

言葉の属性を修得する自然言語理解系

松田晃一・高田正之・
青島正明・小谷善行
(東京農工大学工学部数理情報工学科)

概要

人間とのコミュニケーションを通じて、世界に関する知識及び言葉に関する知識を自動的に修得するシステムについて研究している[1]。この研究の基本的な理念は「人間の子供における知的能力の発達柔軟性」を、知的システムとして実現することである。本発表は、この研究のうち、「言語に関する知識の修得」及びそのベースとなる自然言語理解システムについて述べる。

本システムは日本語で対話を行う。言語知識の修得の対象は自立語(名詞、動詞、形容詞、形容動詞)である。基本的な語(助詞、助動詞、一部の自立語)は初期辞書に含むものとしている。修得されるものは語の品詞、意味属性、格支配構造である。システムは未知の単語を含んだ入力文を構文解析する。本研究の基本的な考え方は、未知の単語の品詞・意味属性を、既知の単語の格支配構造から推論すること(また、その逆に品詞・意味属性から格支配構造を推論すること)である。

システムの知識ベースはフレーム構造を用いており、読みフレーム、単語フレーム、実体フレームの三領域構造で表現される。

1 はじめに

今日、固定的な知識・言語処理システムは様々な方面で実用化の域に達している。知識・言語処理システムの最大の目的はシステムから人間に対して知識をもたらすことであった。したがってシステムの持つ知識は固定的に最初から組み込んでおくだけで十分であった。

一方、動的に知識や言語を修得するシステムの研究はほとんど未踏の領域である。この分野で研究が進まなかったのは、人間からの知識を蓄積していくのは不確実であり効率が悪く、上記の目的には一致しないからである。しかし、長期的あるいは学問的見地からはこの分野の研究の意義は見出せると考える。すなわち、(1) 認知科学としての意義、(2) 人間への情緒的で多様な対応をすべき分野への応用、が考へうる。

本システムではこの分野での実験的な研究を目標としている。ここではその言語的な知識の修得について述べる。

人間の知識・言語修得をみると決して体系的な情報が与えられてはいない。ある程度誤りを含む

入力を受けている。本研究はこうした不確実な状況でも動作しうるシステムをめざしている。人間の場合は入力を蓄積しながら、内部に知的構造を作り出す。その際、相手の人間から知識を引き出すような本能的ともいえる行動を取る。また自分の未知のものを知らうとすることに対する動機付けが起こる。以上の点をとらえ、言語修得の設計に当たって次のような性質をシステムが持つべきものとした。

(1) 自動性

- ・知識を自動的に集積するシステムであること。
- ・基本データ以外は諸知識は対話などの言語活動を通して修得するものとする。
- ・したがって人間が単語の品詞名を教えることはない。
- ・ANIMAL、EPAMのような自動的集積を高いレベルでの言語や知識について行う。

(2) 入力の非体系性

- ・教師なし学習であり、YES/NOなどの

特定のパターンの応答だけを想定することはない。

・どんな場合にも様々な非体系的または不完全な入力に対して、おのおのそれなりの知識を取り入れるものとする。

(3) 自己組織性

・言語・知識データを一貫性を持つように推論し、また対話の相手に働きかける。

さらに、知識データの修得の部分に関する性質として次の性質を持つべきものとした。

(4) 誘導性

- ・相手から自然に知識を引き出すための「からくり」を設計する。
- ・おしゃべりで日常的な情報をやりとりする。また適切な状況での質問を行う。
- ・相手への印象付けから入力方式までヒューマンインターフェースを考慮する。

(5) 自立性

- ・はっきりした動機（システムでは未知情報、矛盾情報の満足を動機とする）を持って行動すること。
- ・動機を持ち、必要な情報はシステムが自分から要求する。また、対話のリードは対話者もシステムもとれる。

2 システムの構成

知識・言語修得システムは、自然言語理解システムと知識ベース管理システム[2][8]と知識ベースの三つに分かれる。未知語の属性の推定及び修得は自然言語理解システムが行い、得られた未知語の属性は知識ベースに格納され、知識ベース管理システムがそれを管理する。

対象とする自然言語は、日本語である。

自然言語理解システムは、次の方式を用いることにした。(1) 左再帰規則を用いた構文解析が行いやすいように、構文解析をボトムアップに行う。(2) 意味解釈は日本語に関して意味構造を取り出しやすいと考えられている格文法をもとに格支配構造を用いて行う。

自然言語理解システムが生成する意味構造（知識）は、知識ベースに格納され、知識ベース管理システムが管理する。知識ベース管理システムは、フレームの検索、冗長性の削除、統一性の保証、

矛盾した知識の管理等を行う。知識ベースは、言葉に関する知識と世界に関する知識から成る。

2.1 初期辞書

本システムの初期辞書は、次のものを含む。

- (1) 構文規則（句構造文法）
- (2) すべての助詞・助動詞の辞書（単語名、品詞名、活用型、変形規則、助動詞と用言との接続関係から成る。）
- (3) 一部の名詞の辞書（単語名、品詞名、意味属性から成る。）
- (4) 一部の動詞・形容詞・形容動詞の辞書（単語名、品詞名、活用型、格支配構造から成る。）
- (5) その他の自立語の辞書（代名詞、連体詞、接続詞、感動詞）

(1)、(2)、(5)は固定的知識とし修得しない。
(注：(1)は知識ベース外の機構とし、それ以外はフレーム構造で知識ベースに格納する。)

2.2 知識ベース

本システムでは、言葉に関する知識を次の三つの領域に分けて考える。

- (1) 読み領域
- (2) 単語領域
- (3) 実体領域

これをもとに知識ベースは、〈対象〉というフレームをトップレベルのクラスとして持つ階層構造から成り、この〈対象〉フレームは、〈読み〉フレーム、〈単語〉フレーム、〈実体〉フレームという三つのクラスから成る。

知識ベース全体の構造を図1に示す。

それぞれの領域の中で、フレームは、相互にakoスロットを介して繋がっており、上位フレームの属性を継承できる。また、領域間でもフレームはスロット（正体スロット、名称スロット、実体スロット、単語スロット）を介して繋がっており、どの領域からでも他の領域の関連情報をデモニックの手続きにより取り出すことができる。

(1) 読み領域

読み領域には、システムに入力される文字列に関する知識が格納されている。

この領域は、システムへの入力として用いられ

そのままの文字列（「さとう」、「が」、「かつ」、「た」、「かつ」等。実際には単語分かち書きローマ字を用いている。）をインスタンスとして持ち、そのインスタンスだけから成る。読み領域のインスタンスは、〈正体〉スロットだけを持ち対応する単語領域のインスタンスをポイントしている。

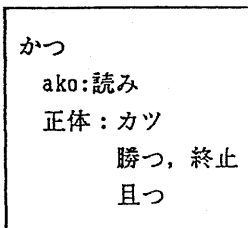


図2 読みのフレーム

ける）から成る。それぞれのクラスは、単語としての属性（品詞名、格支配構造、意味属性、接続情報等）を持つ。単語領域のインスタンスは、〈名称〉と〈実体〉というスロットを持ちそれぞれ対応する読み領域のインスタンスと実体領域のインスタンスをポイントしている。なお、用言の〈名称〉スロットは、〈終止形〉（と〈音便形〉）とする。

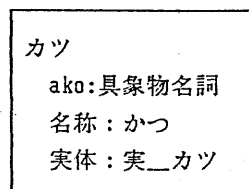


図3 単語のフレーム

(2) 単語領域

単語領域には、読みに対応する単語そのものに関する言語的知識が格納されている。

この領域は、単語に関するクラス（〈動詞〉、〈名詞〉、〈形容詞〉等）とシステムへの入力として用いられる文字列に対応する単語のインスタンス（「佐藤」、「が」、「買う」、「た」、「カツ」等）実際にはローマ字を用い、同音異義語などの区別のためにkatu_1のように識別番号を付

この領域のフレームには、この他に〈品詞〉、〈活用型〉などのスロットがある。また、発音上のアクセント、漢字表記などの属性を扱うことになれば、この領域のフレームに格納することになる。

(3) 実体領域

実体領域には、実体に関する知識が格納されている。入力文から得た意味構造（世界に関する知識）は、ここに格納される。

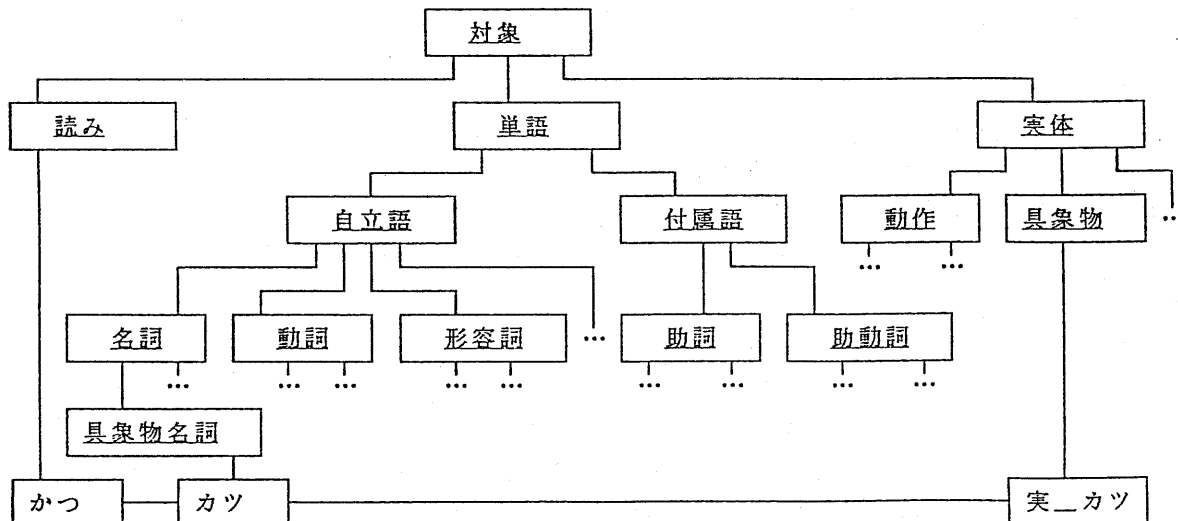


図1 知識ベース全体の構造

この領域は、単語領域のインスタンス（例えば、単語「カツ」）に対応する実体のクラス（「カツ」を表す 実_カツ）とインスタンス（「佐藤が買ったカツ」を表す 実_カツ_1）から成る。実体領域のクラスは、〈単語〉スロットを持ち、対応する単語領域のインスタンスをポイントしている。

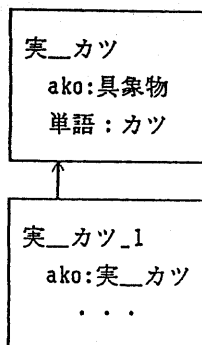


図4 実体のフレーム

動詞の辞書、名詞の辞書の例を付録に示す。

2. 3 自然言語理解システム

自然言語理解システムは、言葉に関する知識を利用して入力文を構文解析、意味解釈し、世界に関する知識を抽出する。構文木を扱うモジュール以外の処理は、知識ベース内に書かれているデモンによって行う。

2. 3. 1 意味解釈

意味解釈で得られる意味構造は、実体領域における実体フレームの繋がりとして表される。この構造は次のような処理で生成される。

(1) まず、基本的に格文法にもとづく意味構造を生成する。

つまり入力文に対して

- ・入力文中の述部を中心とした格補語と述部の実体とを適当なスロットを介して結び付ける。

- ・文中の名詞句が連体修飾されている場合には、名詞句の実体と連体修飾句との関係を格補語と述部との関係と同じ構造で表す。

これらの処理では、格補語と述部との係り受けは、右再帰を原則に必須格を優先して処理する。また、連体修飾は、被修飾名

詞が連体修飾句の述部の一つの格となることを原則として処理している。

この構造は、連体修飾という形で埋め込まれた句（文）の情報が取り出しやすく、形容詞、形容動詞の扱いも簡単になる。

次の文の意味構造を図5に示す。

(例) 佐藤が買ったカツを青島が食べた。

(2) 次に、知識ベース管理システムの持つデモンが起動して(1)で生成した構造に対して以下のことを行う。

- ・各実体に関する情報を取り出す。
- ・問い合わせにより、既にある実体フレームのネットワークの中に繋ぐ。
- ・知識の一貫性の処理を行う。

本システムは、言葉の属性の修得が主な目的なので、実体の同定などの問題は、何等かのデモンがうまく処理してくれるものと想定している。

2. 3. 2 参照回数

本システムでは、述語とその格補語との共起の回数を得るために、それぞれの述語の格補語のスロットに参照回数を保存する。この参照回数は、「の」を伴う名詞句中の形容詞、動詞の係り受けやあいまいな任意格の係り受けの決定に利用できる。また、この参照回数は、特定の用言と共起する名詞のカテゴリ分けにも利用できる。

例えば、「長い少女の髪」の場合、「長い」が「少女」と「髪」のいずれかに係るかは、「長い」と「少女」、「長い」と「髪」の共起回数から係り先を決めることができる。

3 修得される単語及びその内容

本システムは、自立語（名詞、動詞、形容詞、形容動詞）を修得する。それぞれの品詞に関して修得される属性は次のとおりである。（以下、例文中下線を引いた語は未知語とし、[]の中は名詞の意味属性を書いている。）

(1) 名詞の場合：名称、品詞名、意味属性

(例) 大学：だいがく、名詞、場所名詞
(沢田が大学に行く。)

(例) 大学：だいがく、名詞、具象物名詞
(大学が火事で燃える。)

(2) 形容詞／形容動詞の場合：

名称、品詞名、格支配構造

(例) 黒い: くろい、形容詞、[具象物]が
(黒い手帳/あの手帳は黒い。)

(3) 動詞の場合: 名称、品詞名、格支配構造

(例) 帰る: かえる、動詞、
[動物]が+[場所]から+[場所]に
(沢田が大学から家に帰る。)

新しく修得した未知語は、読み領域、単語領域、実体領域にフレームが新しく作られてスロットで相互に結び付けられる。そして、それぞれ適当な階層の子として知識ベースに格納される。

これらの修得された未知語は、2、3、2と同様に参照回数(初めは、値が低い)とともに修得される。この参照回数は、その単語を用いた処理が成功することに値が増し、その値が低いうちはその単語に関する情報の信頼度が低いとみなす。

4 単語の修得方式

未知語の品詞名等の属性の修得は、次の二段階で行われる。(1) 構文解析では、品詞名の推定を行う。(2) 意味解釈では、名詞の意味属性や用言の格支配構造の推定を行う。

意味解釈が成功した時点でそれぞれの属性を修得する、すなわち、知識データとして知識ベースに格納する。

4.1 品詞名の修得

品詞名の修得は、入力文に含まれる未知語の品詞名が、名詞、動詞、形容詞、形容動詞のいずれかであるとして構文解析し、構文解析が成功した場合の品詞名をその単語の品詞名として推定する。

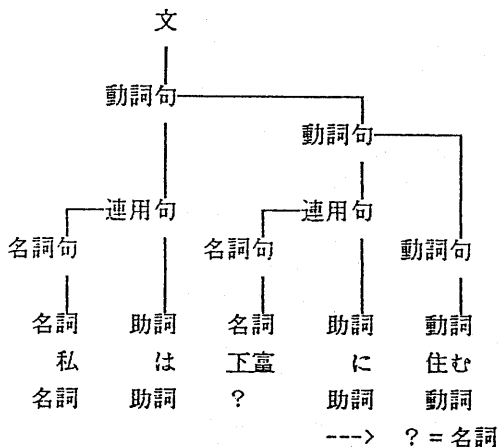


図6 構文解析による品詞名の修得

ただし、用言(動詞、形容詞、形容動詞)が述語として用いられるときは、構文解析からだけでは品詞名が決まらない場合がある。この場合は、その用言の直後にある単語の品詞名、助動詞の接続情報、用言の活用語尾の情報を補助情報として品詞名の推定を行う。それでも決まらないときは、品詞名を動詞として推定する。

(例) ニワトリは、飛ばない。

助動詞「ない」の接続条件と五段動詞の活用語尾から「飛ば」は動詞の未然形であると判断する。

4.2 意味属性と格支配構造の修得

意味属性と格支配構造の修得は、意味解釈で行う。意味解釈は、格支配構造を用いて大よそ次のように行われる。

(1) 入力文の持つ用言(動詞、形容詞、形容動詞)の格支配構造を取り出す。

(2) 入力文に含まれる助詞、名詞の意味属性をもとに、用言の格支配構造に書かれている名詞の意味属性と助詞をチェックし用言のフレームの格スロットを埋めながら、意味構造を生成する。

(なお、この意味解釈は、入力文から用言の実体のインスタンスを作り、知識ベースの実体領域の子として加えif_addedデモンを起動して、単語領域にある用言のフレームから格支配情報を引き出しながら行われる。付録2参照)

4.2.1 名詞の意味属性の修得

(2)で、名詞の意味属性をチェックして格スロットを埋める際に、入力文中の名詞の意味属性が未知である場合、その名詞がその格スロットに適合する意味属性を持つとしてそれを修得する。

(例) 投げる: [人間]が+[具象物]を
佐藤が ボールを 投げる。
[人間] [?] ----> ? = 具象物名詞

4.2.2 用言の格支配構造の修得

(1)で、入力文の用言に対応する(単語)フレームがなく格支配構造が得られない場合、また、同音異義語等のため、すでに知識ベースにある格支配構造を用いての意味解釈が失敗した場合、格

支配構造の修得が行われる。未知の用言の格支配構造は、入力文と同じ格パターンを持つ既知の用言の格支配構造と同じであると類推しそれを修得する。この処理は、知識ベースの単語領域にある〈用言〉フレームのako階層を下に探索しながら行われる。

(例)

行く:[動物]が+([場所] から)+ [場所] に
 ---> 自動詞11

佐藤 が 大学 から 家 に 帰る。
 [動物] [場所] [場所] [?]

---> ? = 自動詞11

(例) 赤い: [具象物]が
 黒い 手帳
 [?] [具象物]

---> ? = 形容詞5

あの手帳 は 黒い。
 [具象物] [?]

---> ? = 形容詞5

(注:()で囲まれた格は、任意格を表す。)

5 特別な場合の意味属性の修得方式

ほとんどの入力文では、名詞の意味属性は用言の持つ格支配構造を用いて修得することができるが、次のような場合は、修得できない。

- (1) 入力文の述部の用言が未知語である場合
- (2) 入力文に含まれる複数の未知語(名詞)がそれぞれ同じ助詞を取る場合

このような場合、システムはユーザに質問して意味属性を修得する。

(2)の場合に関して例を挙げて述べる。次の場合は、それぞれの未知語に対して、誕生日[時間]・居酒屋[場所]、誕生日[場所]・居酒屋[時間]の二通りが考えられる。

(例) 行く:[動物]が+([時間] に)+ [場所] に
 荒木 が 誕生日 に 居酒屋 に 行く。
 [動物] [?] [?]

まず「居酒屋」に対して単語領域、実体領域に新しくフレームを作成する。そのとき、実体フレームにおかれたデモンが起動し、ユーザに質問する。質問文は、単語スロットと名称スロットを介して文字領域を参照して生成する。

システム:「居酒屋は何ですか?」

ユーザ:「お酒を飲む所です。」

次に、ユーザの応答文に含まれる名詞「所」を抽出し、「所」が知識ベースに存在するか否かを調べる。

・それが知識ベースに存在するとき、すなわち「ところ」が読み領域に存在し、単語領域のフレームが〈場所名詞〉の子であり、実体領域のフレームが〈場所〉フレームの子孫であるとき、実体領域における「所」を「居酒屋」の親とし、「居酒屋」の先祖が「場所」となる。そして、〈場所〉フレームの属性スロット値「場所名詞」が「居酒屋」フレームに伝わり、単語領域において「居酒屋」は場所名詞の子となり、意味属性が場所名詞と判断される。

・知識ベースに存在しないとき、または抽出された名詞が多義語として知識ベースに存在するとき、同様な手続きをし、意味属性が判断できるまで質問を繰り返す。ただし、適切なako階層につながらないときは、処理を中断し、階層構造は一応〈実体〉フレームの子孫として修得する。

6 おわりに

本報告では、知識・言語修得システムでの知識ベースの構造と入力文に含まれる未知語の属性の修得について述べた。

本報告にもとづいた学習機能を持つ自然言語理解システムのプロトタイプをPC9801上にProlog-KABAを用いて試作中である。

7 参考文献

- [1] 小谷, 松田, 青島, 高田: 知識・言語修得システムの設計, 第31回情報処理学会全国大会講演論文集 7H-2, 1985
- [2] 青島, 松田, 高田, 小谷: デモンによる知識修得システム, 第31回情報処理学会全国大会講演論文集 7H-3, 1985
- [3] 井上和子: 変形文法と日本語 上・下, 大修館書店, 1976
- [4] 水谷静夫他: 朝倉日本語新講座 3文法と意味 I, 朝倉書店, 1983
- [5] 田中徳積: 計算機による自然言語の意味解釈に関する研究, 電子技術総合研究所研究報告797, 1979
- [6] 大沢一郎他: オブジェクト指向方式によ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
- [7] 安室, 青島, 松田, 高田, 小谷: フレームを用いたPROLOGによる日本語構文解析, 第30回情報処理学会全国大会講演論文集, 1985
- [8] 青島, 安室, 高田, 小谷: フレームに

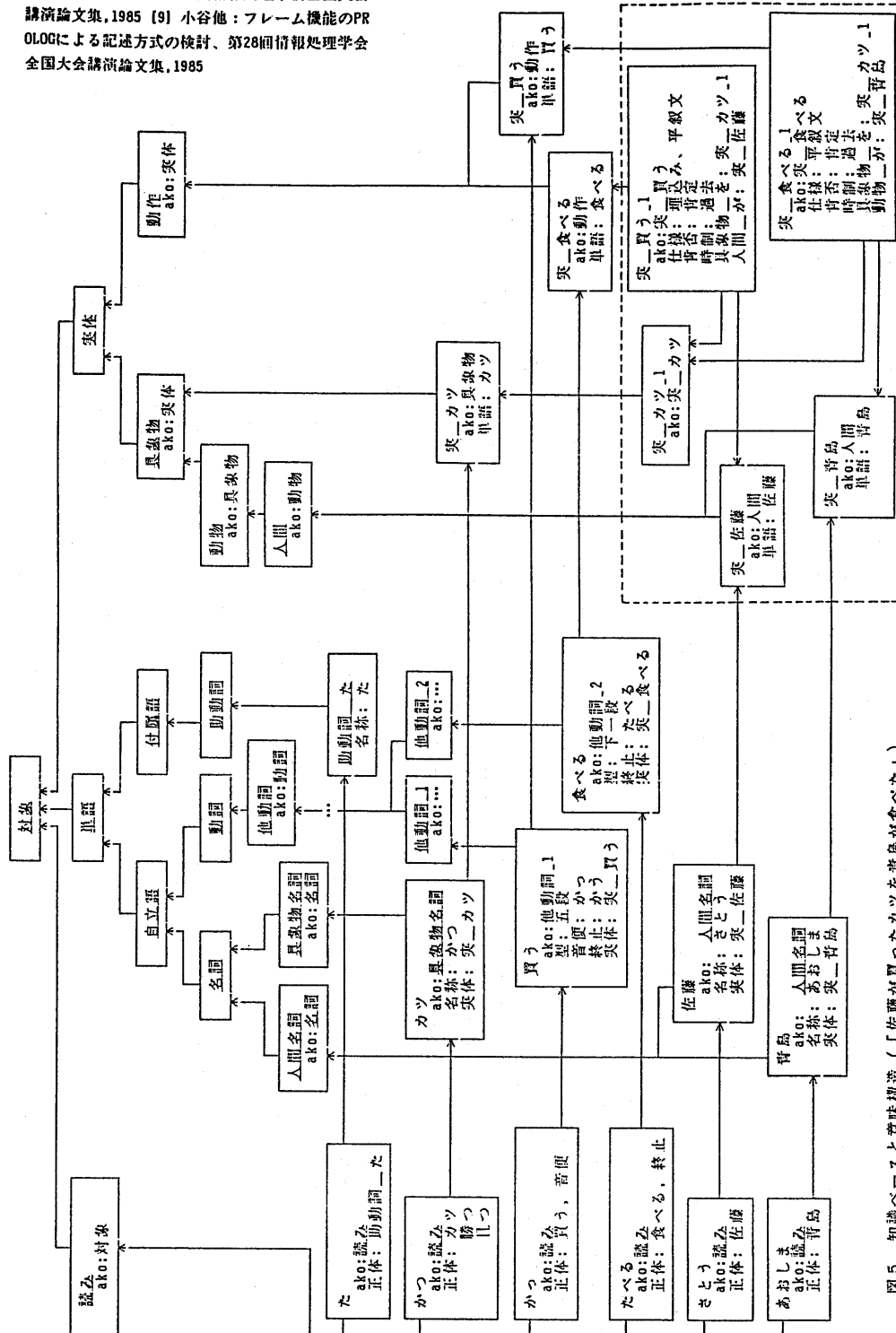
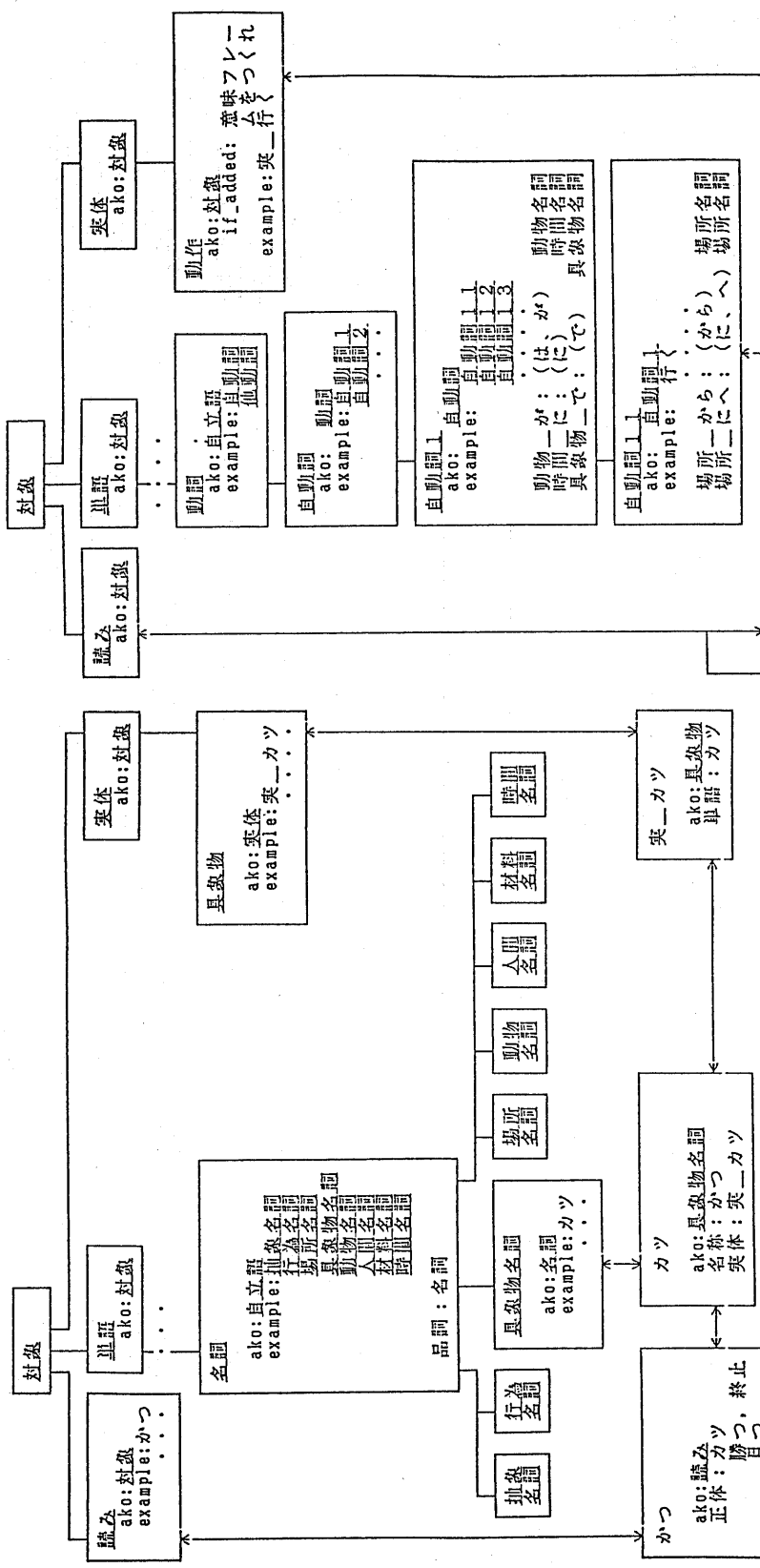


図5 知識ベースと意味構造 (「佐藤が買ったカツを青島が食べた」)



付録 1 知識ベースにおける名詞辞書

付録 2 知識ベースにおける動詞辞書