並列推論マシン上の自然言語解析システムについて

4F－5

山崎重一郎，杉山健司，鈴木香緒里，玉田郁子
富士通

1．はじめに

現在，第5世代コンピュータプロジェクトの一環として並列推論マシンのための自然言語解析システムの研究開発を行っている．その第1歩として並列処理の観点から見て自然と思われる自然言語処理のモデルを考え，そのモデルに基づいてマルチＰＳＩ上にプロトタイプシステムLaPutaを構築した．本稿では，LaPutaの処理モデル，解析機構の基本原理及び実装方式について述べる．

2．並列推論マシンの特性と問題点

並列推論マシンを利用する意義には，大きく2つの考え方があるように思われる．その一つは処理の高速性でありもう一つは人間の認知過程に内在する並列性への探究である．我々は後者についても興味を持っているが，現在は，高速性の追求に焦点を絞って研究を進めている．

しかし，並列推論マシンにシステムをのせるだけで本当に処理速度が劇的に向上するのだろうか．並列構文解析システムＰＡＸをマルチＰＳＩ上に実装した実験では，６４台のプロセッサを使っても1台のプロセッサのときと比較して最高で僅か2～3倍しか速くならなかった．［1］

並列推論マシンは，個々のプロセッサエレメントが非常に強力な処理能力を持つ反面，プロセッサ間の通信には大きなコストを必要とする．それに対して，文脈自由文法の構文解析のように，個々の処理は単純であるにもかかわらず，いろいろな場所で発生した情報が互いに関連を持つような問題は，負荷の分散と通信の節約が両立しにくい．

3．並列自然言語解析の処理モデル

2で述べたような問題点は，文脈自由文法の構文解析という処理対象がたまたま並列推論マシンに不利な処理対象であったというだけで，自然言語解析全体からみると文脈自由の構文解析はごく一部の処理でしかない．並列推論マシンを有効に利用するには処理対象をより大きく捉えた方が有利である．構文解析や形態素解析のような個々の処理レベルにこだわらずに自然言語解析全体を一つの処理対象とすれば，通信コストとプロセッサの負荷のバランスがとりやすくなり，より並列推論マシンに適したものになる．

しかし，このような処理モデルを考えると別の問題が生じる．典型的な自然言語解析システムでは，形態素解析→構文意味解析→文脈処理，というように処理が逐次的に進行していくが，我々が提案する処理モデルでは自然言語処理システム全体を並列的に協調させたい．しかし，この処理モデルを実現しようとして，形態素解析モジュールや構文解析モジュールなどの自然言語解析の全てのモジュールを処理過程で相互に並列的に協調させようとすると，モジュール同士の入力と出力の関係が組み合わせ的に増加するためにモジュールの構造が爆発的に複雑化する．

この問題を解決するために，我々は，形態素解析や構文解析といった処理レベルの色付けの無い汎用的な処理機構を中心とした自然言語解析システムを構築することを考えた．

4．並列解析機構の基本原理

自然言語解析における処理のレベルに依存しない，汎用的な処理機構の基本原理として型推論と部分項の単一化による制約解消機構を利用することにした．

4．1　型推論系

型推論系はものに対して型を割り当てる推論系［3,4］である．文字列をものとするとある文字列が形態素と判定されるということは，その文字列の型が形態素であることと見ることができる．同様に形態素の列の型が語であり，語の列の型が句や文である．また，句構造規則のような文法規則も関数型とみることができる．意味的情報や文脈的情報についても型理論は利用可能である．このように，型理論を利用することによって自然言語解析処理全体を一つの枠組みで見ることができる．

4．2　制約解消系

制約とは，変数の具体値に関する条件を式で表現したものであり，次のような記号法を使用することにする．

X << ［P(X,a),...,P(X,z)］

ここでＰは，制約対象の変数Ｘの具体値を計算する手段が存在するような述語定数である．我々の制約解消系はＰとして部分項の単一化［2］を利用している．

A Natural Language Processing System for Parallel Inference Machine
Shigeichiro YAMASAKI ,Kaori SUZUKI ,Ikuko TAMADA ,Kenji SUGIYAMA
Fujitsu,Ltd.

4．3　型推論系と制約解消系の融合

型推論系と制約解消系の融合は，2階の型と制約付き変数を同一視することによって実現される．

・2 階の型の例

s：{X ｜X!subj＝NP，　X＝VP}

この表記は，sという型が，集合の抽象化の表記で表された型の集合（型）すなわち2 階の型をもつことを表している．さらに，この表記に現れる束縛変数Xを制約の対象とするためにグローバル変数にしたものを，sという型と意図的に混同して，｛s｝と表記する．

walk:｛np｝→｛s｝<<［｛s｝！subj＝｛np｝］

上の記述は，walkという表現はnpが左にあるとsになるという関数の型を持ち，このとき型sの変数｛s｝の具体値のsubjという素性の値は変数｛np｝の具体値と等しいという制約が付けられていることを意味している．

5．　並列自然言語解析システムLaPutaの実装方法

マルチＰＳＩ上に実装したプロトタイプシステムLaPutaの実装方法は次のような特徴を持つ．

(1)　レイヤードストリーム法よりも通信量が少ない

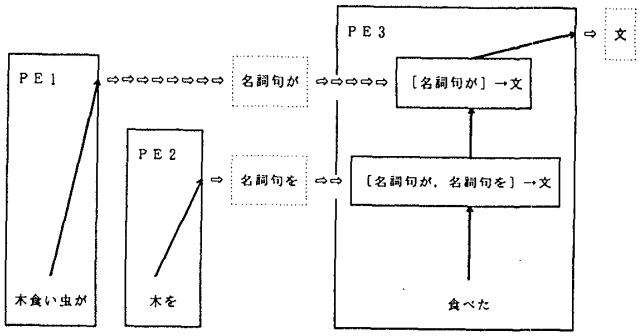
(2)　情報をグローバルに伝播させることができる

LaPuta の並列処理方式は，基本的にはレイヤードストリーム法［5］をもとに考案したものであるが，レイヤードストリーム法では，ストリームの構造で探索空間を表現していたのに対して，プロセスの親子関係で探索空間を表現するものである．この方式はストリームに流すデーターとして探索する候補を流すのではなく，候補が充足されて実際に構成が成功した部分解のみを候補プロセスに同報することによってプロセス間の通信を減らすものである．

もうひとつの特徴は，制約解消系が制約付きのグローバル変数を持つので情報をグローバルに伝播させることができるということである．例えば，談話の環境や文脈などのようなグローバルな情報を，あらゆるレベルの処理がその処理過程で参照・追加ができるということである．

LaPuta の基本動作は大きく分けると二つの動きから構成されている．その第1 のものは，関数型の簡約化を実現するものであり，プロセスから次のプロセスを生成する．また，関数型の簡約化のときに制約条件のマージと制約解決処理が行われる．

第2 のものは関数型の簡約化の結果，部分解が完成したときの処理で，この部分解を次の候補に送信する処理を行う．



PEi：processor element

6．今後の課題

今後，並列解析機構の性能評価を行っていくとともに，効率的な文法記述法についても研究を進め，特に処理レベルを越えた情報の利用が解析速度に与える影響などについて評価を行い，このアプローチの妥当性を検証していきたい．

謝辞

本研究はＩＣＯＴからの受託テーマ「並列処理に基づく自然言語解析ツール」の一環としておこなわれ，ＩＣＯＴ第2研究室の内田室長，吉岡室長代理をはじめとする方々に御支援を頂きました．ここに印して感謝いたします．

【参考文献】

［1］寿崎，佐藤，杉村，赤坂，瀧，山崎，弘田：マルチPSI における並列構文解析プログラムPAX の実現および評価(1988)

［2］向井国昭：半群作用を持つマージ構造とその上の単一化理論(1989)

［3］Reynolds,J.C:Three approaches to type structur. Lecture Notes in Computer Science 185 Mathematical Foundations of Software Development:Springer Verlag (1985)

［4］Hindley,J.R;Seldin,J.P: Intriduction to Combinators and λ-Calculus. Cambridge university press(1986)

［5］松本裕治：並列構文解析．自然言語処理53―2(1986)

［6］橋田浩一，杉村領一，田中裕一：談話理解システムDUALS．情報処理 VOL.30 No.10(1989)