

自然言語の文法と意味解析規則の帰納的学習システム†

中川聖一† 山本幹雄† 若原一彰†

自然言語処理は、1970年代から盛んに研究されているが、実用化するためには多くの問題点がある。その中の一つに、自然言語処理に必要な知識量が膨大であるという問題点がある。ここで、もし必要な知識データをある程度自動的に獲得できれば、データを作る負担が軽減できる。そこで、本研究では、自然言語の文法と意味解析規則を、文とその意味の例から帰納的に学習するシステムを開発した。本システムは、Andersonの言語獲得の理論を用いており、彼が理論を実証するために作った、LAS (Language Acquisition System) を改良・拡張したものとなっている。改良・拡張したのは以下の点である。(1)単語の意味も知らない状態から学習できるようにした。(2)意味表現にセマンティック・ネットを用い、複数の意味を持つ単語も学習できるようにした。(3)学習のためのヒューリスティックを改良した。第一に、一般化しきりにしきりがない場合を、第二に、意味を持たない単語に関する文法を学習するときの問題を改良した。(4)時制の学習と疑問文の学習ができるようにした。

1. まえがき

自然言語処理は1970年代からさかんに研究されているが、実用化されるためには多くの問題がある。問題点の一つに、自然言語処理に必要な知識の膨大さがある。そのため現在の自然言語処理システムは、領域を限定して限られた知識で動くようを作られており、その知識を入れ替えることによって多くの領域に適用できる。しかし、このように領域を限定しても、なお知識作るのは膨大な作業となる。ここで、もし必要な知識をある程度自動的に獲得できればユーザの負担は軽減できる。自然言語処理の応用システムで、ユーザが「こういう文が入力されたら、このような意味表現になってほしい」と考えた場合、例の入力として個別の単語の意味まで入れる必要がなく、時制、疑問文、多義語まで学習できることは有意義である。そこで、われわれは自然言語処理システムに必要な、構文・意味解析規則を解析例から帰納的に推論し学習するシステムを試作した。以前にも、構文規則を帰納的に学習する研究は行われている。

AndersonのLAS^{1),2)}は、文と意味表現のペアを入力されると、現在ある文法(ATNG)を一般化して、文から意味表現への解析、あるいは意味表現から文の生成がより一般的にできるようになるシステムである。Reeker³⁾のPSTは赤ちゃんが聞き取った文と、

同じ意味で赤ちゃんが内的に発話した文との差異を縮めるという問題を解決することによって、文法を学習する。高木ら⁵⁾のシステムは視覚入力と言語入力の2系統の入力を備え、視覚情報を手がかりにして単語の意味と文法を学習していくシステムである。

本研究では、学習用の例として、自然言語文とその意味表現を入力する。

本研究の理論的背景としては、認知心理学における赤ちゃんの言語獲得のAndersonの理論がある¹⁾。彼の理論の中心となる主張は文法規則のほとんどが意味構造から引き出されるということである。すなわち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言っている。彼のLAS (Language Acquisition System) というシステムは文とその意味が入力されると、その文の解析を試みる。解析が失敗したら、ある精密度のヒューリスティックによって、文法は訓練例を解析でき正しい意味を付与できるように拡張される。しかし、このシステムをそのまま自然言語処理に応用するには、以下の問題点がある²⁾。

(a) 文法を学習するシステムであり、単語の学習ができない。

(b) 意味表現としてHAM構造を用いている。HAM構造は命題を2進木の形で表現する。2進木の葉は有意義単語である。有意義単語は一つの概念を表しており、それが意味表現の中で使われるということは複数の意味を持つ単語が扱えない。

(c) 規則を一般化しきりにしきりがない場合や、意味をもたない単語に関する規則を学習する場合など、ヒューリスティックにいくらかの問題点がある。

† An Inductive Learning System of Syntactic and Semantic Analysis Rules of Natural Language by SEIICHI NAKAGAWA, MIKIO YAMAMOTO and KAZUAKI WAKAHARA (Department of Information and Computer Sciences, Toyohashi University of Technology).

†† 豊橋技術科学大学情報工学系

さらに本システムでは、時制の学習（現在形・過去形）と疑問文の学習ができる。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 理論がある⁷⁾。

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、この理論にもいくつかの問題点がある。

(a) システムは、ユーザからのフィードバックを受けて学習しているが、実際問題としてこのような CHILD のやり方で推論し、文法や単語の意味を決めているのかどうか疑問である。

(b) CHILD が学習する文法は単語特定のものであり、言語的な一般化をしていない。

(c) “the”などの意味のない単語は扱っていない。

本研究では意味表現に一般的なセマンティック・ネットワーク、構文解析に文脈自由文法 (CFG) を用い、単語の意味も知らない状態から構文・意味解析規則を、解析例から帰納的に学習するヒューリスティックを開発した。このヒューリスティックでは、LAS や CHILD の問題点を改良してある。どんな言語でも学習できることを目標としている。

2. システムの構成

図1に本システムの構成を示した。学習部は文と意味のペアを受け取り、新しい構文・意味解析規則を推論して古い構文・意味解析規則を変更する。推論するときに構文・意味解析部を使用するヒューリスティックもあるので、図1のような表現になっている。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。

意味表現には現在の自然言語処理で一般的に使用されているセマンティック・ネットワーク⁴⁾を用いた。

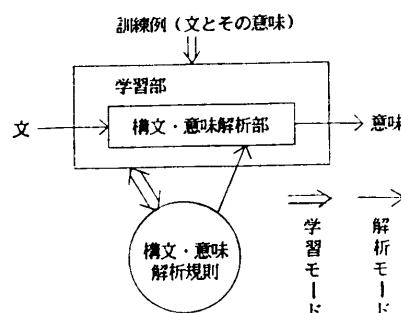


図1 本システムの構成
Fig. 1 Blockdiagram of learning system.

さらに、学習部に送られる意味表現は、解析部が意味表現を構成するときを使われる命令に分解される。基本的には、ノードを発生する命令と、ノードを接続するリンクの発生の命令に分解される。図2に命令の種類をあげ、図3に分解の例を示す。解析部には拡張

ノードを発生する命令

GENDR :	述語を発生 (図4の " " に対応)
GENO :	対象を発生 (図4の "< >" に対応)
GENE :	概念を発生 (図4の "[]" に対応)

リンクする命令

PUTN :	2つのノードがある ラベルでつなげる。
--------	------------------------

図2 意味表現を作る命令

Fig. 2 Operators for semantic representation.

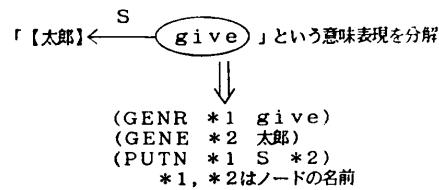


図3 意味表現の分解

Fig. 3 Example of semantic representation.

解析規則

表層部	意味部
S→NP VP	NPをVPにラベルSでリンク
VP→V NP	NPをVにラベルobjでリンク
V→gave	giveノードを発生
NP→the N	Nの意味
NP→N	Nの意味
N→Mary	[Mary]ノードを発生
N→dog	<dog>ノードを発生

解析例 (上の規則で)

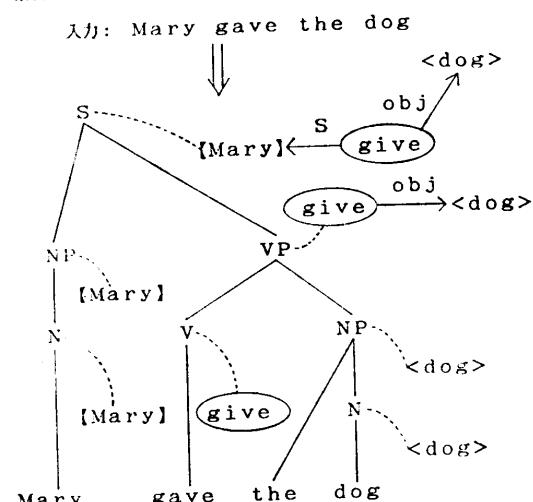


図3 意味表現の分解

Fig. 3 Example of semantic representation.

LINGOL⁶⁾ を改造したものを使用している。主な改造は以下のとおりである。

(1) 意味解析部が構文解析と同時に実行できるようにし、意味解析の失敗からもバック・トラックがかかるようにした。

(2) 解析結果としての意味表現に、セマンティック・ネットワークを用いるために、そのように意味解析部を改造した。

解析規則と解析例を図4に示す。

学習部は文とその意味としてのセマンティック・ネットワークを受け取り、まず文の解析を試みる。この解析は与えられた意味とは関係なしに、その時システムが持っている規則のみで行われる。その結果が与えられたセマンティック・ネットワークと同じになれば、学習部はなにもしない。しかし解析が失敗するか、または解析結果が違っていればシステムはまず与えられた文だけを解析できる特殊な規則を作る（暗記学習）。その後にその暗記学習したものと、もともと持っていた規則をデータとして、三つの学習ヒューリスティックを使って一般化を行い、より広い範囲の文を解析できるようにする。この一般化の手続きは「学習のヒューリスティック」の章で詳しく論じる。またシステムは、文例が一つでも手に入りしだい処理を行えるように、1文例ごとの処理を行っている。

3. 学習のヒューリスティック

本研究で使った学習部の一般化のヒューリスティックを具体的に述べる。以下に挙げるよう三つのヒューリスティックがある。

- (1) AND ヒューリスティック
- (2) 部分解析ヒューリスティック
- (3) マージヒューリスティック

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三つである。

- (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時にループしてはならない。
- (3) 意味表現中の視点となっているノード（トップ・ノード）は明示的に示されなければならない。

帰納的推論では可能な推論結果の制約がきびしいほど早く、正確に推論できる。CFG は完全ではないが、ほとんどの自然言語の文を解析できるので、自然言語処理ではよく使われている。また CFG を文法とする

ことで、自動的に 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。もしこの仮定をしなければ、文法が学習不能になることを Anderson は示唆している²⁾。

本システムは意味構造と文の解析木が似ているという Anderson の主張を採用しており、それを利用して文法を学習するため、意味構造は木構造であることとする。たとえば、意味構造としてネットワーク構造を採用した場合、解析木との比較が困難になる。

本システムは意味表現の形から文の木構造を推論するが、そのとき木構造の根が意味表現中のどのノードに対応するか知らないと、推論される木構造が曖昧になる。それをさけるためにトップ・ノードが木構造の根に対応するものとして^{1), 2)}、学習を行う。このため、トップ・ノードは意味表現中に示されていなければならぬ。

3.2 AND ヒューリスティック

このヒューリスティックによって、単語の意味を決めることができる。解析できない文は、まず暗記学習される。すなわち、その文しか解析できない1個の文法が作られる。そのような二つの文が暗記学習されており、それらの表層文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考えてよい。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、文全体の意味のトップ・ノードを共通部分の意味が含んでいる場合は、そのトップ・ノードが共通部分と他の部分をつなぐリンクに接続していなければならない。こうして、より小さな単語列へ分解してゆき、最後には一つの単語の意味がわかる。

共通部分の意味と単語列が結びついたので、次に各々の文の共通部分でない意味は共通部分でない単語列から生成されたものだとみなす。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付ける文法を作る。分割されていれば、そのような文法は作らない。

さらに、分解した単語列を一つにまとめる文法を作り、このヒューリスティックは終わる。図5に例を示す。

す。

3.3 部分解析ヒューリスティック

文を全部は解析できないが部分的には解析できるとき、このヒューリスティックが使われる。一つの部分単語列が解析され残りの部分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。解

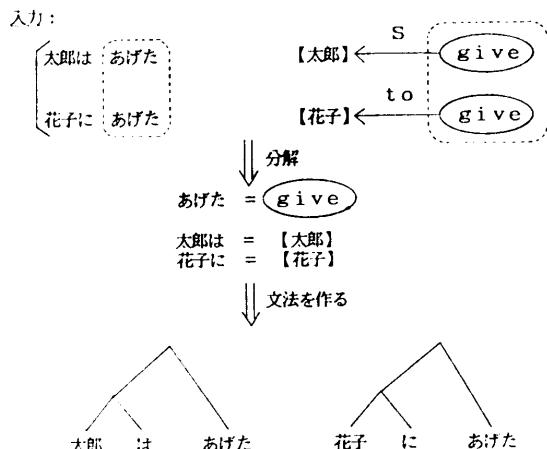
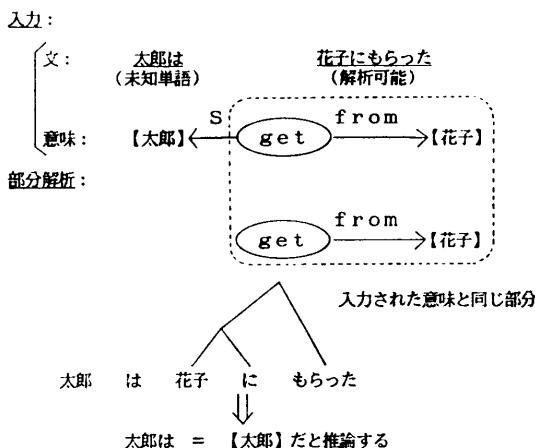


図 5 AND ヒューリスティックの動作例
Fig. 5 Example of AND heuristics.



文法の提案：

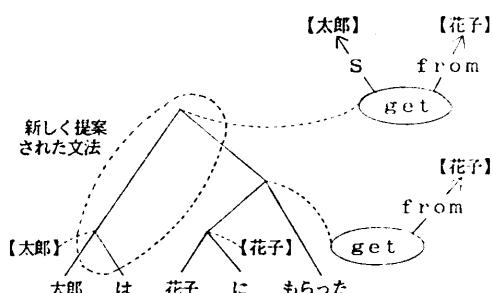


図 6 部分解析ヒューリスティックの動作例
Fig. 6 Example of partial analysis heuristics.

析できた単語列と意味が処理の対象になる。また、二つの部分単語列が解析されそれぞれの意味でもとの文の意味を作ることができる時は、それらをまとめて、その文の意味になるような文法を作る。その時、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。これらの過程の具体例を図 6 に示す。

意味を持たない単語が発見された時は、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・ネットワーク上で動作など関係を表すノード以外に対応する（すなわち名詞などに対応するノード）単語の方に意味を持たない単語を結びつける。これは、日本語の助詞、英語の前置詞がそれぞれ名詞句あるいは前置詞句を形成することが多いからである。また、両どなりのどちらも指されていたら、解析木の枝がその意味を持たない単語のすぐ上をとおる単語に結びつける。これは LAS のように常に右の単語に結びつけるより、妥当性のある文法を生み出す可能性が、日本語と英語の場合は高い。

3.4 マージヒューリスティック

AND ヒューリスティックと部分解析ヒューリスティックは、暗記学習された文法をより構造的に解析できるように特殊な文法を付け加えていくだけである。ある程度の一般化はそのヒューリスティック内に内在しているものの、まだ一般化の余地はある。その一般化をするのがマージヒューリスティックである。これは、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文法がある時、それらの文法を一つの文法にまとめるヒューリスティックである。具体的に、どうなるか以下に示す。つぎの四つの文法があるとする。

- (1) $S = NP\ V$ (意味部 : NP を V にラベル S のリンクでつなげる)
- (2) $S = NP\ V1$ (意味部 : NP を $V1$ にラベル S のリンクでつなげる)
- (3) $V = give$ (意味部 : $give$ の概念)
- (4) $V1 = have$ (意味部 : $have$ の概念)

これは $give$ と $have$ が、まだ一つのカテゴリーになっていない状態である。マージ・ヒューリスティックによって、一つのカテゴリーにまとめられる。(2)の $V1$ を V に変えると、文法も意味も完全に(1)と同じになる。このことは二つの文法を一つにしてよいと

マージ・ヒューリスティックは判断する。V1をVに規則全体にわたって書き換えて、(2)を消す。文法は次のように変更される。

- (1) $S = NP \ V$ (意味部: NPをVにラベルSのリンクでつなげる)
- (3) $V = give$ (意味部: giveの概念)
- (4) $V = have$ (意味部: haveの概念)

4. 時制と疑問文の学習法

ここでは、時制の学習と疑問文の学習についてその方法を示す。前章で述べたANDヒューリスティック、部分解析ヒューリスティック、マージヒューリスティックの三つのヒューリスティックを時制と疑問文が学習できるように拡張する。

4.1 意味表現と入力形式

まず、時制（現在形・過去形）と平叙文、疑問文を区別する意味表現を示す。

- (# (TENSE PRES)) : 現在形
- (# (TENSE PAST)) : 過去形
- (DECLAR #) : 平叙文
- (QUEST #) : 疑問文

ここで、#はノードの名前を表す。ノード“#”は、現在形、過去形あるいは平叙的、疑問的であることを意味する。

次に、英語の規則変化する動詞をもつ過去形の文を入力する場合の形式は、

I look ed

のように、“looked”を“look”と“ed”に分割して“look ed”とする。これは、SelfridgeのCHILD理論⁷と同じ考え方を採用している。つまり、実際に単語が話されるときは単語間にスペースがないので、“looked”と“look ed”との間には、相違はないと考えられる。

また、日本語の場合は、動詞の現在形から過去形への変化において、あまり規則的な変化が見られないので、本研究では英語の場合の“give”と“gave”的に不規則変化した別の単語として扱っている。別の単語とすると効率は悪くなるが、意味に関する規則は同じものを学習するので問題はない。

4.2 AND・部分解析ヒューリスティックを用いた時制の学習例

まず、ANDヒューリスティックによる“ed”的学習を図7に示す。

次に、部分解析ヒューリスティックによる“ed”的

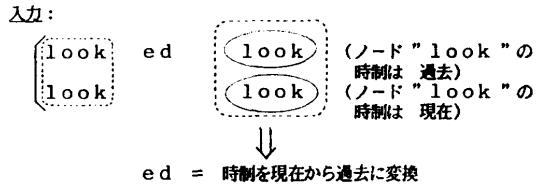


図7 ANDヒューリスティックによる“ed”的学習
Fig. 7 Learning example of “ed” by AND heuristics.

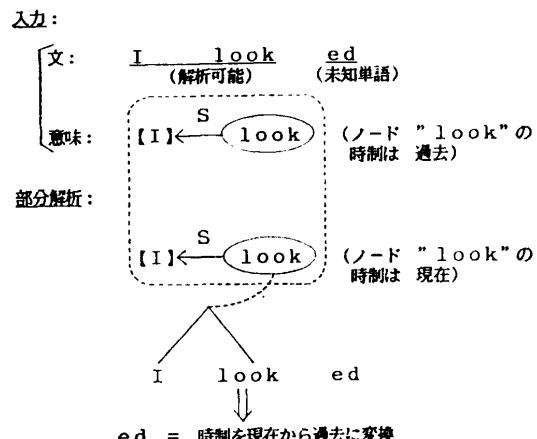


図8 部分解析ヒューリスティックによる“ed”的学習
Fig. 8 Learning example of “ed” by partial analysis heuristics.

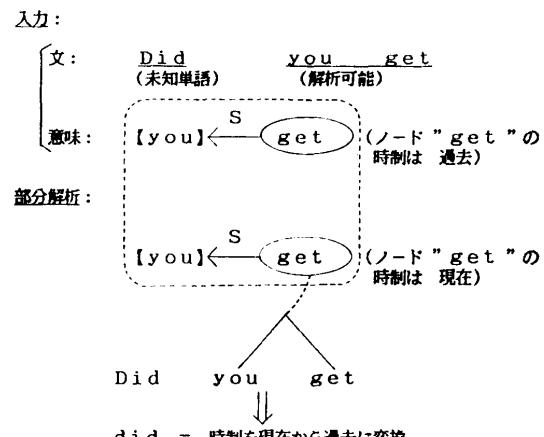


図9 部分解析ヒューリスティックによる疑問文の“did”的学習

Fig. 9 Learning example of “did” by partial analysis heuristics.

学習を図8に示す。解析できた単語“look”に時制の違いを生じ、解析できなかった単語は“ed”であるから、“ed”は時制を現在から過去に変えるものであると推論する。

また、部分解析ヒューリスティックによる疑問文の“did”的学習を図9に示す。解析できた単語“get”に

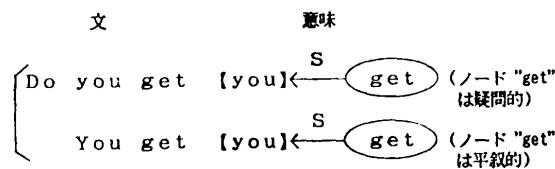


図 10 疑問文の学習

Fig. 10 Learning example of an interrogative sentence.

解析規則

表層部

* V → V1 ED1
 V1 → look
 ED1 → ed

*の規則の意味部の情報

```
(% (*kind* (V1 (tense (pres))))  

(*trans*  

(ED1 :  

(V1 (tense (pres * past))))))%)
```

図 11 情報 *kind*, *trans* を用いた実際の例

Fig. 11 Action example of *kind* and *trans*.

時制の違いを生じ、解析できなかった単語は “did” であるから、“did” は時制を現在から過去に変えるものであると推論する。

4.3 平叙文と疑問文の相違の学習例

学習の例を図 10 に示す。暗記学習されている図 11 の二つの文を比較した時、両文の意味は、ノード “get” が疑問的、平叙的であることを除けば同じである。また、文中において異なるのは、文頭の意味のない単語 “do” だけである。このことから単語 “do” は、文の意味を平叙的なものから疑問的なものに変えると推論する。ただし本システムでは yes/no 型の疑問文のみを対象としている。

4.4 時制に関する情報

学習された構文・意味解析規則に時制に関する情報 *kind* と *trans* をもたせる。その役割を以下に示す。

kind：その規則がもつ時制を記述

trans：どの規則が、どの規則の時制をどのように変換するのかを記述

これらを使った具体例を図 11 に示す。

以上二つの情報を意味部に取り入れたことにより、マージヒューリスティックの基本動作に以下に示すような条件が加わる。

マージの対象となった二つの規則の意味部において

- ・両方とも *kind* の情報を持っていない場合
- 両方とも *kind* の情報を持っていて、しかも同じ規則のもつ時制の情報である場合、マージで

きる。

- ・両方とも *trans* の情報を持っていない場合か、両方とも *trans* の情報を持っていて、しかも全く同じ内容である場合、マージできる。

この結果、二つの情報を意味部に加えることによって、入力文を解析するときに用いる規則に制約を与えることができる。たとえば、図 11において、規則 V → V1 ED1 の *kind* の情報は、V1 の規則の時制が現在でなければならないことを意味する。また、ED1 の *trans* の情報は、V1 の時制を現在から過去に変換することを意味する。

5. 学習ヒューリスティックの評価

帰納的学習とは、特殊なデータから一般的なデータを推論する過程であるから、その評価は、正しい一般化ができているか否かと、一般化の程度 (over generalization と under-generalization) によってなされる。さらに Anderson の LAS との比較も行う。

5.1 AND ヒューリスティックの評価

このヒューリスティックの誤った動作は以下の三つが考えられる。

- (1) 単語列に意味を付けすぎる。
- (2) 単語列に付けた意味が足りない。
- (3) 単語列に間違った意味を付ける。

表層と意味が 1 対 1 で対応している単語と多義語だけの時はうまくいくが、同義語をいくつか持つ単語が単語列にはいっている場合に上であげた誤った動作をする可能性がある⁹⁾。たとえば「みかんをあげた」と「オレンジをあげた」という文がまったく同じ意味表現を持っていたとするとき、その意味は「をあげた」にすべて対応づけられ（意味を付けすぎる）、「みかん」と「オレンジ」は意味を持たない（意味が足りない）と結論してしまう。また、同じ文で、動詞「あげた」の意味がそれぞれ「持ち上げる」と「与える」という異なる意味を持っており、「みかん」と「オレンジ」が同じ意味だとすると「をあげた」に「みかん」の意味が、「みかん」に「持ち上げる」の意味が、「オレンジ」に「与える」の意味が対応づけられる（間違った意味）。

この問題を解決するためには、AND ヒューリスティックで使用するデータを多くすることが考えられる。すなわち、いまのヒューリスティックでは二つの文しか使っていないが三つ以上の文を使用すると安全性が高まる。しかし現実には、使用するデータを増や

すとすべての文に対してこのヒューリスティックを使う時に、組み合わせ的爆発がおこるため、よい方法とはいがたい。本システムでは、以前すでに入力されている単語の同義語となる単語を学習する時は、以前に入力された単語が十分に学習され意味が確立しているか、あるいはその単語が入っている単語列の他の単語はすでに解析可能となっていることを条件として、この問題を回避している。要するに本システムが持っている文法が増えれば増えるほどこの条件を犯す確率は減るし、AND ヒューリスティックのまえに必ず部分解析ヒューリスティックが起動されるようにすれば、問題は起きなくなる。

5.2 部分解析ヒューリスティックの評価

部分解析ヒューリスティックは、基本的には特殊な文法を提案するヒューリスティックであり、大きな一般化は行わないで一般化しすぎる問題は起きない。

一般化し足りない場合は、部分解析の能力を制限から二つ生じる。一つは、一つの文を三つ以上に分けて部分解析を行わぬために生じる。一つの文が三つ以上に分解できそれが部分解析できる場合は、三つ以上の部分を一つにまとめるべきだが、本システムでは行っていない。他の一つは、同じ部分単語列が文法的に曖昧さを持っており、かつ導かれる意味表現は同じで正しいものだった場合に生じる。本システムは曖昧さのうち、はじめに解析に成功したものだけしか処理しない。他の解析による文法の提案の可能性を無視することになる。ただし、これらの問題点は実験システムの複雑さの制限で解決を試みなかった問題であり、本質的な問題ではないとわれわれは考えている。

このヒューリスティックは、Anderson のヒューリスティックの一般化が足りない場合を改善している。Anderson のヒューリスティックが一般化できない例は次の文のような場合である。

The cop frightens the thief.

The cop is frightened by the thief.

The cop tends to like thieves.

“The cop”は文上の同じ構成要素でありながら意味表現での役割が違うため、LAS の「意味表現での役割が同じ文法は一つにまとめてよい」というヒューリスティックでは一般化できない²⁾。そのため、上の例文の “The cop” は全部違う文法でそれぞれ解析される。しかし部分解析ヒューリスティックは、前に学習した文法が使えるときはそれを利用して新しい文法を作っていくのでこのような問題は起こらない。

5.3 マージヒューリスティックの評価

まず、一般化のしすぎから検討していく。このヒューリスティックが一般化しすぎる例をあげる。次のような文のための文法が学習されていたとする。

S + gave + 人 + 物 (1)

S + gave + 物 + to + 人 (2)

この文法は同じ意味を生成する。ここで次のような文のための文法が学習されたとする。

S + donated + 物 + to + 人 (3)

(2) と (3) の文法が比較され gave と donated が同じクラスにマージされる。すると、次のような文も許されることになる。

S + donated + 人 + 物 (4)

(4) のような文は英語では許されず、一般化をしそうになる。この問題は微妙な意味の違い、あるいは感覚の問題であるため解決はむずかしく、LAS でも解決はされていない。LAS で問題となった、他の一般化のしすぎは次のような場合である。名詞句を学習している時、学習している言語では square は名詞にも形容詞にもなるが red は形容詞にしかならないと仮定する。the square thing, the red thing, the square は名詞句になるが the red は名詞句にならない言語である。Anderson のシステムでは the red が許される文法を作ってしまう²⁾。本システムでは、この一般化のしすぎは起こらない。しかし、“ed”を学習したために不規則変化する動詞、たとえば “go” という単語においてその過去形が “goed” となる規則を獲得してしまう一般化のしすぎが起こる。これは、Selfridge の CHILD 理論の中で、3歳ぐらいの子供は動詞の過去形を獲得していく過程において、一時期、不規則変化する動詞に “ed” をつけて言葉を話すと述べられている現象と同じである。これは、“go” に対してその過去形 “went” の例が与えられたとき、この特殊な規則を優先すれば解決できると考えられる。

次に一般化が足りない場合であるが、これは意味ネットワークのリンクのラベル名の不統一から生じる。たとえば、give のリンクのラベル名と、have のリンクのラベル名を全部違えたら、give と have は永遠に一つのクラスにはマージされない。

6. システムの学習例

英語の実際の学習例を説明する。図 12 にシステムへの入力、図 13 にその入力より得られた文法を示

文 1 : I GAVE
 意味 : ((GENE *1 I)(GENR *2 GIVE)(TOP *2)
 (PUTN *2 S *1)(*2 (TENSE PAST))(DECLAR *2))
 文 2 : YOU GAVE
 意味 : ((GENE *1 YOU)(GENR *2 GIVE)(TOP *2)
 (PUTN *2 S *1)(*2 (TENSE PAST))(DECLAR *2))
 文 3 : YOU LOOK ED
 意味 : ((GENE *1 YOU)(GENR *2 LOOK)(TOP *2)
 (PUTN *2 S *1)(*2 (TENSE PAST))(DECLAR *2))
 文 4 : I LOOK
 意味 : ((GENE *1 I)(GENR *2 LOOK)(TOP *2)
 (PUTN *2 S *1)(*2 (TENSE PRES))(DECLAR *2))
 文 5 : YOU GAVE THE RED DOG
 意味 : ((GENE *1 YOU)(GENR *2 GIVE)(GENE *3 RED)
 (GENO *4 DOG)(TOP *2)(PUTN *2 S *1)
 (PUTN *2 OBJ *4)(PUTN *4 COLOR *3)
 (*2 (TENSE PAST))(DECLAR *2))
 文 6 : I GAVE A BLACK CAT
 文 7 : YOU HAVE A BLACK CAT
 文 8 : YOU GIVE THE RED DOG
 文 9 : I HAVE THE CAT
 文 10 : I HAD THE CAT
 文 11 : YOU GIVE THE BLACK CAT
 文 12 : I GIVE A RED CAT
 文 13 : I HAD THE DOG
 文 14 : DID YOU LOOK THE DOG
 文 15 : DID YOU WANT THE BLACK DOG
 (文 6 から文 15 については意味表現は略)

図 12 入力データ
 Fig. 12 Input data.

```

n1 -> I ((TOP *1)(GENE *1 I))
n1 -> YOU ((TOP *1)(GENE *1 YOU))
n2 -> DOG ((TOP *4)(GENO *4 DOG))
n2 -> CAT ((TOP *3)(GENO *3 CAT))
v1 -> LOOK ((TOP *2)(GENR *2 LOOK)(*2 (TENSE PRES)))
v1 -> WANT ((TOP *2)(GENR *2 WANT)(*2 (TENSE PRES)))
v1 -> HAVE ((TOP *2)(GENR *2 HAVE)(*2 (TENSE PRES)))
v1 -> HAD ((TOP *2)(GENR *2 HAVE)(*2 (TENSE PAST)))
v1 -> GIVE ((TOP *2)(GENR *2 GIVE)(*2 (TENSE PRES)))
v1 -> GAVE ((TOP *2)(GENR *2 GIVE)(*2 (TENSE PAST)))
a1 -> A ()
a1 -> THE ()
adj1 -> RED ((TOP *3)(GENE *3 RED))
adj1 -> BLACK ((TOP *3)(GENE *3 BLACK))
did1 -> DID ()
ed1 -> ED ()
npl -> x1 n2 ((TOP n2)(PUTN n2 COLOR x1))
x1 -> a1 adj1 ((TOP adj1))
npl -> a1 n2 ((TOP n2))
n2 -> adj1 n2 ((TOP n2)(PUTN n2 COLOR adj1))
npl -> a1 n2 ((TOP n2))
v1 -> v1 ed1 ((TOP v1)(%(*KIND* (v1(TENSE(PRES))))  

    (:TRANS* (ed1:(v1(TENSE(PRES * PAST)))))))%)
s -> n1 v1 ((TOP v1)(PUTN v1 S n1)(DECLAR v1)  

    (%(*KIND* (v1(TENSE(PRES PAST))))%))
s -> n1 v1 npl ((TOP v1)(PUTN v1 S n1)(PUTN v1 OBJ npl)  

    (DECLAR v1)  

    (%(*KIND* (v1(TENSE(PRES PAST))))%))
s -> x2 v1 npl ((TOP v1)(PUTN v1 S x2)(PUTN v1 OBJ npl)  

    (QUEST v1)(%(*KIND* (v1(TENSE(PRES))))  

    (:TRANS* (x2:(v1(TENSE(PRES * PAST)))))))%)
x2 -> did1 n1
    
```

()の中は意味部。
 カテゴリー名は、分かりやすいように変えてある。

図 13 学習された文法
 Fig. 13 Learning results of grammar.

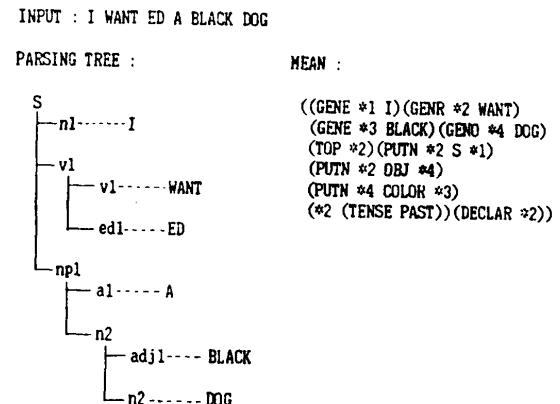


図 14 図 13 の文法を使った解析例
 Fig. 14 Analysis example by using acquired grammar in Fig. 13.

す。文 1 と文 2 で、AND ヒューリスティックを使って、S+V という文型と “I”, “you”, “gave” を学習している。文 3 と文 4 で、部分解析・AND ヒューリスティックを使って、“look”, “ed” を学習している。このとき、“ed”的役割も学習している。文 5 では、S+V+O という文型を学習している。文 5 から文 13 で、冠詞 (“a”, “the”) と形容詞 (“black”, “red”) と名詞 (dog”, “cat”) を学習し、それらを使った名詞句の文法を学習している。また、“have”, “had”, “give” も学習し、不規則変化の動詞を獲得している。文 14 で、Did+S+V+O という文型と、“did” を学習している。ただし、疑問文は “did”+“平叙文” という規則ではなく、“did 名詞句”+“動詞句 名詞句” という規則で学習されている。これは、未知語 “did” が、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。文 15 で、“did” の情報を用いて、動詞 “want” を現在形として学習している。以上の学習の後、獲得された文法と意味解析規則を用いて、入力文 “I wanted a black dog” を解析した結果を図 14 に示す。システムにとって非文法的な文である “I haded a dog” や “You have a cat red” などは解析不能というメッセージが出される。

7. 結び

以上、Anderson の理論に基づいた LAS の問題点に対し、改良・拡張を加えたシステムについて述べた。結果として以下のことがわかった。

- (1) 意味に基づく一般化は、かなりの文法を推論できる。
- (2) 単語を知らないところからはじめても、あま

り大きな問題はおきなかった。

(3) 否定、疑問文などの文も学習できる。

また、本システムの問題点は4章の学習ヒューリスティックの評価のところで述べたとおりであるが、他にも問題点が残されている。

たとえば、本研究では日本語の場合、動詞の過去形はすべて不規則変化しているものとして扱っているが、実際には、規則的な変化をしている動詞も部分的にはある。また英語の場合、“looked”のように完全に分解できる規則変化の動詞しか扱っていないが、これも実際には、“studied”のように動詞の語尾を変化させて“ed”を付ける場合もある。以上の問題を解決するには、さらに高度な学習機能が必要である。

もう一つは、意味表現から規則を引き出すので、意味表現に一般化が大きく依存することである。たとえば、マージ・ヒューリスティックの評価のところで述べた give と have の例のように、ラベルの付け方に一般性がないと、よい一般化ができない。しかし、これは Schank の CD 理論¹⁰⁾のように意味素を用いて、意味表現をしっかり規定すると解決されるとわれわれは考えている。

参考文献

- 1) Anderson, J.: Computer Simulation of a Language Acquisition System, in Solso, R. (ed.), *Information Processing and Cognition*, Erlbaum, Washington (1975).
- 2) Pinker, S.: Formal Models of Language Learning, *Cognition*, Vol. 7, pp. 217-283 (1979).
- 3) Reeker, L. H.: The Computational Study of Language Acquisition, in Rubinoff, M. and Yovits, M. C. (eds.), *Advances in Computers* (Vol. 15), pp. 181-237, Academic Press, New York (1976).
- 4) Norman, D. A. and Rumelhart, D. E. (eds.): *Explorations in Cognition*, W. H. Freeman, San Francisco (1975).
- 5) 高木 朗ほか：簡単な日本語の意味と文法を学習するシステム、情報処理学会知識工学と人工知能研究会資料, 39-31 (1985).
- 6) 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).

- 7) Selfridge, M.: A Computer Model of Child Language Learning, *Artificial Intelligence*, Vol. 29, No. 2, pp. 171-216 (1986).
- 8) Anderson, J.: Induction of Augmented Transition Networks, *Cognitive Science*, Vol. 1, pp. 125-157 (1977).
- 9) 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研究会資料, 55-5 (1986).
- 10) Schank, R. C.: *Conceptual Information Processing*, North-Holland, Amsterdam (1975).

(昭和 62 年 4 月 28 日受付)
(昭和 63 年 11 月 14 日採録)

中川 聰一（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、日本音響学会、人工知能学会、IEEE, INNS 各会員。

山本 幹雄

昭和 59 年豊橋技術科学大学情報工学課程卒業。昭和 61 年同大学大学院修士課程修了。同年(株)沖テクノシステムズラボラトリ入社。現在豊橋技術科学大学情報工学系教務職員。自然言語処理、人工知能に関する研究に従事。情報処理学会、人工知能学会、認知科学会各会員。

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)応用機器技術部勤務。在学中は自然言語処理の研究に従事。

