

日本語文の読みの偏好のモデル化

Preference Reading Models of Japanese Sentences

島津 明 内藤昭三
Akira Shimazu Shozo Naito
NTT 基礎研究所
NTT Basic Research Laboratories

あらまし 日本語文の特徴は左枝分かれ構造である。これを反映する読みを第1解として出力するチャート解析法を示す。まず、アジェンダのエッジの優先度を与える方略を示す。次に、チャート解析法に構成素間の位置を予測する機能を付加して、確実に左枝分かれ構造を第1解として解析する拡張チャート解析法を示す。次に、これらの方法を基本文例に適用した結果を示す。

Abstract Japanese sentences tend to have left branching structures. This paper describes chart parsing methods which produce a left branching structure as the first solution. First, several heuristics which compute priorities of edges in the agenda are given. Second, an extended chart parsing method which deterministically produces a left branching structure as the first solution is explained. Third, their experimental results are shown.

1 まえがき

日本語文の読みの偏好 [12] を反映する文構造を第1解として出力する解析法について述べる。

一般に、文は、1文だけを見ると、曖昧であるが、言語構造の特徴から読み手はある構文構造を優先してとることが心理言語学により指摘されている。英語では、右連合、最小付加、語彙的偏好という読みの偏向現象 [1][2][7] が指摘されている。一方、日本語の文構造の特徴は左枝分かれ構造であるということが指摘されている [4]。認知科学の観点からは、このような言語構造の特徴を反映する解析モデルを明らかにすることは、興味のある問題である。工学の観点からは、偏好モデルを反映する解析法は、1文の範囲内で処理する場合に、可能性の高い解を第1解とするという点で有用である。また、1度にすべての解を求めるのではなく第1解から順に適切

な解を求めることができれば効率がよい。

英語文の読みの偏好のモデル化に関しては Shieber の報告がある [7][8]。LR 解析法において、シフト/レジュースの選択に対しシフトを優先しレジュース/レジュースの選択に対し右辺の長い規則を優先する方略を述べている。さらに、これに対応するチャート解析法 [3] の方略を提案している。チャート解析法は、種々の解析方略がとれ、効率的な解析法として広く知られている解析法である。チャート法に適用される方略はアジェンダからのエッジ選択の順位を与える。しかしながら、Shieber の方略を日本語文に適用した場合、左枝分かれ構造が第1解として必ずしも出力されない。

本報告では、上述の工学的な観点から、まず、日本語文を対象に左枝分かれ構造を優先して求めるチャート解析法の方略を示す。次に、いくつかの方略の中で効率的なものを示す。さらに、チャート法を修正し、確実に左枝分かれ構造を第1解として優先して求める方法を示す。なお、日本語の LR 解析では、レジュース操作を優先すればよい。可能な限りレジュースをすることにより近い構成素どうしがまとめられて行くからである。これは以前から日本語の解析において行なわれている [9]。レジュース/レジュースの選択については、一般に日本語の文構造が係り受け (2進木) 関係により表わされるので、特に問題ではない。

2 英語文の読みの偏好のモデル化

2.1 英語文の読みの偏好現象

心理言語学において、英語文の読みの偏好現象が報告されている。それらは次のようなものである。

(1) right association (右連合)

前置詞句や副詞句が左側の動詞句や名詞句を修飾するとき、より近い項を修飾する構造を優先する。こ

これは解析木に関しては低位節付加となっている。日本語では左枝分かれ構造が優先するとすると、日本語や英語の現象に共通する読みの偏向は低位付加構造となる。

例) Joe took the book that he bought for Susan.

(2) minimal attachment (最小付加)

解析木における非終端接点の個数が最小となるような構造を優先する [2]。次の例に見るように、これはよりフラットな木を優先することを意味する。

例) Joe bought the book for Susan.

(3) lexical preference (語彙的偏好)

句をどの節に付加する読みが好まれるかは語彙的選択による [1]。例はそれぞれ低位付加と高位付加の読みの偏好を示す。

例) The woman wanted the dress on that rack.

例) The woman positioned the dress on that rack.

2.2 LR 法によるモデル化

Shieber[7] は英語文を対象に読みの偏好をモデル化する構文解析法を提案している。構文解析法は LR 法を用いる。この構文解析ではスタックを用いてシフト (shift) とレジュース (reduce) の操作により解析を進めていく。Shieber は、シフトとレジュースの操作に競合が起きたときに、どのように判断して解析を進めるかの制御方略を与えることにより読みの偏好現象をモデル化している。制御方略は以下の 2 つである。

(1) シフトとレジュースの競合の時はシフトを優先する (解析を右に右に進めていく)。

(2) レジュースとレジュースの競合の時は長い規則 (右辺の項が多い) を優先する。

一般に、(1) が右連合の読みを、(2) が最小付加の読みをもたらす。語彙的偏好に関しては、(2) を修正する。動詞が weak form か strong form かを辞書項目にマークしておき、長い規則は strong form の時に選ばれらる。なお、(1) は右に右に解析を進めていくという LR 解析の貪欲さを表わす。

日本語では、従来から試みられている [9] ように、レジュースをシフトに優先すれば、左枝分かれ構造が得られる¹。この方略および日本語規則は一般に 2 進木で表現できることから、解析表はいらぬ。

¹David Carter が同様のことを確認している。

2.3 チャート法によるモデル化

シフトレジュース法 (LR 法) は心理学的実在性の観点から興味のある方法である。一方、チャート法は、種々の解析方略がとれる効率的な優れた解析法として知られている。Shieber[8] は、チャート法により LR 法のシミュレートを試み、それが上記の制御規則に沿った動作をすることを報告している。具体的には、下降型チャート法とシフトレジュース法の操作を対応させ、この対応に基づいて、シフトレジュース的に動作するように、エッジに優先度を与え、その順に従ってエッジをアジェンダから選びながら解析を進める。

まず、LR 法とチャート法の操作は次のように対応付けられる。

(1) LR 法で作成時につくる閉包 (closure) はチャート法の予測操作 (prediction, ドットが右辺の左端にある活性エッジを作る) に対応する。

(2) LR 法のシフト操作は、活性エッジに不活性エッジを加えて活性エッジを作るチャート法の完了操作 (completion) に対応する。

(3) LR 法のレジュース操作は、不活性エッジを作るチャート法の完了操作 (completion) に対応する。

次に、この対応のもとに、次の順にアジェンダからエッジを選択する。

(p1) 予測エッジ > 活性エッジ >
語彙エッジ > 不活性エッジ

LR 法では閉包の作成がシフトとレジュースに優先するから、予測エッジは活性エッジおよび不活性エッジの付加に優先する。ただし、LR 法では前もって閉包が作られるが、これと違ってチャート法では動的に閉包が作られる。読みの優先の制御方略の「シフトはレジュースに優先する」から、活性エッジは不活性エッジに優先する。不活性エッジは語彙的 (lexical) なものとそうでないものに分け、語彙的エッジによる操作がよりシフトに近いとみなし、他の不活性エッジに優先する。語彙的エッジとは NP → John という語彙規則に対するエッジ [i, NP → John •, i+1] である。

次に、(p1) により順序付けられないとき、次の優先度を用いる。

(p2) ドットの位置がより右にあるエッジを優先する。

- (1) 角から3軒目(の家)に居ます。
- (2) 外国へ／に出張の先生が帰られた。
- (3) 人口が1億以上(の国)を言いなさい。
- (4) ここより上(の方)に行きましょう。
- (5) 3個で百円のリンゴを買いました。
- (6) 白い煙を合図に出発した。

図 1: 格助詞の付いた後置詞句が名詞を修飾する例

これは、右に右に解析を進めていく貪欲さ、すなわちシフトをレジュースに優先させることの反映である。

次に、(p2)により順序付けられないとき、次の優先度を用いる。

(p3)より長い規則により作られたエッジを優先する。

これはレジュース / レジュース競合の時の優先方略の反映である。

3 日本語文の読みの偏好のチャート法によるモデル化

日本語構造の性質に関しては左枝分かれ構造が指摘されている [4]。これは前節に述べたように解析木に関して低位付加を優先することに対応する。左枝分かれ構造は、文が左から右に書かれるとして、修飾部が先に主部が後にくる性質と、近い要素同士が結び付く(自然な読み)ことからでてくる。ところで、「が」、「を」、「に」等の格助詞が後置する句は句節を飛び越して動詞に掛かる場合、全体は左枝分かれ構造ではない。しかしながら、それらの句は名詞にも掛かる場合がある(図 1) [6]。また、それらの後置詞の次に動詞の埋め込み句がくる場合もある。従って、近い要素同士を先ず結び付けてみてもし適切でないなら次の構造を求める方略は妥当であろう。

3.1 Shieber の方略の日本語解析への適用

Shieber の優先順位を日本語文に適用した場合、必ずしも左枝分かれ構造が第 1 解として得られるとは限らない。例えば、「角から 3 軒目に行く」では、[角から [3 軒目] [行く]]となる。ちなみに、「先生が飼っている猫が殺したネズミが食べた魚はくさっていた」 [4] は [[[先生が飼っている]][猫が殺した][ネズミが食べた]魚]はくさっていた」となる。これら確かめた実験で用いた文法を付録 A に示す。チャート法および LR 法のトレースを付録 B、付録 C に示す。うまくいかない理由は日本

語の構文構造と読みの偏向に関する優先方略とが合わないためである。日本語では修飾部が左に主部が右にくる。特に、文の主部である用言は文末にくる。一方、優先方略はシフトを優先する。上の例で言うと、小さな単位で後置詞句が固まっては ([角から], [3 軒目]), 予測あるいはシフトの優先により解析が右に右に進み、文末の用言まできて、もう右には進めないから完了あるいはレジュースにより置いて置かれた後置詞句が用言と次々にまとまって、右枝分かれ構造が先に完成する。逆に、英語の場合は、左にある主部に右にある修飾部がまとめ上げられていき、右連合の読みにならうことになる。ただし、複数のまとめ上げが可能なときは(レジュース / レジュースの競合)、長い方のまとめ上げが優先され、最小付加の読みがなされる。

例えば、「角から 3 軒目に行く」の解析では、[NP → PP NP ·] ([角から 3 軒目]) と [PP → NP · P] ([3 軒目]_) とがアジェンダにあるとき、優先方略 (p1) により [PP → NP · P] が優先して選択される。その結果、[3 軒目] と [に] とを結合したエッジ ([PP → NP P ·]) が作られ、アジェンダに加えられる。このときアジェンダには [[角から] 3 軒目] および [3 軒目]に] に対するエッジがある。これらはどちらも不活性エッジのため、どちらが先に選択されるかは任意である。実験では [3 軒目]の方が先に選択された。その後 [U → PP · U] ([3 軒目]_) というエッジが作られる。これは優先方略 (p1) により [[角から] 3 軒目] のエッジよりも先に選択され、[[3 軒目]に行く] に対するエッジがつけられる。

3.2 Shieber の優先順位の変更

上述(付録 C) したように、LR 法でも Shieber の方略が日本語文に対してうまく機能しない。付録 C から分かるように、また、従来から試みられているように [9]、LR 法ではシフトとレジュースの優先順を換えればよい。チャート法でも同様の方策が考えられるが、実際に試みしてみるとうまく機能しない。不活性エッジが活性エッジに優先するようにしても同様にうまく機能しない。この場合、活性エッジの優先度が小さくなるので、修飾部を含む活性エッジがアジェンダに残され、文の右側の方が先にまとまってしまうからである。予測エッジの順位を最下位にする方略も考えられる。これは、LR 法では閉包を最初に作るので、LR 法のシミュレートにはならないが、上例はうまくいく。しかし、複合文(例えば、「信号を見て道を渡る。」)がうまくいかない ([[信号] [見て [道] 渡る]])。さらに、この方

表 1: 実験に用いた基本文の例

	文 (品詞列)	特徴
(1)	今度の学会の子稿 (N P N P N)	名詞句
(2)	外国に出張の先生が帰る. (N P N P N P V)	単位文, 名詞句
(3)	今度の学会に出す原稿を作る. (N P N P V N P V)	単位文, 埋め込み文
(4)	信号を見て横断歩道を渡る. (N P V P N P V)	複合文
(5)	先生の家の庭に咲いたバラをもらって教室の部屋を飾る. N P N P N P V N P V P N P N P V	複合文, 名詞句 埋め込み文

略は、予測の順位を下けているため、右への解析の進み方が悪く、解析の能率もよくない。

3.3 日本語文向きの解析方略

左枝分かれ構造を第1解として出力するチャート法の優先方略を示す。それは、Shieberの方略にエッジ生成順情報を加えるものである。

Shieberの方略が日本語文に適用されたときの分析をみると、アジェンダにある低位付加構造が先にとられず、予測エッジと活性エッジの優先により解析が右に進んでいくため、結果として左枝分かれ構造が最初の解とならない。すなわち、文頭の低位付加構造が文の右側の構造より先に作られていることがうまく利用されていない。そこでエッジが生成される順も考慮する。修正方略として、Shieberの方略では順序が付けられないとき、すなわち、(p3)により順序づけられないとき、次の優先度を用いる。

(p4) 先に作られたエッジを優先する。

この方略により、表1の基本文について解析を試みたところどれも左枝分かれ構造が第1解として得られた。実験に用いた文は、単位文、複合文、埋め込み文、複合名詞句を組み合わせた文等の基本文型である。

3.4 他の方略

上記の他に考えられる日本語向け方略を以下に示す。

(1) 逆方向の解析方略

日本語は修飾部が前に主部が後ろにある。英語は概ねその逆である。そこでShieberの方略を用い右から左に解析する方法が考えられる。実際、表1の文等に対し第1解として左枝分かれ構造が得られる。

(2) 上昇型の方略

上述の解析方略はいずれも下降型のチャート法に對

するものである。上昇型にShieberの方略をそのまま適用しても次の理由でうまくいかない。上昇型の場合、予測操作がなく、3.1節、3.2節の場合のように小さな単位でまとまりやすい。すなわち、この場合も、先に作られた活性エッジがアジェンダに取り残されるために小さな単位が先にまとめられてしまう。そこで、先と同様にShieberの方略に生成順を加味した優先順位方略にすると、表1の文について第1解として左枝分かれ構造が得られる。

4 左枝分かれ構造を確実に優先する拡張チャート解析

上述の方略は一種のヒューリスティックスであり、左枝分かれ構造を確実に優先するという保証がない。そこで、エッジ表現を拡張して、上述のような方略を用いなくても、確実に左枝分かれ構造を優先して求める解析法を示す。すなわち、チャート解析において、単語境界を予測しながら解析する。特に、最長の予測エッジを優先して選んでいくと、左枝分かれ構造が優先して解析される。

(例) 自由の女神の写真
0 N 1 P 2 N 3 P 4 N 5

以下に示す解析過程の(1)において、PPとNPの境界が4または2と予測できる。ここで4の予測を選んで解析を進めると左枝分かれ構造が優先して解析される。

- (1) NP → ・0 PP 4 NP 5 (自由の女神の | 写真)
NP → ・0 PP 2 NP 5 (自由の | 女神の写真)
- (2) PP → ・0 NP 3 P 4 (自由の女神 | の)
- (3) NP → ・0 PP 2 NP 3 (自由の | 女神)
- (4) PP → ・0 NP 1 P 2 (自由 | の)
- (5) PP → 0 NP・1 P 2 ((自由) の)

- (6) $PP \rightarrow 0 NP 1 P \cdot 2$ ([[自由] の])
 (7) $NP \rightarrow 0 PP \cdot 2 NP 3$ ([[自由] | の] | 女神)
 (8) $NP \rightarrow 0 PP 2 NP \cdot 3$ ([[自由] の] 女神)

 (12) $NP \rightarrow 0 PP 4 NP 5$ ([[[[自由] の] 女神] の] 写真])

4.1 分割位置情報付きエッジ表現

A を非終端構成素, $B_i (i = 1 \sim n)$ を非終端構成素または終端構成素, $c_i (i = 0 \sim n)$, i, j を単語境界の位置として, $[A, (c_0) B_1 (c_1) \dots (c_{k-1}) B_k (c_k) \cdot B_{k+1}(c_{k+1}) \dots (c_{n-1}) B_n (c_n), i, j]$ を分割位置情報付きエッジ (以後, 簡単に予測 / 活性 / 不活性エッジ) と呼ぶ. 解析が i から j の間まで進んでおり, B_h がカバーする単語列が $ch-1$ と ch の間にあることを示す. $c_0 \sim c_n$ を予測位置と呼ぶ. ここで, $c_0 = i, c_k = j$ である.

- (例) 自由 の 女神 の 写真
 0 N 1 P 2 N 3 P 4 N 5
 (1) $[NP, 0 \cdot PP 4 NP 5, 0, 0]$ (自由の女神の | 写真)
 (2) $[NP, 0 \cdot PP 2 NP 5, 0, 0]$ (自由の | 女神の写真)

なお, 例の (1) は (2) に較べて左構成素のカバーする単語列が長いという意味で「左が長いエッジ」と呼ぶ.

4.2 分割位置情報付きエッジを用いる下降型 Chart 法

チャート解析法を次のように拡張する. 予測のステップでは構成素間の位置を予測し, 分割位置付きエッジを作る. 完了のステップでは新たなドット位置と予測された位置が等しい場合にだけ完了操作を行なう. 詳細を付録 D に示す.

4.3 左枝分かれ構造を優先する解析

予測のステップにおいて, 一般に, 複数の予測エッジをアジェンダに置きうる. そこで, 左が最も長いエッジを優先して取ると (実際のプログラムでは効率のため, エッジをアジェンダに付加するときにソートする), 上例のように左枝分かれ構造が優先して解析される.

4.4 構造パターンの適用による構成素境界の予測

予測エッジは構造パターンと呼ぶ規則により求めることができる [10]. 構造パターンは, 構成素の特徴パターンを付加した文脈自由規則である. 予測位置を決める制約として特徴パターンが予測ステップ時に用いられる.

- (例) $U \rightarrow PP(\text{pattern1}) U(\text{pattern2})$
 $\text{pattern1} = (*x \text{ *particle})$
 $\text{pattern2} = (*y >\text{predicate-part})$

4.5 エッジのジェネレータ化

予測のステップにおいて複数のエッジが予測される場合, 左最長エッジだけをアジェンダに置き, 後のエッジは, 解析のやり直しにより必要になったときに用意するようにすると効率が良い. そこで, アジェンダに置くエッジでドットが左端のエッジをジェネレータとする [10][11]. これをジェネレータ・エッジと呼ぶ.

- (例) 自由 の 女神 の 写真
 0 N 1 P 2 N 3 P 4 N 5
 $\text{gen}(NP, \text{自由の女神の写真})$
 $= [NP, 0 \cdot PP 4 NP 5, 0, 0]$ (自由の女神の | 写真)
 $\text{gen}([NP, 0 \cdot PP 4 NP 5, 0, 0])$
 $= [NP, 0 \cdot PP 2 NP 5, 0, 0]$ (自由の | 女神の写真)
 $\text{gen}([NP, 0 \cdot PP 2 NP 5, 0, 0])$
 $= \phi$

上記の解析において, アジェンダから取り出したエッジがジェネレータであれば, 次に予測されるエッジを求めて (次のエッジが求められないなら次のエッジをアジェンダから求める), それをアジェンダに残すとともに, 予測操作を行なう. ジェネレータとしてのエッジは, 解析過程の環境情報として, 入力単語列と構成素境界との対応情報 (構造パターンでは変数と単語列との束縛情報) を保持しておいて次のエッジの生成に用いる.

4.6 意味解析

構造パターンに意味解析関数を付加し, チャート解析の完了ステップにおいて意味解析関数を適用することにより, 部分構造の適切性を調べるとともに意味構造をまとめ上げることができる.

- (例) $U \rightarrow PP(\text{pattern1}) U(\text{pattern2}) | SF$
 $\text{pattern1} = (*x \text{ *particle})$
 $\text{pattern2} = (*y >\text{predicate-part})$
 $SF = \text{case-analysis}(PP, U)$

意味解析を併用するために, 予測エッジおよびジェネレータ・エッジは次のように拡張する. A を非終端構成素, $B_i (i = 1 \sim n)$ を非終端構成素または終端構成素, f を意味解析関数, $c_i (i = 0 \sim n)$, i, j を単語境界の位

表 2: 単文に対する実験結果

方法	順番	#EC	#EA	時間(秒)	解(個)
生成順	F	111	245	3.55	1
	L	137	293	4.30	2
逆方向	F	80	175	1.96	1
	L	120	267	3.22	2
上昇型	F	31	34	0.13	1
	L	41	43	0.25	2
拡張チャート解析	F	26	27	5.05	1
	L	48	53	9.42	2

文: 角から3軒目に行く。

#EC: チャートに入れられたエッジ数

#EA: アジェンダに入れられたエッジ数

F: 第1解が求められた時点

L: アジェンダが空になった時点

注: プログラムは原理的なものである。

チャートの効率化などはしていない。

拡張チャート解析には意味解析等が入っている。

計算機は Symbolics 3600.

置, $B1', \dots, Bk'$ をそれぞれ $B1, \dots, Bk$ の意味構造とするとき², $[A, (c_0) B1 (c_1) \dots (c_{k-1}) Bk (c_k) \cdot Bk+1(c_{k+1}) \dots (c_{n-1}) Bn (c_n) \mid f, i, j, _ , (B1' \dots Bk')]$ を改めて予測 (ジェネレータ) / 活性 / 不活性エッジと呼ぶ。完了操作においてドットが c_n の位置まで進んだとき, 意味解析により意味構造ができれば, $_$ の位置に意味構造 A' を加える。できなければエッジは作らない。

4.7 再解析

上述の最長一致の方略により正しい構造がストレートに求められるとは限らない。しかしながら, 部分的に構造をまとめ上げながら解析を進める上述の方法により, 構造の不適切さを, あるいはまとめ上げが出来ないことを検出することができる。問題が検出されたとき, アジェンダ上には不適切な構造を導くに至ったエッジの親であるジェネレータ・エッジがある。そのジェネレータ・エッジを再度用いて次の予測をすることにより次の解析を進めることが出来る [11]。部分構造の不適切性の検出には, 格構造を示す用言, 文構造 [5] を規定する接続助詞等の語彙的な制約を利用することができる [11]。

²実験では $B1'$ の位置に複数の意味構造を書けるようにし曖昧な場合のエッジの個数をまとめている。

表 3: 複合文に対する実験結果

方法	順番	#EC	#EA	時間(秒)	解(個)
生成順	F	932	2063	81.98	1
	L	1236	2617	145.67	112
逆方向	F	250	526	10.72	1
	L	1075	2265	140.88	112
上昇型	F	660	705	34.18	1
	L	964	1076	82.03	112
拡張チャート解析	F	349	411	127.03	1
	L	609	691	218.97	10

文: 先生の家の庭に咲いたバラをもらって教室の部屋を飾る。

他の事項は表2と同様である。

上述したジェネレータ・エッジ表現および意味解析関数の適用により左枝分かれ構造を優先して解析する解析を拡張チャート解析法と呼ぶ。作成したプログラムでは, 形態素解析結果をラティス表現し, 構造パターンにより語を最長一致的に選択している。例えば, 「くるまでまつ」は, 第1番目の解が「車で待つ」, 第2番目の解が「来るまで待つ」となる。

5 実験結果

4種類の方略による実験結果を表2, 表3に示す。紙面の都合で単文, 複合文に対する結果だけを示すが, 他の文型の実験結果も, 方略間の相対的關係はほぼ同様である。実験プログラムは原理的なものであり, チャートの効率化などはしていない。実験は, 拡張チャート解析法を除き, 分かち書きした語列が入力である。拡張チャート解析法はラティス構造表現された形態素解析結果を入力とする。拡張チャート解析の時間には, 意味解析および予測エッジ作成 (ラティス構造の形態素解析結果と構造パターンとのマッチ) の時間が余分に含まれている。意味解析をする拡張チャート解析を除くと, 無駄な予測がないだけ, 全般に上昇型がよい結果となっている。

6 あとがき

日本語文は左枝分かれ構造の特徴を持つ。これをモデル化するチャート解析法の方略 (アジェンダにおけるエッジの優先度) および基本文型の実験結果を示した。次にチャート解析法に構成素間の位置を予測する機能を

付加して、確実に左枝分かれ構造を第1解として解析する拡張チャート解析法を示した。これは意味解析を加えて語彙情報を利用しより適切な解析をするとともに、ジェネレータ機能を持たせたエッジ表現により不用な解析を避けて第1解に向け直線的に解析する。

参考文献

- [1] Marilyn Ford, Joan Bresnan, and Ronald M. Kaplan. A Competence-Based Theory of Syntactic Closure. in *The Mental Representation of Grammatical Relation*, J. Bresnan, ed., MIT Press, 1982.
- [2] Lyn Frazier and Janet D. Fodor. The Sausage machine: A new two-stage parsing model. *Cognition*, 6, 1978.
- [3] Martin Kay. Algorithm Schemata and Data Structures in Syntactic Processing. Technical Report CSL-80-12, Xerox PARC, 1980.
- [4] 久野暲. 日本語文法. 大修館, 1973.
- [5] 南不二夫. 現代日本語の構造. 大修館, 1974.
- [6] 荻野紫穂. 連用修飾を受ける体言. 東京女子大学卒業論文, 1986.
- [7] Stuart M. Shieber. Sentence Disambiguation by a Shift-Reduce Parsing Technique. in the proceedings of the 21st ACL, 1983.
- [8] Stuart M. Shieber. A Uniform Architecture for Parsing and Generation. in proceedings of the 12th Coling, 1988.
- [9] Akira Shimazu and Hitoshi Iida. Experimental Japanese Language Question Answering System MSSS78. in proceedings of the sixth IJCAI, 1979.
- [10] 島津明, 内藤昭三, 野村浩郷. 構造予測を用いた日本語文の意味解析法. 情報処理, Vol.27, 1986.
- [11] 島津明, 内藤昭三. 適切な解を順に求める解析法. 情報処理学会第37回全国大会, 1988.
- [12] 辻井潤一. 文解析方式. 情報処理, Vol.27, No.8, 1986.

付録

A 実験に用いた規則の例

$S \rightarrow U, S \rightarrow C, S \rightarrow PP\ C, S \rightarrow S\ AUX,$
 $U \rightarrow PP\ U, U \rightarrow NP\ U, U \rightarrow ADV\ U,$
 $U \rightarrow V, U \rightarrow A$
 $PP \rightarrow NP\ P, PP \rightarrow PP\ P,$
 $NP \rightarrow PP\ NP, NP \rightarrow S\ NP, NP \rightarrow N,$
 $C \rightarrow S- S,$
 $S- \rightarrow S\ P$

B 第1解を得るまでの解析のトレース 1

例文: 角から3軒目に行く。

方法: Shieberの方略によるチャート解析

記号: A: = アジェンダへ付加

G: = アジェンダから選択

1. A: [N → 角・] ([角])
2. A: [P → から・] ([から])
-
62. G: [P → から・] ([から])
63. G: [N → 角・] ([角])
64. A: [NP → N・] ([角])
-
124. A: [PP → NP P・] ([角から])
125. G: [NP → NP P・] ([角から])
-
129. A: [U → PP・U] ([角から]_)
-
170. A: [NP → N・] ([3軒目])
-
188. G: [NP → PP・NP] ([角から]_)
-
193. G: [NP → N・] ([3軒目])
194. A: [U → NP・U] ([3軒目]_)
195. A: [PP → NP・P] ([3軒目]_)
196. A: [NP → PP NP・] ([角から] 3軒目)
197. G: [NP → NP・P] ([3軒目]_)
198. A: [P → NP P・] ([3軒目]C)
-
254. G: [NP → NP P・] ([3軒目]C) (注)
-
256. A: [U → PP・U] ([3軒目]C]_)
-
306. A: [U → V・] ([行く])
-
314. G: [U → PP・U] ([3軒目]C]_)
-
322. G: [U → V・] ([行く])
323. A: [S → U・] ([行く])
324. A: [U → PP U・] ([3軒目]C [行く])
325. G: [U → PP U・] ([3軒目]C [行く])
326. A: [S → U・] ([3軒目]C [行く])
327. A: [U → PP U・] ([角から] 3軒目]C [行く])

328. G: [U → PP U ·] ((角から [3 軒目に [行く]])])
 329. A: [S → U ·] ((角から [3 軒目に [行く]])])

(注) この時点で、アジェンダには [角から [3 軒目]] および [3 軒目に] に対するエッジがある。これらはどちらも不活性エッジのため、どちらが先に選択されるかは任意である。実験では [3 軒目に] が先に選択された。エッジの生成順も加味した方略を用いると、先に作られた [[角から] 3 軒目] が先に選択され、これが「に」、「行く」と順に結びついて、左枝分かれ構造が得られる。

C 第1解を得るまでの解析のトレース 2

例文: 角から3軒目に行く。

方法: Shieber の方略による LR 法解析

スタック	入力
1 0	角から 3 軒目に行く
2 0 角 8	から 3 軒目に行く
3 0 N 9	から 3 軒目に行く
4 0 NP 5	から 3 軒目に行く
5 0 NP 5 から	3 軒目に行く
6 0 NP 5 P 7	3 軒目に行く
7 0 PP 3	3 軒目に行く
8 0 PP 3 3 軒目 8	に行く
9 0 PP 3 N 9	に行く
10 0 PP 3 NP 5	に行く (注)
11 0 PP 3 NP 5 に 6	行く
12 0 PP 3 NP 5 P 7	行く
13 0 PP 3 PP 3	行く
14 0 PP 3 PP 3 行く 10	
15 0 PP 3 PP 3 V 2	
16 0 PP 3 PP 3 U 4	
17 0 PP 3 U 4	
18 0 U 1	
19 -	
11' 0 NP 5	に行く
12' 0 NP 5 に 6	行く
13' 0 NP 5 P 7	行く
14' 0 PP 3	行く
15' 0 PP 3 行く 10	
16' 0 PP 3 V 2	
17' 0 PP 3 U 4	
18' 0 U 1	
19' -	

(注) 決定表により、10 において、NP → PP NP のレジュースをするか「に」をシフトするかの選択枝が生ずる。Shieber の方略に従えばシフトが優先するので、11 から 18 のように解析が進み、左枝分かれ構造が最初の解とならない (レジュースを優先すると 11' 以降のように解析が進み左枝分かれ構造が得られる)。

D 分割位置情報付きエッジを用いる下降型チャート解析法

<準備> 構成素が S (文) であるエッジをアジェンダに加える。ここで、入力単語列の範囲が 0 から n として、エッジを [S, (0) · B1 (p₁) ... (p_{k-1}) Bk (n), 0, 0] とする。位置 p_i (i = 1 ~ k-1) は何らかの方法 (例えば、4.4 で述べるような構造パターン) により求められるとする。位置 p₁, ..., p_{k-1} の組み合わせが複数あれば、一般にエッジは複数個できる。また、単語列を w₁, ..., w_n, 対応する規則を A1 → w₁, ..., A_n → w_n とするとき、エッジ [A1, w₁ ·, 0, 1], ..., [A_n, w_n ·, n-1, n] をアジェンダに加える。

次に、アジェンダからエッジ E を取り出し、それをチャートに加えると共に、E に対して、以下の 3 つの手続きのいずれかをアジェンダのエッジがなくなるまで繰り返し適用する。なお、完了操作において [S, (0) B1 (p₁) ... (p_{k-1}) Bk · (n), 0, 0] ができれば、それは解の一つである。

<予測> E が、解析済み位置が ij とする活性エッジ Ea のとき、Ea のなかでドットの右側の非終端構成素が B のとき ([A, (i) α (j) · B (c_k) β (c_n), i, j]), B → C1 ... Ck というすべての規則に対して、[B, · (j) C1 (c₁) ... (c_{k-1}) Ck (c_k), j, j] というエッジをアジェンダに加える。c₁, ..., c_k の組み合わせが複数あれば、作られるエッジも複数できる。ただし、既に同じエッジがチャートに存在するときにはその限りではない。

<完了> Ea と Ei が解析済み位置がそれぞれ ij, jk である活性エッジおよび不活性エッジとして、E が Ea あるいは Ei のとき、すなわち、E が Ea(Ei) なら、Ei(Ea) はチャートから取り出す。チャートになければ、単に E をチャートに入れる。このとき、Ea のドットの右側の構成素が B であり ([A, (i) α (j) · B (k) β (n), i, j]), Ei が構成素 B のエッジ ([B, (j) r (k) ·, j, k]) であるなら、ドットを B の右側、解析済み位置を ik とする他は Ea と同じ内容のエッジ ([A, (i) α (j) B (k) · β (n), i, k]) を新しくアジェンダに加える。

<走査> E が、解析済み位置が ij でありドットの右側が終端構成素 w_{j+1} である活性エッジ Ea のとき、w_{j+1} の右側位置を j+1 として ([A, (i) α (j) · w_{j+1} (j+1) β (n), i, j]), ドットを w_{j+1} の右側、解析済み位置を i, j+1 とする他は Ea と同じエッジ ([A, (i) α (j) w_{j+1} (j+1) · β (n), i, j+1]) を新しく作り、チャートに加える。