

自然言語の構文・意味解析規則の主観的確率を用いた帰納的学習システム

An Inductive Learning System of Natural Language Grammar with the Epistemic Probability

中川 聖一* 若原 一彰*
Seiichi Nakagawa Kazuaki Wakahara

* 豊橋技術科学大学情報工学系
Dept. of Information and Computer Sciences, Toyohashi University of Technology, Toyohashi 440, Japan.

1988年3月22日 受理

Keywords: inductive learning, probabilistic learning, Dempster-Shafer's theory, learning of syntax and semantics.

Summary

Natural language has many ambiguities, so the consideration of ambiguities is important for learning syntactic construction and semantics. In order to acquire the grammar and the meaning of words, it is necessary to have the flexibility and to gain the experience of learning by introducing not truth values (0 and 1) but a degree of uncertainty. So, for the purpose of modeling such a learning process, we introduced Dempster-Shafer's theory into our inductive learning system of natural language grammar.

Our system, using CFG (Context-Free Grammar) as the syntactic representation and the semantic network as the semantic representation, can learn the grammar and the meaning of words with the epistemic (subjective) probability inductively, initially without presenting the grammar and the meaning of words.

1. はじめに

自然言語は、曖昧性に富んでいるので、統語的な構造や意味を学習するのに、曖昧性を考慮に入れる必要がある。文法や単語の意味を獲得していくためには、真か偽かの2値で処理するのではなく、確からしさの度合を導入することによって柔軟性をもたせ、連続的に学習経験を積み重ねることが必要である。そこで、このような学習プロセスをモデル化するために、自然言語の構文・意味解析規則の帰納的学習⁽¹⁾に、Dempster-Shaferの主観的確率の理論⁽²⁾を導入した。

本研究の言語の帰納的学習法の理論的背景としては、J. Andersonの理論⁽³⁾がある。彼の理論の中心となる主張は、入力として文と意味のペアをいれてやれば解析規則のほとんどが推論可能であるということである。彼の理論は、LASというシステムで実現さ

れている。ここで、Andersonは、意味の使用以外に次の三つの仮定をしている。

- (1) 子供は単語の意味を知っている。
- (2) 子供は意味構造のトップ・ノードが、どれか知っている。
- (3) 解析木の枝は交差してはならない。

しかし、このシステムは文法を学習するものであり、単語の学習ができない。また、“the”などの意味をもたない単語や多義語を扱うことができないなどの問題点がある⁽⁴⁾。本研究では上記の(2)と(3)の仮定はしているが、(1)の仮定はしていない。

さらに本システムは、時制の学習(現在形・過去形)と否定文・疑問文の学習ができる⁽¹⁰⁾。時制の学習や命令文・平叙文の理解、平叙文の生成の研究にはSelfridgeのCHILD理論⁽⁵⁾がある。

彼のCHILD理論は、子供の言語学習の六つの事実を取り上げ、それらの事実と矛盾しない方法で、言

語を理解・生成することを目的としている。しかし、この理論にも、ユーザからのフィードバックを受けたり、言語的な一般化をしていないなどの問題点がある。

本システムでは、構文解析規則に文脈自由文法を、意味表現にセマンティック・ネットワークを仮定し、入力として文と意味のペアを与えながら、単語の意味や解析規則を知らない状態から構文・意味解析規則を帰納的に学習し、時制（現在・過去）と否定文・疑問文も扱えるように開発されているが、さらに確率的に学習できるようにシステムを拡張、改良した。

2. システムの構成

Fig. 1 に本システムの構成を示す。学習部は文と意味のペアを受け取り、文を解析できなければその文を解析できる新しい構文・意味解析規則を推論して、すでに学習している構文・意味解析規則との間で一般化し変更する。学習が終わった状態では構文・意味解析部に文を入力することによって、その文の意味を得ることができる。

構文解析規則には文脈自由文法を、文の意味表現に

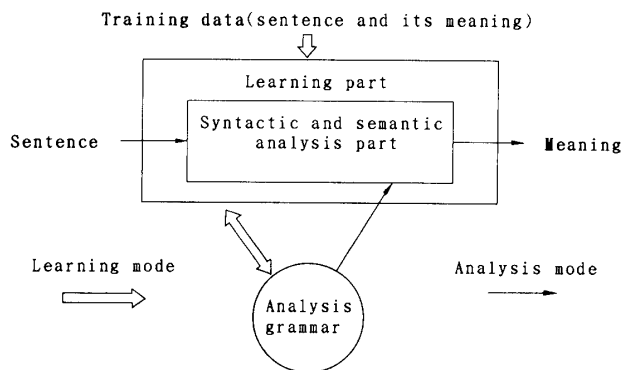


Fig. 1 Blockdiagram of learning system.

Operator for generating nodes

- GENR : Generation of the predicate (Correspondence to "give" in Fig. 4)
- GENO : Generation of the object (Correspondence to "< dog >" in Fig. 4)
- GENE : Generation of the concept (Correspondence to "[Mary]" in Fig. 4)

Operator for linking

- PUTN : linking between two nodes by a label

Fig. 2 Operator for semantic representation.

Decomposition of semantic representation

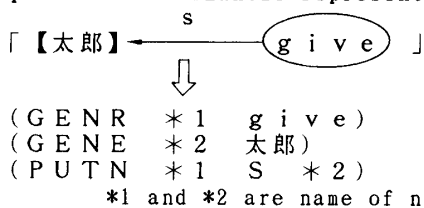


Fig. 3 An example of semantic representation.

はセマンティック・ネットワーク⁽⁶⁾を用いた。学習部に送られる文の意味表現は、解析部が意味表現を構成するときに使われる命令に分解される。Fig. 2に命令の種類をあげ、Fig. 3に分解の例を示す。解析部に

Analysis grammar		
Confidence score	Surface structure	Semantics
0. 8	S → NP1 VP	Linking NP to VP by label "s"
0. 6	VP → V NP2	Linking NP2 to V by label "obj"
0. 3	V → gave	Generation of node "give"
0. 4	NP2 → A N2	Concept "N2"
0. 3	NP1 → N1	Concept "N1"
0. 2	N1 → Mary	Generation of node " [Mary] "
0. 5	N2 → dog	Generation of node "<dog>"
0. 6	A → the	No meaning

Analysis grammar (by using grammar above)

Input : Mary gave the dog

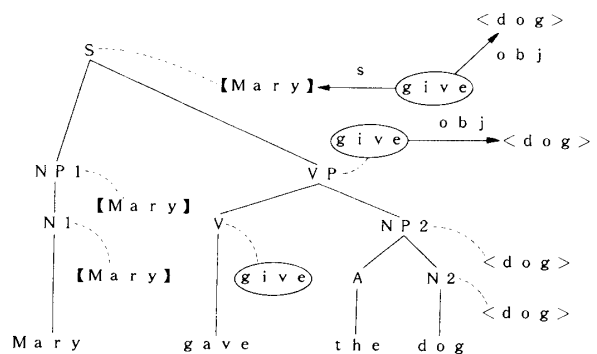


Fig. 4 Analysis grammar and analysis example.

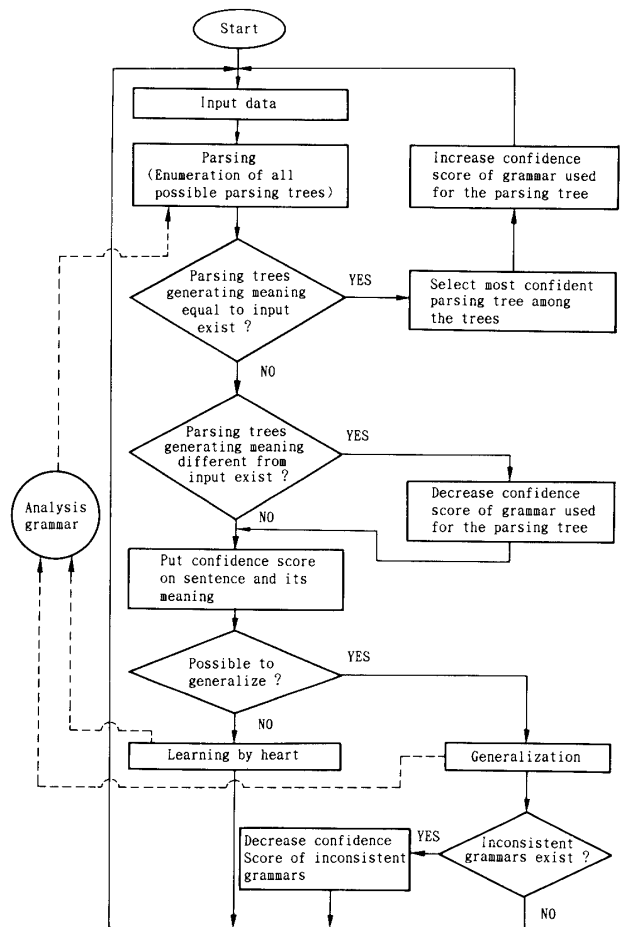


Fig. 5 Algorithm of learning.

は拡張 LINGOL⁽⁷⁾ を改造したものを使用している⁽¹⁾。解析規則と解析例を Fig. 4 に示す。また、学習部の詳細なアルゴリズムを Fig. 5 に示す。システムは、いつでも規則の一般化をできるように逐次的な処理になっている。一般化ヒューリスティックは3章で、信頼度の定義・計算方法は4章で詳しく述べる。本システムは、Mulisp を使って書かれ、PC-9801 UV 上で実現されている。

3. 一般化ヒューリスティック⁽¹⁾

本研究で使った学習部の一般化のヒューリスティックについて述べる。

3.1 言語と意味の制約

本研究で仮定した、学習される言語と意味の制約は以下の二つである。

- (1) CFG で解析できる言語であること。
- (2) 意味構造のノードをたどっていったときにループしてはならない。

CFG を文法とすることで、Anderson の仮定した「解析木の枝が交差してはならない」という条件を本研究でも仮定していることになる。また、意味構造から文法を引き出すためには、意味構造が CFG と同じものでなければ推論不可能となるので、ノードをたどっていったときにループしないことが要求される。そのほかに意味構造の、どのノードがトップ・ノードであるかの情報も入れてやる必要がある⁽³⁾⁽⁴⁾。

3.2 AND ヒューリスティック

解析できなかった文(または単語列)は暗記学習されているが、それらの暗記学習されている文(または単語列)を比較した場合、表層部に共通な単語列がただ一つだけあり、かつ、それらの意味構造にも共通な構造がただ一つだけあるとき、その単語列がその意味を持っていると推論できる。ただし、その意味構造は、残りの意味構造と一つだけのリンクによってつながっていないなければならない。そのリンクでつながっているノードがその単語列のトップ・ノードとなる。二つのリンクでつながっているとトップ・ノードが決定できない。また、もとの意味のトップ・ノードを含んでいる場合は、そのトップ・ノードがリンクしていなければならない。次に、共通部分以外の残りの部分に対しても規則を作ろうとする。残りの単語列が、共通部分によって二つに分割されずに連続して残っているならば、残りの単語列を残りの意味に対応づける規則を作

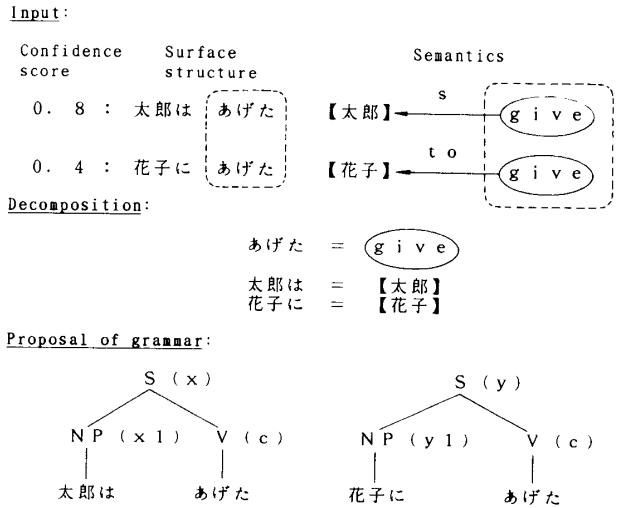


Fig. 6 An example of AND heuristics.

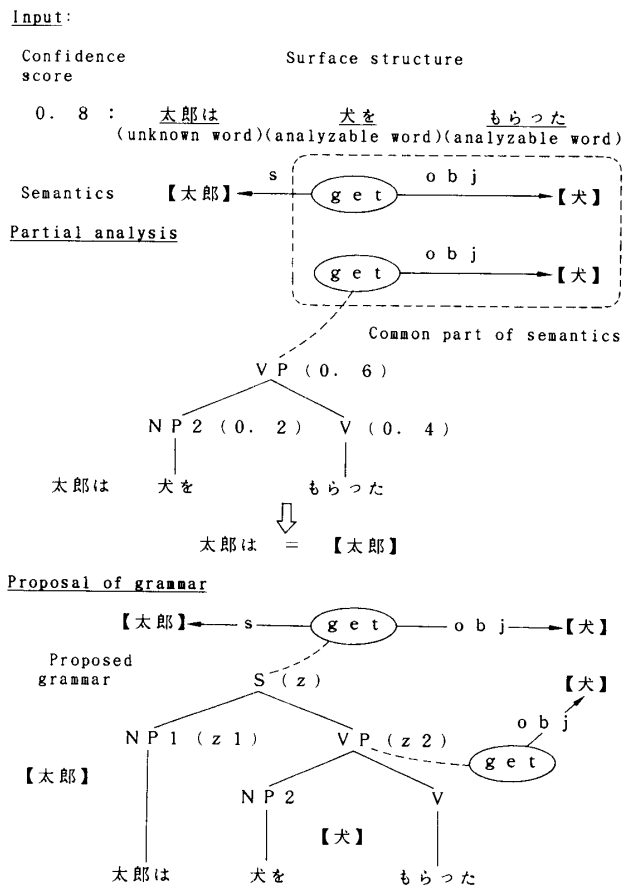


Fig. 7 An example of partial analysis heuristics.

る。分割されていればそのような規則は作らない。さらに、分解した単語列を一つにまとめる規則を作って終わる。Fig. 6 に例を示す。

3.3 部分解析ヒューリスティック

文(または単語列)を全部は解析できないが、部分的に解析できるときに使われる。一つの部分単語列が解析され残りの部分単語列が解析不能の場合には、AND ヒューリスティックと全く同じ働きをする。ま

た二つの部分単語列が解析され、解析されたそれぞれの意味を単に1本のリンクで結合することにより、もとの文の意味を作ることができるときは、それらを結合してもとの文の意味を作るような規則を作る。そのとき、二つの意味のトップ・ノードを結ぶことによって意味が完成されなければならない。Fig. 7 に例を示す。意味を持たない単語(例えば、冠詞“the”, 助詞“は”や“が”)が発見されたときは、その単語の表層上の両どなりのどちらの単語にその単語を付けるか問題となる。本システムでは、セマンティック・ネットワーク上で示されている単語の方に意味を持たない単語を結びつける。また、両どなりのどちらも示されていたら、解析木の枝がその意味を持たない単語のすぐ上を通る単語に結びつける。

3.4 マージヒューリスティック⁽¹⁰⁾

文法の右辺の、ある非終端記号の名前を変えると文法もそれに付いている意味規則も同じになる二つの規則があるとき、それらの規則を一つにまとめるために使われる。ただし、*kind* 情報は同じでなくてもよ

Merging of grammar:

Confidence score	Surface structure	Semantics
0.7	S = NP V	(Linking NP to V by label "s". *kind* information of V is {AFFIRMATION} and {PRESENT}.)
0.6	S = NP V1	(Linking NP to V1 by label "s". *kind* information of V1 is {AFFIRMATION} and {PRESENT}.)
0.3	V = give	(Concept "give", {AFFIRMATION} and {PRESENT}.)
0.4	V1 = have	(Concept "have", {AFFIRMATION} and {PRESENT}.)

Merged grammar:

w	S = NP V	(Linking NP to V by label "s". *kind* information of V is {AFFIRMATION} and {PRESENT}.)
0.3	V = give	(Concept "give", {AFFIRMATION} and {PRESENT}.)
0.4	V = have	(Concept "have", {AFFIRMATION} and {PRESENT}.)

Fig. 8 An example of MERGE heuristics(1).

Merging of grammar:

Confidence score	Surface structure	Semantics
0.7	S = NP V	(Linking NP to V by label "s". *kind* information of V is {AFFIRMATION, NEGATION} and {PRESENT, PAST}.)
0.5	S = NP V1	(Linking NP to V1 by label "s". *kind* information of V1 is {AFFIRMATION} and {PRESENT}.)
0.2	V = give	(Concept "give", {AFFIRMATION} and {PRESENT}.)
0.4	V1 = have	(Concept "have", {AFFIRMATION} and {PRESENT}.)

Merged grammar:

u	S = NP V	(Linking NP to V by label "s". *kind* information of V is {AFFIRMATION, NEGATION} and {PRESENT, PAST}.)
0.2	V = give	(Concept "give", {AFFIRMATION} and {PRESENT}.)
0.4	V = have	(Concept "have", {AFFIRMATION} and {PRESENT}.)

Fig. 9 An example of MERGE heuristics(2).

い、*kind* 情報とは、動詞の{肯定, 否定}や{現在, 過去}の区別を記録しておくものである。Fig. 8 と Fig. 9 に例を示す。Fig. 8 では、非終端記号 V1 を V に変えると表層部も意味部も全く同じになる。このことから二つの規則を一つにしてよいと判断する。すでに学習されている規則全体にわたって、V1 を V に書き換えて S=NP V1 の規則を削除する。Fig. 9 では *kind* 情報が異なっている場合を示してある。

4. 信頼度の定義とその計算方法

まず最初に、Dempster-Shafer 理論⁽²⁾について述べて、それから信頼度の定義とその計算方法について述べる。主観的確率値の計算には Dempster の結合規則を用い、これで信頼度を増減させる。

4.1 Dempster-Shafer 理論

Dempster は、Bayes 確率に適さない主観にかかわる不確実性を扱うために、下界、上界確率と名付けた新しい考え方を示した⁽⁸⁾。さらに Shafer は、Dempster 理論を洗練し、主観的な意味を付与するために元来の下界、上界確率をそれぞれ belief function, plausibility と言い換え、それらを基本確率を通じて定義した⁽⁹⁾。基本確率 $m(A_i)$ は部分集合 A_i に閉じ込められているが、Fig. 10 のような A_i 内の各点に自由に動ける半可動確率質量としてイメージを描くことができる⁽²⁾。

A_0 を有限な全集合、 $A_i (i = 1, 2 \dots)$ をその部分集合とすると、基本確率 $m(A_i) (i = 0, 1 \dots)$ は $[0, 1]$ の値をとり、次の条件を満たす。

$$m(\phi) = 0 \quad (\phi : \text{空集合})$$

$$\sum_{A_i \subseteq A_0} m(A_i) = 1 \quad (1)$$

$m(A_i) > 0$ のとき、 A_i は焦点要素と呼ばれる。

下界確率は基本確率を用いて

$$P(A_i) = \sum_{A_j \subseteq A_i} m(A_j) \quad (2)$$

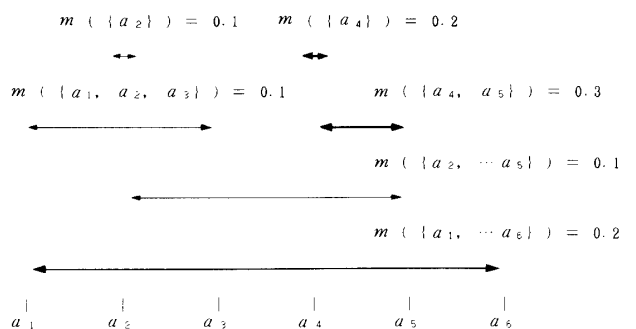


Fig. 10 Image of basic probability $m(A_i)$ ⁽²⁾. $A_0 = \{a_1, \dots, a_6\}$

と定義される。一方、上界確率は、

$$P^*, P.(A_i) = 1 - P.(\bar{A}_i) \\ = 1 - \sum_{A_j \in \bar{A}_i} m(A_j) \quad (3)$$

で定義される。

基本確率が、 $m(\{a_2\})$ と $m(\{a_4\})$ のように静的なものばかりであると、Dempster や Shafer の考え方は、Bayes 確率に縮退する。基本確率、下界、上界確率は加法性を満たさない。

重要な法則として Dempster の結合法則がある。これは、独立な証拠から推論された基本確率を統合する方法を与えている。 m_1, m_2 を独立な証拠に基づいて得られた基本確率とし、 A_{1i}, A_{2j} ($i, j = 0, 1, 2 \dots$) をそれぞれの焦点確率とする。この場合、以下に示す結合規則によって統合され、新たな基本確率が求まるとしている。このような確率の更新は、Bayes 確率ではできない。

$$m(A_k) = \frac{\sum_{A_{1i} \cap A_{2j} = A_k} m_1(A_{1i}) m_2(A_{2j})}{1 - \sum_{A_{1i} \cap A_{2j} = \phi} m_1(A_{1i}) m_2(A_{2j})} \\ (\text{ただし, } A_k \neq \phi) \quad (4)$$

4.2 信頼度の定義

主観的確率値 (Dempster-Shafer 理論の基本確率) のことを信頼度 (confidence score) と呼ぶことにする。信頼度は、 $[0, 1]$ の値をとるものと定義する。例えば、

信頼度	表層部	意味部
0.8:	V→get	(get の概念)

の規則では 0.8 だけ正しいと信頼でき、残り 0.2 は、(無知、不信) という意味を含んでいる。

なお本研究では、Dempster-Shafer 理論の中の下界、上界確率の概念は適用していない。

本研究では、信頼度を用いて最も確からしい解析木を選び出すメカニズムを採用し、また、信頼度が 0 に近い場合は 0 とみなして、その規則をメモリーから削除、信頼度が 1 に近いときは 1 とみなして、その規則は正しいと判断している。

4.3 信頼度の初期値と信頼度の増減の計算方法

信頼度の初期値の使用と信頼度を増加させたり減少させたりする場合の計算方法を(1)~(4)に示す。

(1)は、目の前で起こっている状況と、そのときに聞いた言葉の意味が一致しているかどうかについて扱ったものである。もちろん言葉の聞き間違いとか、見ている状況が別のものであったりすることも考えられる。(2)は、聞いた言葉の意味を理解するために知識(規

則)が使われ、またその言葉の意味を裏付ける別の状況などがあった場合に、その知識(規則)の経験数は高まり信用性も増すということを扱ったものである。(3)は、(2)の場合と逆で、目の前で起こっている状況と、聞いた言葉の意味が一致しなかった別の状況があった場合について扱ったものである。しかし、この原因を見つけることは正確には困難である。なぜなら、言葉の聞き間違いとか、異なった状況を見ていたとか、それとももっと複雑な文脈や要因などが考えられるからである。本研究では、入力される文と意味は正しいとし、解析した結果の意味が間違っているとして制限している。(4)は、知識(規則)を学習したときに、すでに学習している知識(規則)と“矛盾”するものであるかどうかについて扱ったものである。この“矛盾”は、学習時において、どれか間違った知識(規則)があるときや、多義語などのときに生ずる。これについても、正確に判断するのは困難である。なぜなら、その知識(規則)のもつ信用性や人間の確信によって判断され、さらにこのとき、いろいろな文脈や要因などを加味するからである。本研究では、シンプルな方法で処理している。実際に人間は、このような(1)~(4)に設定した種類のものに近い、感覚あるいは経験能力などをもっていて、これらによる判断基準で知識を学習しているように思われる。

(1) 文と意味のペアが入力された場合、構文・意味解析部がその文を解析できなかったときは、文と意味のペアに信頼度 α ($0 < \alpha < 1$) の初期値を付ける。以下に例を示す。

信頼度	表層部	意味部
α :	S→I GET	(I を GET にラベル s でリンク)

(2) 文と意味のペアが入力された場合、構文・意味解析部がその文を解析でき、入力の意味と一致した意味を生成する解析木が存在するときは、その解析木に適用されたすべての規則の信頼度を上げる。信頼度を上げるための初期値として、 β ($0 < \beta < 1$) を用いる。これは、対象となっている規則を肯定する、別の信頼度 β の証拠があったと仮定することに等しい。以下に計算方法を示す。

信頼度	表層部	意味部
0.7:	S→N V	(N を V にラベル s でリンク)

上の規則を“ A ”とすると、Dempster の式より、

$$m_1(A) = 0.7 \quad m_1(A, \bar{A}) = 0.3 \\ m_2(A) = \beta \quad m_2(A, \bar{A}) = 1 - \beta$$

例えば、 $\beta = 0.2$ とすると、

$$m_2(A) = 0.2 \quad m_2(A, \bar{A}) = 0.8$$

$$m(A) = \frac{0.7 * 0.2 + 0.7 * 0.8 + 0.3 * 0.2}{1.0 - 0.0}$$

$$= 0.76$$

よって規則は、

0.76: S→N V (NをVにラベルsでリンク)

ただし、入力の意味と一致した意味を生成する解析木が複数存在するときは、以下の計算により最も値の高い解析木を選び、その解析木に適用されたすべての規則の信頼度を上げる。Fig. 4を例にして以下に計算方法を示す。

$$NP1 = 0.3 * N1 = 0.3 * 0.2 = 0.06$$

$$NP2 = 0.4 * \min\{A, N2\}$$

$$= 0.4 * \min\{0.6, 0.5\} = 0.2$$

$$VP = 0.6 * \min\{V, NP2\}$$

$$= 0.6 * \min\{0.3, 0.2\} = 0.12$$

$$S = 0.8 * \min\{NP1, VP\}$$

$$= 0.8 * \min\{0.06, 0.12\} = 0.048$$

- (3) 文と意味のペアが入力された場合、構文・意味解析部がその文を解析できるが、入力の意味と一致した意味を生成する解析木が存在しないときは、その異なった意味を生成する解析木に適用されたすべての規則の信頼度を下げる。信頼度を下げるための初期値として、 γ ($0 < \gamma < 1$) を用いる。これは、対象となっている規則を否定する、別の信頼度 γ の証拠があったと仮定することに等しい。

以下に計算方法を示す。

信頼度 表層部 意味部
0.7: S→N V (NをVにラベルsでリンク)

上の規則を“A”とすると、Dempsterの式より、

$$m_1(A) = 0.7 \quad m_1(A, \bar{A}) = 0.3$$

$$m_2(\bar{A}) = \gamma \quad m_2(A, \bar{A}) = 1 - \gamma$$

例えば、 $\gamma = 0.5$ とすると、

$$m_2(\bar{A}) = 0.5 \quad m_2(A, \bar{A}) = 0.5$$

$$m(A) = \frac{0.7 * 0.5}{1.0 - 0.7 * 0.5} = 0.54$$

よって規則は、

0.54: S→N V (NをVにラベルsでリンク)

- (4) 互いに矛盾する規則が見つかった場合、それらの規則の信頼度を下げる。矛盾する規則とは、辞書的規則では表層部の右辺の単語が同じで意味部が異なっているもの、それ以外の規則では表層部の左辺も右辺も同じで意味部が異なっているものを示す。信頼度を下げるための初期値として、 γ ($0 < \gamma < 1$) を用いる。計算方法は各矛盾する

規則に対して(3)で示した計算をする。ただしこの計算が行われるのは、一般化するとき、新しく得られた規則とすでに学習している規則との間で矛盾が生じた場合に限られる。

ここで用いられた初期値 α , β , γ は、もし人間が同じ知識を二、三度学習したら、信頼度が1に近づくであろうという仮定に基づいた値を用いている。

4・4 一般化における信頼度の計算方法

学習部の一般化ヒューリスティックにおける信頼度の計算方法について述べる。

(1) AND ヒューリスティックにおける信頼度の計算

AND ヒューリスティックにおける動作の信頼度として初期値 CFand ($0 < CFand < 1$) を用いている。一般的に、AND ヒューリスティックは、文(単語列)を比較することによって、その中に共通する単語(列)の意味を推論するので誤動作を生じやすい。また、新しく得られた規則には、それを導出するために用いられた規則の信頼度の最小値を割り当てる、という基本的な考え方で計算をしている。Fig. 6を例にして以下に計算方法を示す。

入力の文と意味のペアの信頼度より、

$$x = 0.8, \quad y = 0.4$$

V→あげた(意味部: “あげた”の概念。以下【あげた】で表記)の規則の信頼度 c は、

$$c = \min\{x, y\} * CFand$$

$$= \min\{0.8, 0.4\} * 0.9 = 0.36$$

(CFand=0.9)

よって、

信頼度 表層部 意味部
0.36: V→あげた 【あげた】

NP→太郎は(意味部: “太郎”の概念。以下【太郎】で表記)の規則の信頼度 $x1$ は、

$$x1 = \min\{x, c\} = \min\{0.8, 0.36\} = 0.36$$

よって、

信頼度 表層部 意味部
0.36: NP→太郎は 【太郎】

NP→花子に(意味部: 【花子】)の規則の信頼度 $y1$ は、

$$y1 = \min\{y, c\} = \min\{0.4, 0.36\} = 0.36$$

よって、

信頼度 表層部 意味部
0.36: NP→花子に 【花子】

また、二つの文(単語列)の骨格を構成する規則は、それぞれ以下のとおりになる。

信頼度 表層部 意味部
0.8: S→NP V (NPをVにラベルsでリンク)

信頼度 表層部 意味部

0.4: S→NP V (NPをVにラベルsでリンク)
 なお、比較した文(単語列)の一方が一つの単語であるときは、CFandの値は、1.0として計算する。

(2) 部分解析ヒューリスティックにおける信頼度の計算

部分解析ヒューリスティックは、すでに学習している規則を使って部分的に解析を行うので誤動作を生じない。また、新しく得られた規則には、それを導出するために用いられた規則の信頼度の最小値を割り当てる、という基本的な考え方で計算をしている。Fig. 7を例にして以下に計算方法を示す。

入力文と意味のペアの信頼度より、

$$z = 0.8$$

VPまでの解析木に適用された規則、つまり、VP→NP2 V(意味部: NP2をVにラベルsでリンク)、NP2→犬を(意味部:【犬】)、V→もらった(意味部:【もらった】)の中で、最も信頼度の低い値を選び、その値をz2とする。

$$z2 = \min \{0.6, 0.2, 0.4\} = 0.2$$

NP1→太郎は(意味部:【太郎】)の規則の信頼度z1は、

$$z1 = \min \{z, z2\} = \min \{0.8, 0.2\} = 0.2$$

よって、

信頼度	表層部	意味部
0.2:	NP1→太郎は	【太郎】

また、対象となった文(単語列)の骨格を構成する規則は、以下のとおりになる。

信頼度	表層部	意味部
0.8:	S→NP1 VP	(NP1をVPにラベルsでリンク)

(3) マージヒューリスティックにおける信頼度の計算

マージヒューリスティックにおける信頼度の計算方法には2通りがある。Fig. 8とFig. 9を例にして以下に計算方法を示す。

Fig. 8では、信頼度0.6をもつ規則の表層部と意味部のカテゴリー名V1をVに変えると、信頼度0.7をもつ規則と全く同じになる。

信頼度0.7の規則を“A”とすると、Dempsterの式より

$$\begin{aligned} m_1(A) &= 0.7 & m_1(A, \bar{A}) &= 0.3 \\ m_2(A) &= 0.6 & m_2(A, \bar{A}) &= 0.4 \\ w &= m(A) \\ &= \frac{0.7*0.6+0.7*0.4+0.3*0.6}{1.0-0.0} = 0.88 \end{aligned}$$

Fig. 9では、信頼度0.5をもつ規則の表層部のカテゴリー名V1をVに変えると、信頼度0.7をもつ規則

の表層部と同じになり意味部も包含されている。つまり、意味部にあるV1やVの*kind*情報はOR関係でマージされるので、各規則の信頼度のmaxを計算する。

$$u = \max \{0.7, 0.5\} = 0.7$$

4.5 忘却係数の導入

人間は、その知識について信頼性がなかったり、使われなくなった場合、その知識を忘れてしまうことがある。そこで、このプロセスを示すために忘却係数を導入した。忘却は、学習が終わり次の文が入力されるまでの間に起こり、各規則の信頼度が下がる。ただし、文と意味の入力は連続的に行われるので、入力されるまでの時間に関係なく1度だけ信頼度を下げるというシンプルな構成になっている。忘却係数として、 δ ($0 < \delta < 1$)を用いる。計算方法は、規則の信頼度を下げる方法と同じであるが、 δ の値はわずかな値を用いる。

5. システムの実験例

ここでは、英語の例文を学習させた実験例について説明する。忘却係数 δ を低く設定して、連続的に規則の学習経験を積み重ねることで、規則の信頼度が増加していくプロセスを示したものである。入力データを与えるとき、文の時制や文型に関係なく、ランダムな順序で入力することが可能である。間違った規則や単語に関する情報を獲得しても最終的には、正しい規則を学習していく。Fig. 11にシステムへの入力データ(Snは第n文を、SEMは意味表現を示す。S7~S100については意味表現は略してある)、Fig. 12にその入力データから獲得された規則を示す(()の中は意味部を示す。カテゴリー名はわかりやすいように変えてある)。係数の初期値として、 $\alpha=0.7$, $\beta=0.5$, $\gamma=0.5$, $\delta=0.02$ を用いた。英文50までは、主に新しい単語や規則の獲得をしていて、それらの単語や規則の信頼度を上げるというような学習経験は少ない。S1とS2で、ANDヒューリスティックにより単語“I”, “you”, “went”と文型S+Vの規則を学習している。S3とS5で、単語“go”, “don’t”を学習している。このとき、“don’t”の役割も学習している。S4とS7で、単語“came”を学習している。このようにして、S50までには単語数32, 規則数20を獲得している。ここで、疑問文の文型Do+S+V(+NP1), Did+S+V(+NP1), Don’t+S+V(+NP1), Didn’t+S+V(+NP1)は、例えば、“Do+名詞句”+“動詞句”という規則で獲得されている。これは未知単語“do”が、

S 1 : I WENT	1.00: n1	-> I ((TOP *1)(GENE *1 I))
S E M : ((GENE *1 I)(GENR *2 GO)(TOP *2)	1.00: n1	-> YOU ((TOP *1)(GENE *1 YOU))
(PUTN *2 S *1)(*(2 (FORM AFFIRM)(TENSE PAST))(TYPE DECLAR))	1.00: n1	-> THEY ((TOP *1)(GENE *1 THEY))
S 2 : YOU WENT	1.00: n1	-> WE ((TOP *1)(GENE *1 WE))
S E M : ((GENE *1 YOU)(GENR *2 GO)(TOP *2)	0.98: n2	-> BOOK ((TOP *3)(GENO *4 BOOK))
(PUTN *2 S *1)(*(2 (FORM AFFIRM)(TENSE PAST))(TYPE DECLAR))	0.99: n2	-> PEN ((TOP *4)(GENO *4 PEN))
S 3 : I DON'T GO	0.99: n2	-> DOG ((TOP *4)(GENO *4 DOG))
S E M : ((GENE *1 I)(GENR *2 GO)(TOP *2)	0.96: n2	-> CAT ((TOP *4)(GENO *4 CAT))
(PUTN *2 S *1)(*(2 (FORM NEGA)(TENSE PRES))(TYPE DECLAR))	1.00: n2	-> BIRD ((TOP *4)(GENO *4 BIRD))
S 4 : WE CAME	0.99: v1	-> PUT ((TOP *2)(GENR *2 PUT)(*(2 (FORM AFFIRM)(TENSE PRES)))
S E M : ((GENE *1 WE)(GENR *2 COME)(TOP *2)	0.82: v1	-> PUT ((TOP *2)(GENR *2 PUT)(*(2 (FORM AFFIRM)(TENSE PAST)))
(PUTN *2 S *1)(*(2 (FORM AFFIRM)(TENSE PAST))(TYPE DECLAR))	0.94: v1	-> GIVE ((TOP *2)(GENR *2 GIVE)(*(2 (FORM AFFIRM)(TENSE PRES)))
S 5 : YOU GO	0.88: v1	-> GAVE ((TOP *2)(GENR *2 GIVE)(*(2 (FORM AFFIRM)(TENSE PAST)))
S E M : ((GENE *1 YOU)(GENR *2 GO)(TOP *2)(PUTN *2 S *1)	0.99: v1	-> WALK ((TOP *2)(GENR *2 WALK)(*(2 (FORM AFFIRM)(TENSE PRES)))
(*(2 (FORM AFFIRM)(TENSE PRES))(TYPE QUEST))	0.97: v1	-> LOOK ((TOP *2)(GENR *2 LOOK)(*(2 (FORM AFFIRM)(TENSE PRES)))
S 6 : DO THEY COME	0.97: v1	-> GET ((TOP *2)(GENR *2 GET)(*(2 (FORM AFFIRM)(TENSE PRES)))
S E M : ((GENE *1 THEY)(GENR *2 COME)(TOP *2)	0.88: v1	-> GOT ((TOP *2)(GENR *2 GET)(*(2 (FORM AFFIRM)(TENSE PAST)))
(PUTN *2 S *1)(*(2 (FORM AFFIRM)(TENSE PRES))(TYPE DECLAR))	0.99: v1	-> WANT ((TOP *2)(GENR *2 WANT)(*(2 (FORM AFFIRM)(TENSE PRES)))
S 7 : I CAME	0.98: v1	-> COME ((TOP *2)(GENR *2 COME)(*(2 (FORM AFFIRM)(TENSE PAST)))
S 8 : WE WALK	0.90: v1	-> CAME ((TOP *2)(GENR *2 COME)(*(2 (FORM AFFIRM)(TENSE PAST)))
S 9 : YOU DON'T GO	0.94: v1	-> GO ((TOP *2)(GENR *2 GO)(*(2 (FORM AFFIRM)(TENSE PRES)))
S 10 : I WALK	0.98: v1	-> WENT ((TOP *2)(GENR *2 GO)(*(2 (FORM AFFIRM)(TENSE PAST)))
S 11 : YOU WANT ED	1.00: ai	-> A ()
S 12 : YOU GET	1.00: ai	-> THE ()
S 13 : DID THEY COME	0.99: adj1	-> BLACK ((TOP *3)(GENE *3 BLACK))
S 14 : I WANT	0.98: adj1	-> WHITE ((TOP *3)(GENE *3 WHITE))
S 15 : I GET A RED PEN	0.96: adj1	-> BLUE ((TOP *3)(GENE *3 BLUE))
S 16 : DON'T THEY COME	0.96: adj1	-> RED ((TOP *3)(GENE *3 RED))
S 17 : YOU WANT THE WHITE DOG	0.98: dol	-> DO ()
S 18 : WE DON'T WANT	0.94: did1	-> DID ()
S 19 : I DIDN'T GIVE A RED PEN	1.00: didn't1	-> DIDN'T ()
S 20 : WE GIVE	1.00: don't1	-> DON'T ()
S 21 : DIDN'T THEY COME	0.97: ed1	-> ED ()
S 22 : YOU GOT	1.00: npl	-> x1 n2 ((TOP n2)(PUTN n2 COLOR x1))
S 23 : WE WENT	1.00: xl	-> al adj1 ((TOP adj1))
S 24 : I GOT THE WHITE BOOK	1.00: npl	-> al n2 ((TOP n2))
S 25 : THEY WALK	0.97: vl	-> vl ed1 ((TOP v1)(*(KIND* (v1(FORM(AFFIRM))(TENSE(PRES))))
S 26 : THEY COME	0.96: vl	-> didn't1 vl
S 27 : WE DIDN'T GO		((TOP v1)(*(KIND* (v1(FORM(AFFIRM))(TENSE(PRES))))
S 28 : YOU GAVE		(*TRANS* (didn't1:(v1(FORM(AFFIRM * NEGA))
S 29 : I COME		(TENSE(PRES * PAST))))%))
S 30 : THEY GAVE A RED DOG		
S 31 : YOU CAME		
S 32 : WE WANT THE RED PEN		
S 33 : I GIVE A BOOK		
S 34 : DO YOU GO		
S 35 : I GET THE CAT		
S 36 : THEY PUT		
S 37 : YOU GIVE A BLACK DOG		
S 38 : WE WALK		
S 39 : WE WANT THE DOG		
S 40 : I GOT THE BLUE PEN		
S 41 : DID YOU GET A CAT		
S 42 : THEY PUT THE PEN		
S 43 : I DON'T COME		
S 44 : DO THEY WANT A BOOK		
S 45 : WE GAVE THE BLACK DOG		
S 46 : DIDN'T YOU GET THE CAT		
S 47 : THEY PUT A WHITE BOOK		
S 48 : DON'T YOU GIVE A PEN		
S 49 : WE CAME		
S 50 : I GOT THE BIRD		
S 51 : THEY DIDN'T PUT A RED PEN		
S 52 : DID YOU WALK		
S 53 : I GET THE BLUE BOOK		
S 54 : WE WENT		
S 55 : THEY WANT ED A WHITE CAT		
S 56 : DON'T THEY GIVE A BIRD		
S 57 : YOU CAME		
S 58 : DO YOU COME		
S 59 : I PUT THE BLACK BOOK		
S 60 : WE GOT A DOG		
S 61 : THEY DON'T GO		
S 62 : DIDN'T YOU WANT THE BLACK DOG		
S 63 : I GAVE A WHITE BIRD		
S 64 : DON'T YOU WALK		
S 65 : THEY PUT A RED PEN		
S 66 : WE WENT		
S 67 : DIDN'T YOU GET		
S 68 : I WENT		
S 69 : DO THEY GIVE THE CAT		
S 70 : DID YOU PUT A PEN		
S 71 : WE PUT THE BLUE PEN		
S 72 : I GOT A RED BIRD		
S 73 : THEY WANT ED A DOG		
S 74 : DIDN'T YOU COME		
S 75 : I CAME		
S 76 : WE DIDN'T GET A BLUE BOOK		
S 77 : DO YOU GIVE THE RED BIRD		
S 78 : THEY DON'T PUT A PEN		
S 79 : DON'T YOU WALK		
S 80 : WE WENT		
S 81 : I LOOK A BIRD		
S 82 : DIDN'T THEY PUT THE BLACK CAT		
S 83 : WE GAVE THE CAT		
S 84 : DO YOU COME		
S 85 : I LOOK ED A BLACK BIRD		
S 86 : DID THEY GO		
S 87 : WE PUT THE WHITE PEN		
S 88 : DON'T YOU GET A DOG		
S 89 : DID THEY LOOK THE WHITE BIRD		
S 90 : I DIDN'T WALK		
S 91 : I GOT THE BLACK BOOK		
S 92 : YOU WENT		
S 93 : THEY LOOK ED A BLUE PEN		
S 94 : WE DIDN'T GIVE A DOG		
S 95 : WE WANT THE RED BIRD		
S 96 : DON'T YOU WALK		
S 97 : DO YOU PUT A BLUE BOOK		
S 98 : I LOOK ED A WHITE CAT		
S 99 : THEY DON'T COME		
S 100 : DIDN'T YOU GET A BLACK BIRD		

[$\alpha = 0.7, \beta = 0.5, \gamma = 0.5, \delta = 0.02$]

Fig. 11 Input data.

部分解析ヒューリスティックで冠詞と同じ扱いを受けたためである。50文の時点で、規則：名詞句→X1名詞(ただし、X1は冠詞+形容詞)と規則：文→名詞 動詞句の信頼度は1になっている。

Fig. 12は、英文100までに獲得された規則の結果である。単語“look”を獲得している以外はすべて、学習経験を積ませて規則の信頼度を上げることをしている。その結果、規則の信頼度は高くなっている。特に頻繁に使った単語“I”, “you”, “they”, “we”, “a”, “the”や規則：文→名詞 動詞句, 文→名詞 動詞句 名詞句, 名詞句関係の規則などは信頼度が1.0になっている。S100までに獲得された単語数は33, 規則数は20である。学習時間は、S1からS50までが64分, S51からS100までが17分であった。ト

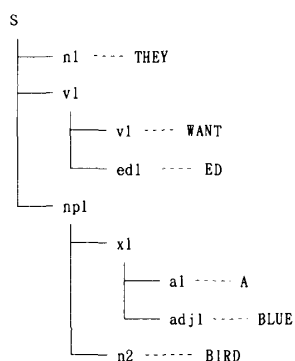
Fig. 12 Learning results of grammar.

タル時間は81分である。なお、信頼度は小数点以下3桁で四捨五入してあるので、信頼度が1.0になれば下がることはないが、より柔軟な学習プロセスを考えるとときは四捨五入をしないほうがよいと思われる。

この実験例では正しい規則を学習していくが、もし間違った規則を学習しても忘却できるようになっている。つまり、その規則が使われなくなれば信頼度は下がり、0に近づく。ここで信頼度を0とみなせば間違った規則を忘却できるわけである。

INPUT : THEY WANT ED A BLUE BIRD

PARSING TREE :



MEANING :

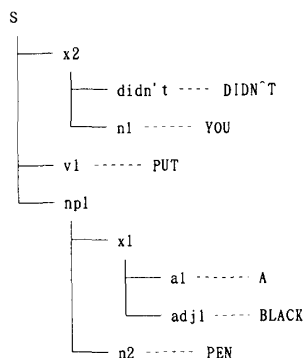
((GENE *1 THEY)(GENR *2 WANT)
(GENE *3 BLUE)(GENO *4 BIRD)
(TOP *2)(PUTN *2 S *1)
(PUTN *2 OBJ *4)
(PUTN *4 COLOR *3)(TYPE DECLAR)
(*2 (FORM AFFIRM)(TENSE PAST)))

CONFIDENCE SCORE : 0.94

Fig. 13 An analysis example by using acquired grammar in Fig.12 Input : "They wanted a blue bird."

INPUT : DIDN'T YOU PUT A BLACK PEN

PARSING TREE :



MEANING :

((GENE *1 YOU)(GENR *2 PUT)
(GENE *3 BLACK)(GENO *4 PEN)
(TOP *2)(PUTN *2 S *1)
(PUTN *2 OBJ *4)
(PUTN *4 COLOR *3)(TYPE QUEST)
(*2 (FORM NEGA)(TENSE PAST)))

CONFIDENCE SCORE : 0.88

Fig. 14 An analysis example by using acquired grammar in Fig.12 Input : "Didn't you put a black pen?"

学習が終わった状態で、システムに文を入力するとその文の意味が得られる。獲得された Fig. 12 の規則を用いて入力文を解析した例を、Fig. 13, Fig. 14 に示す。もちろん、"I GOT ED A BLACK PEN", "YOU GIVE THE CAT RED", "WE DIDN'T GAVE A BOOK" などの文は、解析不能というメッセージが出される。

この実験例では簡単な文型を扱っているが、例えば "give" という単語が二重目的語をとったり、"to" + 名詞の文型をとっても、その文を解析できる規則を獲得可能である。

6. む す び

本研究は、基本的に Anderson の理論や Selfridge

の理論に基づいていて、それらの問題点を改良し、さらに信頼度を導入することによって柔軟性をもった確率的・帰納的学習システムを実現し、その方法と実験結果について述べた。研究の結果、わかったことを以下に示す。

- (1) 意味に基づく一般化は、かなりの文法を推論でき、単語を知らない状態から始めても、あまり大きな問題は起きなかった。また、時制や否定・疑問の規則も学習できた。
- (2) 確率的に規則を学習することによって、従来の真か偽かの2値で処理するのは違い、曖昧性を考慮しながら連続的に学習経験を積むプロセスをモデル化することができた。
- (3) 本稿では示さなかったが、英語以外に日本語の構文・意味解析規則も同様に帰納的に学習できることがわかった⁽¹⁾。

また、信頼度付き学習システムの問題点は、以下のとおりである。

- (1) 文と意味のペアが入力された場合、その文を解析できるが、入力の意味と異なった意味を生成する解析木が存在したとき、この原因となった規則を正確に見つけることは容易ではなく、解析木に適用されたすべての規則の信頼度を一律に下げている。
- (2) 規則の信頼度の増減や忘却に初期値を設定したが、厳密には人間が持つ主観や学習能力、環境などによって影響されるものであり、固定された値を持つというような単純なものではない。
- (3) 一般化ヒューリスティックでは、新しく獲得された規則の信頼度を常に最小限に見積っているが、人間は悲観的に考えるときもあれば、楽観的に考えるときもある。
- (4) 一般化ヒューリスティックが適用できるときはいつでもすぐに適用してしまい、すべての可能な規則の仮定をするに至っていない。

以上の問題を解決するには、三つの一般化ヒューリスティックの改良のほかに、さらに高度な学習機能が必要である。しかし、規則の信頼度を利用して、文のパーズングや生成、階層的なメモリー構造(高速データベース検索)、一般化ヒューリスティックを動作させる判断基準へ応用でき、それによって本システムの問題点も改善できると思われる。

◇ 参 考 文 献 ◇

- (1) 山本, 中川 : 自然言語の構文・意味解析規則の帰納的学習システム, 情報処理学会「自然言語処理」, Vol. 55-5 (1986. 5).
- (2) 石塚 満 : Dempster & Shafer の確率理論, 電子通信学

- 会誌, Vol. 66, No. 9, pp. 900-903 (1983).
- (3) Anderson, J. : Computer Simulation of a Language Acquisition System, in *Information Processing and Cognition*, R. Solso (ed.), Erlbaum (1975).
- (4) Pinker, S. : Formal models of language learning, *Cognition*, Vol. 7, pp. 217-283 (1979).
- (5) Selfridge, M. : A computer model of child language learning, *Artificial Intelligence*, Vol. 29, No. 2, pp. 171-216 (1986).
- (6) Norman, D. A. and Rumelhart, D. E. (eds.) : *Explorations in cognition*, W. H. Freeman (1975).
- (7) 田中穂積, ほか : 自然言語のためのプログラミングシステム—拡張 LINGOL について—, 信学論 D, Vol. 60-D, No. 12, pp. 1061-1068 (1977).
- (8) Dempster, A. P. : Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*, Vol. 38, pp. 325-339 (1967).
- (9) Shafer, G. : *A Mathematical Theory of Evidence*, Princeton Univ. Press (1976).
- (10) 若原, 中川 : 自然言語の時制・否定・疑問の文法と意味解析規則の帰納的学習システム, 人工知能システムの枠組みシンポジウム論文集, pp. 85-94 (1987. 11).
- (11) 若原一彰 : 自然言語の構文・意味解析規則の主観的確率を用いた帰納的学習法, 豊橋技術科学大学修士論文 (1988. 2)

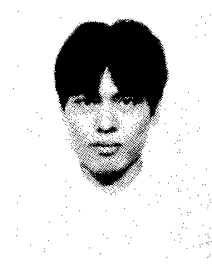
〔担当編集・査読委員：石崎 俊〕

著者紹介



中川 聖一 (正会員)

昭和 51 年京都大学大学院博士課程修了。同年, 京都大学情報助手。昭和 55 年豊橋技術科学大学情報工学系講師。昭和 58 年助教授。工学博士。昭和 60~61 年カーネギー・メロン大学客員研究員。音声情報処理, 自然言語処理, 人工知能の研究に従事。昭和 52 年電子通信学会論文賞受賞。著書: 「情報基礎学詳説」(分担執筆, コロナ社), 「確率モデルによる音声認識」(電子情報通信学会) など。電子情報通信学会, 情報処理学会, 日本音響学会, IEEE 各会員。



若原 一彰

昭和 61 年宇都宮大学情報工学科卒業。昭和 63 年豊橋技術科学大学大学院修士課程情報工学専攻修了。現在, 日本電装(株)勤務。在学中は自然言語処理の研究に従事。情報処理学会会員。