

5C-4 並列自然言語解析システムLaPutaの試作評価

山崎重一郎, 末広香緒里, 長沢郁子, 杉山健司
富士通研究所

1 はじめに

我々は第5世代コンピュータプロジェクトの一環として並列自然言語解析システムLaPutaの研究開発を進めている。第5世代コンピュータプロジェクトの最終目標は並列推論マシンを作ることであり、LaPutaは並列推論マシンでの実行を目的として研究開発を進めている。

本稿では、LaPutaの処理モデル、原理およびマルチブサイ上への実装の実験結果について述べる。

2 研究の目的

我々の研究の目的は、自然言語解析を並列処理の観点から再考し、より自然で効率的な自然言語解析処理モデルを提案することと、その高速性の実証である。

3 並列処理モデル

自然言語解析の並列処理のための処理モデルとして、並列協調と負荷分散の二つの観点からモデルを考えた。

並列協調のモデルは、自然言語処理の全てのレベルを並列的に協調させるものであり、負荷分散のモデルは、優先度制御を候補どうしのプロセッサ獲得競争によって実現する処理モデルである。この二つを組み合わせたものを我々は競争と協調のモデルと呼んでいる。

3.1 自然言語処理レベル間の並列協調

自然言語解析の各レベル、すなわち形態素、構文、意味、文脈などの処理では、それぞれ曖昧性が発生するが、こういった曖昧性の解消に他のレベルからの情報を必要とすることがしばしばある。このため、並列処理の観点から見るとこれらのレベルは並列的に協調させるのが自然である。LaPutaは、自然言語処理の全てのレベルの間の並列協調を実現している。これは、並列解析機構が処理レベルを越えて情報を交換することによって実現している。

3.2 プロセッサ獲得競争による優先度処理

並列処理においても処理効率を高めるために制御は必要である。しかし、自然言語処理のような大規模な並列処理システムでは人間が制御を完全に把握することは不可能な

ので自動的に制御が決定される仕組みが必要になる。我々はそのための仕組みとして、規則の優先度と並列処理の制御を組み合わせることを考えた。この実現手段として、規則どうしが自分の処理に必要なプロセッサをめぐる競争を行う処理モデルを検討している。

4 原理

LaPutaの並列解析機構は型推論と型付き素性構造の単一化に基づいている。

4.1 型推論

型推論の役割はある型宣言のもとで与えられた対象の型に関する判断を検査し、その証明を与えることである。

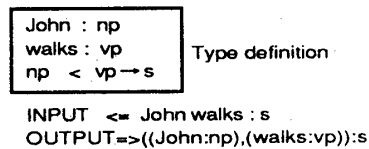


図1 型推論の例

システムの計算結果は、入力 of 型判断に対する証明木であり、これは構文レベルでは構文解析木に対応している。

4.2 並列型推論機構の動作

並列型推論機構はプロセスが通信ストリームを介してデータ交換することによって動作する。

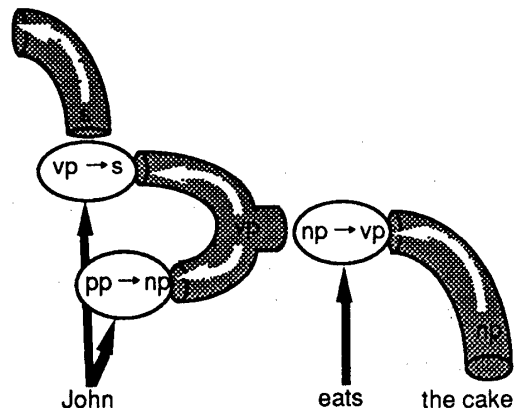


図2 並列型推論機構の動作

4.3 単一化

LaPutaの文法記述では、素性の値に型を与えることができる。このようにして与えられた型は素性の値に対象が単一化されるときに検査される。意味処理は、構文解析規則の素性の値に対する意味範疇の検査を型検査というかたちで実行され、これによって意味処理は構文解析と連携する。この検査に使用される機構は構文解析の例で示した型推論機構と同一のものである。

素性の値に与えられた型にはさらに制約を付与することができる。例えば、食べるという動詞の行為者という素性の値は単に主格の名詞句の対象だけでなく、目的格の名詞句の対象にも影響されて適切さが変化するような現象が取り扱い可能となる。

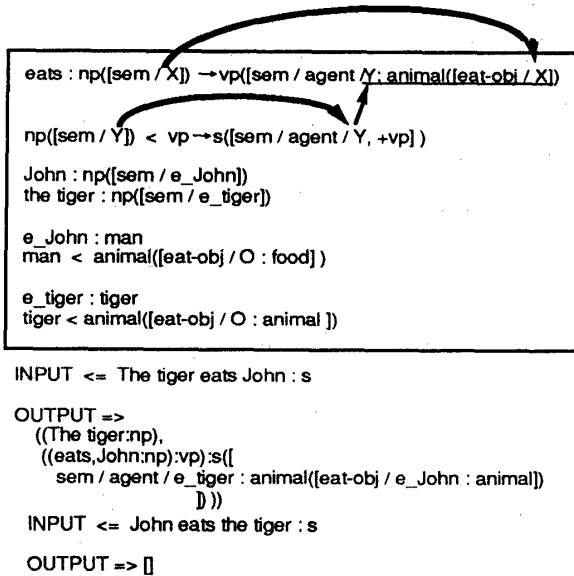


図3 目的語に応じた主語の適格性の変化

5 実験

5.1 実験に用いた文法

測定に用いた日本語構文解析文法は、ICOTで開発されたもので、文法の規模は約300規則である。この文法は本格的な日本語処理を目指しているものであり、解析結果は句構造木だけでなく複雑な素性構造も構成するものである。

5.2 実験の経過

実験は並列解析機構の改良と平行して進めた。並列解析機構の改良は絶対速度の向上を目標としたので、まず逐次実行速度の改善を行い、10倍以上の性能改善が達成できた。しかし、この改良は台数効果を減らす結果となった。

5.3 実験結果

台数効果の実験は逐次実行速度が最も早い版で行った。

セッサ台数を1台から32台に増加させた時の速度の向上率は4.5倍であった。

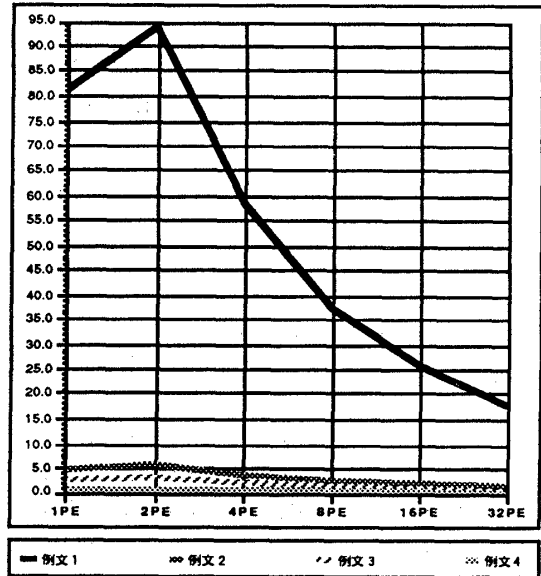


図4 プロセッサ台数と解析時間

	形態素数	解の個数	1PE:32PE
例文1	21	161	4.5
例文2	10	7	3.1
例文3	13	7	2.2
例文4	4	1	1.5

6 今後の予定

今後さらに並列解析機構と並列解析文法の絶対解析速度の向上を目指した改良を進める。

また、並列協調モデルを実証するために形態素、構文、意味の並列協調の実験の準備を進めている。この実験では競争モデルの実装も進める。

謝辞

この研究はICOTからの再委託研究として行った。研究を進める上で、ICOT第6研究室の田中裕一室長をはじめ佐野洋氏、赤坂宏二氏、萩原馨氏、佐藤健氏などに貴重な助言をいただいた。特に佐野氏には文法を提供していただいた。ここに記して感謝いたします。

[参考文献]

[寿崎86] 寿崎かすみ他、マルチPSIにおける並列構文解析システムPAXの実現および評価、並列処理シンポジウムJSP'89予稿集、情報処理学会、Feb.89

[佐藤90] 佐藤裕幸、並列自然言語構文解析システムPAXの改良、KL1プログラミングワークショップ'90予稿集、1990

[松本90] 並列論理型言語における探索問題のプログラミング-Layered Stream法の拡張、KL1プログラミングワークショップ'90予稿集、1990