

並列オブジェクト指向言語 A-NETL による 英文解釈プログラム記述

4F-4

菅原 幸子 馬場 敬信 吉永 努
(宇都宮大学 工学部)

1. はじめに

我々の研究室では、並列オブジェクト指向マシンの実現を目指して、応用分野、プログラミング言語、マシンを総合的に捉えたトータルアーキテクチャ A-NETL の研究を行っている。本研究は、A-NETL の応用分野の一つとして並列オブジェクト指向言語 A-NETL を用いた英文解釈処理の記述とそのシミュレーションを行ったものである^[1]。

A-NETL は、①並列に実行されるオブジェクト単位のプログラム構成、②多数オブジェクトの静的定義、動的生成・消去、③未来型メッセージによる同期、④メッセージの並列送信、複数受信などの特徴を持つ^[2]。

プログラムの記述に当たっては、A-NETL の特徴を生かして処理の並列性を高めることに重点を置いている。

2. 並列オブジェクト指向英文解析

自然言語解析の処理を構文解析と意味解析とに分け、それぞれの解析を並列に行う。

2.1 構文解析

文を構成する単語の品詞に着目した文脈自由文法による文法規則を考える。例えば、規則

$$PP \rightarrow PREP \ N \quad (1)$$

は、前置詞句 PP (Preposition Phrase) が前置詞 PREP (Preposition) と名詞 N (Noun) からなることを示す。

文: 'I want to go to Sendai.' (a)

に規則(1)を適用すると、'to' が前置詞であり、'Sendai' が名詞であるから、(a)の1部分 'to Sendai' はまとめて前置詞句であると見なすことができる。これによって構成される解析木(1)を図1-1に示す。

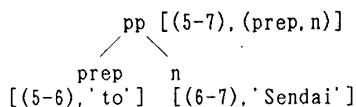


図1-1 解析木(1)

まず前置詞句、前置詞、名詞を扱う機能単位(以下オブジェクト)としてそれぞれ PP, PREP, N を定義し、一般に構文解析オブジェクトと呼ぶ。また、単語 'xxx' に対して辞書オブジェクト xxx を定義する。辞書オブジェクトは、自分の単語の品詞を知っていて、文中でどの位置に出現しているかという情報を与えられると、その品詞の単語を扱う構文解析オブジェクトに位置と単語を送る。

PREP は、to から「5番目、'to'」という情報を受け取ると解析木の節 prep を生成する。これに「位置(5-6)、単語'to'」を持たせ、prep を N に送る。N は Sendai から「6番目、'Sendai'」という情報を受け取って節 n を生成し、かつ PREP から節 prep を受け取ると、位置情報からそれらの単語が連続して出現していることを確認して、PP に位置(5-7)と節(pre, n)を送る。PP は節 pp を生成し、[位置(5-7), 下位節(pre, n)]を持たせることで解析木(1)を構成する。同様に、文法規則

1)を構成する。同様に、文法規則

$$VP \rightarrow V \ PP \quad (2)$$

(VP: 動詞句; V: 動詞)

によって解析木が生成される。この時、規則(1)で生成される節 pp が右辺の PP に送られて、図1-2のようになる。

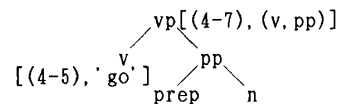


図1-2 解析木(2)

一方、文(a)の 'I' に対して、規則

$$NP \rightarrow PRON \quad (3)$$

(NP: 名詞句; PRON: 代名詞)

によって解析木が生成されるが、これと解析木(1), (2)の生成とは独立であり、並列に行われる。

このようにして規則

$$S \rightarrow NP \ VP \quad (4)$$

(S: 文)

により最終的に文を1つの解析木に変換する。但し、規則(4)のオブジェクト S は、予め文の構成単語数が与えられていて、これと、送られてくる位置情報によって得られる単語数とが一致したときに節 s を生成し、一致しなければ何もしない。

辞書オブジェクト中の単語はそれぞれユニークで、1文中に同一単語が複数回使用された場合はその位置毎に解析オブジェクトに情報が送られる。文(a)で to は PREP に「3番目、単語'to'」と「5番目、単語'to'」の2つの情報を別々に送る。

また、複数の品詞を持つ単語は各々の品詞を扱うオブジェクトに同時に情報を送る。更に、規則(1)及び規則

$$NP \rightarrow N \quad (5)$$

$$NP \rightarrow N \ NP \quad (6)$$

における N のように、複数の規則の右辺に同じ品詞を扱う構文解析オブジェクトがあれば、複数の解析木が生成される可能性がある。

このような複数の解析木は相互関係を持たないことから、処理を並列に行うことが可能である^[3]。

2.2 意味処理

ここで扱う意味は、言葉が実際に表す対象のもつ概念を含む、構文解析では対処しきれない言語の要素である^[4]。

2.1の構文解析において、1つの規則に対して複数の下位節を持つ解析木が構成されるとき、その時点で根となっている節は下位節の組合せが意味的に合うかどうか調べる。意味が通ればそのまま解析を続行し、そうでなければ自身を消去する。このため、その節を用いる以降の解析が止まる。このようにして、意味の合わない単語同士の組合せを早期に除去し、余分な解析を避けている。

例えば、

句: 'on January 3' (b)

に規則(1)が適用されると、'on January'の部分が前置詞句として pp を根とする解析木が構成される。しかし 'January' そのものに対する前置詞は 'on' ではないので、pp を除去する。

このような処理を実現するため、次の様な方法を考える。

各々の単語や解析木の節に、意味を表す標識オブジェクトを持たせる。具体的には、January に MONTH, 3 に DAY, NUMBER へのオブジェクトポインタを持たせておく。また、各標識は自身に関連する標識を、即ち MONTH, DAY はそれぞれ DATE を、NUMBER は THING, PERSON を持っている。規則(5), (6)によって 'January 3' は名詞句となる。このとき左辺の NP が生成する節 np は、n, np (右辺) の2つの下位節に関連意味標識の返答を要求するメッセージを並列送信する。これによって、共通する標識が返って来れば意味が通る組合せであるとする。この場合には双方から DATE が返って来るので、意味が通ると判断する。そして、根の節自身の意味として DATE を持つようにする。

同様に、最終的に'誰がどうする'という文全体の概念 DO を節 s が持つことになって意味解析を含めた構文解析が終了する。

3. A-NETLによる記述

3.1 問題定義

実際の記述に当たって、東北新幹線の座席予約を例として、いつでもどこからどこに行きたいという入力文に限定した^[5]。これに対して、要求に合った座席の日時や駅名等を出力することで、英文を正しく解釈したものとする。

3.2 オブジェクト定義

オブジェクトとして、静的に次の5つを定義する。

- (1) Manager: 英文入力を直接受け付け、辞書オブジェクトへ位置情報を、構文解析オブジェクト S に文の単語数をそれぞれ提供する。
- (2) 辞書オブジェクト: 2.1 の辞書オブジェクトに対応する。Manager から送られる単語の位置と単語をその品詞に対する構文解析オブジェクトに送る。
- (3) 構文解析オブジェクト: 2.1 の解析オブジェクトに対応する。
- (4) 意味標識オブジェクト: 2.2 の意味標識に対応する。あるオブジェクトは自身に関連する標識を持つ。
- (5) Seat: 個々の座席を各々 [日付, 時刻, 列車名, 駅名, 座席名, 既・未予約] で表し、構文解析オブジェクト S から要求を受けると、予約可能な座席情報を Manager に送る。また、クラスを次のように定義する。
- (6) Node: 構文解析木の構成に際してその葉及び節に対応するオブジェクトを動的に生成する。

Node で生成されるオブジェクトは、2.2 で述べた意味標識へのオブジェクトポインタの比較を行う。

これらのオブジェクト同士のメッセージ通信の様子を図2に示す。

3.3 処理の概要

Manager は入力文を単語別に分け、辞書オブジェクトにその位置情報を、構文解析オブジェクトの1つ S に文構成単語数を送る。辞書オブジェクトは自分の単語の位置を格納し、その上で解析を要求されると、位置・品詞別に構文解析オブジェクト群に解析を要求する。

構文解析オブジェクトは、解析木ノードオブジェクトを生成・操作して解析木を構成する。解析木構成は、動的に生成されたオブジェクトが下位節となるべき他のオブジェクトをリストの要素として持つことで実現する。

複数の下位節を持つ解析木が構成される度、意味解析が行われる。解析木ノードオブジェクトは、構文解析オブジェクトから意味解析の依頼を受けて解析木の下位節に対応するオブジェクトの持つ意味標識オブジェクトにメッセー

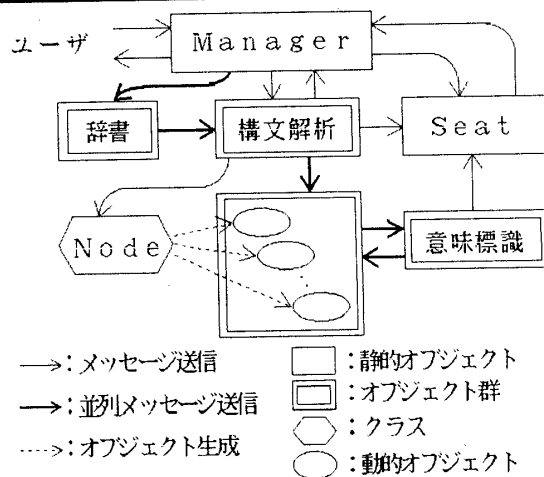


図2 オブジェクト間の関係

ジを送り、その返答によって意味の合う単語の組合せかそうでないかを判断する。またこの時、意味標識オブジェクトのうち、座席予約に直接関係する目的地名や日付などの標識オブジェクトは、意味を持つ単語の組合せが入力文中に存在する場合、解析木の葉を探索して得る情報をユーザの要求条件として Seat に送る。

解析木が完成すると、構文解析オブジェクト S は座席決定オブジェクトに解析完了のメッセージを送る。Seat は Manager を介して解析中に得られた条件に合う予約可能な座席を検索・決定してユーザに返答する。

4. おわりに

本稿では、並列オブジェクト指向言語 A-NETL による英文解釈プログラムの並列問題解決法について述べた。

現在は座席予約の話題を扱うプログラムとなっており、意味処理は解析木の節と意味標識オブジェクト群間、文の解釈は意味標識から Seat への通信によって行われる。このうち、意味標識 - Seat 間には特定オブジェクトによる関連事項の記述があるが、解析木の節は特定意味に影響されることなく、独立している。従って、Seat を話題の一部と考え、[話題-意味標識] 単位でオブジェクトを定義すれば他の話題についても適用が可能になると思われる。

また、シミュレーションの結果、並列度は文法規則数や入力文の単語の個数および品詞数が多ければ並列度が上がることが判った。これは構文解析処理時に生成されるノードオブジェクトの個数によって並列度が左右されることを意味する。今後、プログラムの改良等を行って更に並列度を増す可能性について検討していきたい。

参考文献

- [1] 加賀谷 他: "並列オブジェクト指向言語 A-NETL のプログラミング支援環境", 情報処理学会第38回大会, 4p-2 (1989).
- [2] 岩本 他: "並列オブジェクト指向言語 A-NETL の言語処理系", 情報処理学会第38回大会, 4p-1 (1989).
- [3] I. Ohisawa, A. Yonezawa: "Object-oriented parallel parsing for context-free grammars", Proc. Int. Conf. on Computational Linguistics, Budapest (Aug. 1988).
- [4] Harry Tennant, 森健一訳: "自然言語処理入門", 産業図書, 1984.
- [5] 大沢一郎, 米澤明憲: "オブジェクト指向方式における対話理解システム", コンピュータソフトウェア, Vol. 2, No. 1, pp. 11-28, 1985.