 正解までのシステム主導対話回数

#### 5.4 解析実験

上記の機能制約文法の枠組み上で、基本的現象と実文約 50 文に基づいて、注釈付き構造成規則約 170、機能約 100 についての機能制約の実験規則を書き下し、それを用いて、SUN3/60(8MB メモリ)上で解析実験を行った。

図 5.3, 5.4 に、3~15 文節の別の実文 56 文に対するシステム主導対話を含む解析実験の結果を示す。

対話なしで、正解率(期待値)/同点を含む第 1 位解の正解率([ ] 内は、正解到達可能な文に対する正解率)は 56.4%[60.4%]/80.4%[86.4%]で、3 回までのシステム主導対話により 85.4%[91.9%]/91.1%[98.1%]まで上がった(図 5.3)。解析系単独での正解率は、期待値では充分ではないが、同点を含む第 1 位解では、高い値を示している。対話の効果として、30/10%程度の高い寄与を確認できる。

文長に対しては、3-6/7-9 文節で、平均 0.3/0.6 回程度の対話で正解に達するが、10-12/13-15 文節では、平均 2.5-3.0 回程度を必要とする(図 5.4)。丸山らのユーザ主導対話の平均回数(4-6/7-9/10-12/13-15 文節に対し 0.0-0.1/0.5-0.7/1.5-2.1/2.1-2.8 回)[15]を若干越える値だが、ユーザ主導対話では、解全体あるいは複数箇所注意到注意を払う必要があることを考えると、長文では、システム主導対話の方が、ユーザの負担は小さい。また、文長に応じ、対話回数-正解率を大量のデータに基づいて得れば、予測品質による指標を設定することが期待できる。

表 5.1 に、27 形態素 13 文節の文での、情報量最大及び別の戦略での対話ごとの解の数の変化を示す。解確定までの対話回数は、1 回程しか違わないが、曖昧さを最も減らす情報量最大の戦略が、最も平均的に解を減らしている。一方、文末からの戦略では、最初に第 1 位解も全解の数も一挙に減らしているが、ユーザの選択によっては、逆にほとんど減らない、という結果になる危険性を含む。また、解の数が 10 程度になると、戦略によらず収束はよくない。これらは、本システム主導対話は、解の確定まで行う完全対話より、限られた対話回数で効果的に解を減らす利用法での有効性を示すものになっている。

CPU 処理時間は、図 5.2 の 18 形態素 10 文節の文で 4.8 秒と、全解型解析系としてはある程度速度を確保している。一方、入力との並行解析による待ち時間の節約(最終より 1 区間前までの時間)は、本例では 1.2 秒で、全体の 4.8 秒に対しては、充分な節約にはならなかった。

表 5.1 対話戦略と対話回数ごとの質問対象文節/解の数の減少例

対象文: しかし植物を提供した試験研究機関によって、性質の分類の方法が異なるため比較することが困難だった。

対話戦略\回数	0	1	2	3	4	5	6
情報量最大		しかし	よって	提供	植物	ため	性質
	5/69	5/25	3/10	3/6	2/4	1/2	1/1
文末文節から		ため	性質	よって	提供	しかし	
	5/69	1/14	1/11	1/7	1/5	1/1	
文頭文節から		しかし	植物	提供	よって	性質	ため
	5/69	5/25	3/17	3/11	2/4	1/2	1/1
第 1 位解優先		植物	よって	ため	しかし	提供	性質
	5/69	3/54	2/24	1/10	1/3	1/2	1/1

上段: 質問対象文節/下段: 第 1 位解の数/全解の数

## 6. まとめ

日本語解析系単独での高い解析能力を求めると同時に、曖昧性解消能力を補完するための対話機構—システム主導対話とユーザ主導対話—の検討と実装を、(1) 係り受け関係候補群により曖昧さを内在する表現の上、及び(2) 機能制約文法の枠組みで、構文意味レベルの複数の尤解を順位付きで高速に得る日本語解析系において示した。

(1) では、曖昧さを内在する表現の上で扱うことで、組合せの爆発を避け、文節ノードに関する候補の優先選択と制約適用の繰り返して扱った。対象文節の選択は、システムが行う優先解釈では、確実性の高い候補をもつノードから、ユーザへの問合せは、確実性の低いノードから、という戦略により正解に至る可能性を高めた。

(2) では、全探索ベースながら、SLEEP/AWAKE 機構により効果的に全解性を保持した日本語解析系の上に、構文/意味解析と同じ制約の枠組みで統一的にとらえた対話機構を実装した。対話は、ペナルティ付き複数解から、ペナルティの確率化を通じて情報量の計算を行い、情報量最大の対象について問合せることで、グローバルな視野をもつ対話戦略を示した。その解析実験では、3回までのシステム主導対話で、かなりの解析率の向上を確認し、さらに、文長ごとの予測品質に対する対話回数を得る見込みを得た。また、本システム主導対話では、少ない対話回数で効果的に解を減らす利用法が有効であるとの認識を得た。

上記の2つのアプローチは、原文の再解析を求めるものではなく、高い優先解釈能力をもつ解析処理部と密接に連携した対話機構として位置付けられる。

従来の対話に比しては、これまで見られなかったシステム主導対話での具体的な対話戦略を提案し、また、ユーザ主導対話では、候補一覧にシステムのもつスコアや確率を付加して、ユーザの判断を助ける指針も示した。

また、ここで提案するシステム主導対話とユーザ主導対話の組合せの枠組みは、具体的な利用実験は行っていないが、解析系単独の解析能力と対象文の難易度の関係、要求品質等の状況に応じて、柔軟な利用形態を提供できると考える。即ち、長文が多かったり、高品質を求める場合は、システム主導対話を利用し、解析系単独でも充分解析できそうなら、ユーザ主導対話を利用すればよいし、バッチの処理でもよい。また、対話の選択や対話回数/予測品質の指示を、各契機に変更可能にすれば、ユーザ主導対話の結果を見て、必要に応じシステム主導対話に移行する等、さらに柔軟性を高めることができる。

ここで優先点として導入したスコア、ペナルティは、実際には、高い精度で与えることはできず、これに基づく対話計画の厳密性はない。しかし、この優先点に基づく解析系単独での解析率から見て、その有効性は認められ、その観点から、対話計画の第0次近似の指針としては有効であると考えられる。一方、仮に構文意味レベルでかなり精度の優先点を設定できたとしても、文脈等の寄与の反映は困難であり、この点からは、むしろ、ユーザとの対話は、文脈の知識源として欠かせないものになる。

対話的な文入力の利用形態での、入力終了後の待ち時間の軽減のために、Left-to-Right 処理による文入力と並行した解析処理を実装したが、十分な効果がなかった。これには、日本語解析で効果的と考えられる文末からの処理 [22] との比較による判断が必要であろう。

今後は、日本語解析系としての性能向上とともに、対話計画のモデルの改善が求められる。また、対話でのユーザインターフェースが課題となっている。

### 謝辞

この対話型日本語解析系の研究開発にあたり、その最初の契機を与えて頂いた山内佐敏室長、全般にわたり多くの示唆・助言を与えて頂いた藤田克彦室長、対話型日本語解析系2で利用した形態素解析処理系について、その初版を開発した今郷昭研究員、同辞書を作成した望主雅子研究員、同即時処理機構等の機能追加・改良を行った伊藤篤研究員に感謝する。

### 参考文献

- [1] 内田裕士, 増山顕成, 小部正人: "意味解析向き自然言語パーサ", 情報処理学会 自然言語処理研究会報告 34-3, 1982
- [2] 梶博行, 岡島惇: "日立における機械翻訳システム", 情報処理 Vol.26 No.10, pp1214-1216, 1985
- [3] Tomita, M.: "An Efficient Context-free Parsing Algorithm for Natural Languages and Its Applications", PhD thesis, Computer Science Department, Carnegie-Mellon University, 1985
- [4] 沼崎浩明, 田村直良, 田中穂積: "並列論理型言語による一般化LR構文解析アルゴリズムの実現", 情報処理学会 自然言語処理研究会報告 74-6, 1989
- [5] 林達也, 宮俊司: "横型トップダウン解析システムの実現と評価", 情報処理学会 自然言語処理研究会報告 74-9, 1989
- [6] 水田昌明: "単一化文法の効率的な解析手法", 情報処理学会第42回全国大会 5C-8, 1991
- [7] 平川秀樹, 天野真家: "日本語解析における最適解探索", 情報処理学会 自然言語処理研究会報告 74-2, 1989
- [8] 高橋直人, 板橋秀一: "ニューラルネットによる日本語形態素・係り受け解析", 情報処理学会 自然言語処理研究会報告 80-1, 1990
- [9] 亀田雅之: "組合せ最適化方式", 公開特許公報 平2-297138, 1990
- [10] 吉見毅彦, 西田収: "最良優先型構文解析系と日英機械翻訳システムへの応用", シェア技法 第47号 pp64-68, 1990
- [11] 岡島惇, 新田義彦, 山野文行: "機械翻訳における校正機能", 情報処理学会 自然言語処理研究会報告 40-5, 1983
- [12] 平井章博, 梶博行, 声沢実: "機械翻訳向け前編集のための日本語係り受け構造の曖昧性検出方式", 情報処理学会論文誌 Vol.31 No.19, pp1425-1437, 1990
- [13] 青山昇一, 石川雅彦, 柿ヶ原康二, 加藤達哉, 杉村領一: "対話翻訳の方式について", 電子情報通信学会 言語とコミュニケーション研究会報告 90-14, 1990
- [14] 渡辺日出雄, 丸山宏: "制約依存文法に基づいた日本語解析支援システム", 情報処理学会 自然言語処理研究会報告 69-6, 1988
- [15] Maruyama, H., Watanabe, H. & Ogino, S.: "An Interactive Japanese Parser for Machine Translation", Coling '90, 1990
- [16] Wood, M.M. & Chandler, B.J.: "Machine Translation for Monolinguals", Coling '88, 1988
- [17] Bresnan, J. (ed): "The Mental Representation of Grammatical Relations", The MIT Press, 1982
- [18] 新田善久: "LFGと意味解析の融合に向けて", 情報処理学会 自然言語処理研究会報告 68-2, 1988
- [19] Aho, A.V., Ullman, J.D.: "The Theory of Parsing, Translation, and Compiling Vol.1: Parsing", p.315, Prentice-Hall Inc., 1972
- [20] 宮川洋, 原島博, 今井秀樹: 岩波講座情報科学4 "情報と符号の理論", pp26-43, 1982
- [21] 野村典正, 森健一: "漢字かな変換システムの試作", 電子通信学会論文誌 Vol.J66-D No.7 pp789-795, 1983
- [22] 藤田克彦: "決定的係り受け解析に関する試み", 昭和63年度人工知能学会全国大会(第2回), pp399-401, 1988