

考察：「日本語文法学習システム」

日本語文法、日本語の意味表現の学習の原理を考察していきます。まずは、日本語の文法のパターンを調べて、基本要素は何かを追求します。そして、次に意味と文法の基本的な関係は何かと考えを進めていきます。筆者はプログラマーなので、最後に文法と意味の表現案を提示したいと思います。

なお、本考察は主に次の本に触発されたものです。

- ・「人はいかに学ぶか」 稲垣佳世子、波多野誼余夫著 中公新書
- ・「問題解決の心理学」 安西祐一郎著 中公新書
- ・「日本語練習帳」 大野晋著 岩波新書
- ・「分散人工知能」 片桐恭弘著 コロナ社
- ・「認知意味論のしくみ」 初山洋介著 研究社
- ・「日本語文法のしくみ」 井上優著 研究社
- ・「ここからはじまる日本語文法」 森山卓郎著 ひつじ書房
- ・「認知科学への招待」 淵一博編 NHK ブックス

その他にもいくつかの本が参考になっています。

第1章 はじめに

意味はどこから生まれてくるか。それは脳の情報活動全てが源泉となって生じてくるものでしょう。物を見る。そこには、形ある物があり、それには属性があり、それは衝突したり、飛んだりといった活動とか他者への作用なんかをしています。人間には感情があり、思考して抽象的な心的体験もします。また、人体は頭、胴体とかからなり、頭はさらに目とか、口とかから構成されます。物事には属性の他に構造もあるといえるし、入れ子構造、繰り返し、時間、空間、様々なパターンからなることも見て取れるわけです。そうした物全てが意味体系の土台になるのです。それらは、更に下の土台に還元できない、プリミティブと考えられます。物理の世界では、物事の原理がどこまで還元されていくか分からないですけれども、言語を考える上での土台となるのは、脳の情報活動の中で体験されること全てということで決着できそうです。土台となる要素は無数にあるでしょうけれども、それより下には還元されないという、安心感のある世界です。人は、言葉を学習しながら、その意味の土台を再発見しているといえるのです。

そのことを前提に、日本語の言語現象を考察していったら、文法が実際にこうした意味要素を基本として組み立てられているのを見ます。そうして、このことをカテゴリ化機能としてみて、すべての言語現象をカテゴリ化で生まれる品詞（作業品詞もふくめて）とその上の基本的操作で表し、計算機で言語学習を実現する手法へとつなげていこうと思います。

第2章 様々な文法項目

自然言語は1次元の記号の並びに意味を持たせる仕組みであります。だから、1次元に表現するという基本的な制約が働きます。時空間の4次元の中で人間活動全般を、1次元の記号列の中に情報として盛り込むということで、様々な工夫が必要になってくるでしょう。物理的に大それた事はできるわけもなく、そこには地道な基本的な原理が生じているはずで、それを、例文を追いつながりながら見ていきましょう。

その前に、1次元に記号を並べるのだから、次のことは、物理的に生まれることがらです。そのことは、前提といたしましょう。

(1) 配置

top	:	文頭
tail	:	文末
at	:	その場所
before	:	atの前方
after	:	atの後方
just before	:	atの直前
just after	:	atの直後
N th	:	N番目

そして、記号にはまとまりがあるから、記号としてのまとまり、記号列のまとまりがあるでしょう。

(2) まとまり

- ・音韻
- ・単語
- ・句
- ・文
- ・文章(段落)
- ・品詞あるいはカテゴリ
- ・入れ子

1次元配置のみでなく、記号要素間の関係もあるでしょう。構造ですね。公理として、交差的構造はないものとします。実際になくて、物事はすべてネスト構造のみです。関係としては、修飾関係と係り結びの関係があります。そのほかの関係は著者は知りません。これらは分析前に、前提として、あるものとしましょう。

(3) 関係

- ・修飾
- ・被修飾
- ・係り結び

これらを基盤に文に当たっていきましょう。

2.1 助詞

「は」は主題の提示を表し、「が」は副主題を表します。「が」はまた目的格「を」を持たない動詞の時、主格を示します。この性格上、「は」は文全体を修飾することになり、結局、文の最後の動詞を修飾することが多くなります。「が」は直後の動詞を修飾することになります。

(例文1) サクラにしてもコブシにしても、開花している樹には、ある種の謎めいた妖しさがあり、ひきこまれるような美しさがある。

【説明】

- ・「は」の導く句は、「妖しさがあり」と「美しさがある」を修飾しているのはあきらか。「が」は直後の「あり」とか、「ある」とかを修飾しているのみです。
- ・「は」の導く句は主格になっているとはいえない。「が」の導く句は、明確に主格になっているのがわかります。
- ・「は」は「樹に」を受けて、この句を主題にしているのですが、これは「は」が格情報を本来持たないことを示しています。主題を表すから、たまたま主格を導いたりするだけ。これに比べ、「が」は格情報を担っていますから、「開花している樹にが、」とは言いづらいものがあります。「に」格と主格とを重複してもってしまうからです。

(例文2) 私はリンゴが食べたい。 cf: 私はリンゴを食べたい。

cf: 私はリンゴが好きだ。

cf: 象は鼻が長い。

cf: x 私はリンゴを食べる。

【説明】

・例文1で、「が」が副主題を表すと述べました。それは例文2の「が」を「を」に置換できるからです。でもこれはいつでもできるとは限りません。「食べたい」とか、「好きだ」とか、形容する言葉(形容詞、形容動詞)を修飾するときです。自分の属性を修飾するとき、「・・・は、・・・が xxx。」と言えるようです。それは、属性を表現するとき、動詞みtainな表現とダブル(「好きだ」は英語では動詞ですね)のために捻れて、「を」が使えるようになったのだと考えています。

・「は」は主格にならないときが多々あります。

cf: 椎茸は軸を除きます。(料理の言い方)

主格の同定は難しい意味論的解析問題の世界の中にあります。

「は」と関連して、ことあげ詞を考えてみます。「も」、「だけ」、「のみ」、「など」、「ばかり」、「しか」、「さえ」、「まで」、「でも」、「くらい」などです。これらは、「は」と同じような振る舞いをします。無論それぞれに個性がありますが。格についていえば、

(例文3) 私もリンゴを食べた。 cf: 私はリンゴも食べた。

cf: 私もリンゴは食べた。

- cf: 京都だけ行った。
 cf: 京都へだけ行った。
 cf: 京都だけ美しい。
 cf: 京都だけが美しい。
 cf: x 京都がだけ美しい。

【説明】

・主格を導いたり、その他様々な格を導きます。深層格への解析がものすごく難しい助詞群です。しかも論理を担ったりしますので、意味を何らかの構造（マップ）で表さねばなりませんから、これらの語の学習システムを作るのを困難にしています。

「の」の格助詞兼名詞修飾機能は注意を要します。

- (例文4) 私の昨日作ったプログラムは、巧く動いた。 cf: x 君への作ったプログラム
 (例文5) 彼女の好きなアップルパイを沢山つくった。

【説明】

・「が」格が「の」になって、直後の名詞を修飾しています。例題5で、「彼女の」は「アップルパイ」を修飾するのではなくて、「好きな」の主格となっているのが読み取れます。その他の格接続の場合は直接名詞を修飾しなくてはならないようです。

cf: 君へのプログラム

これは、主格が主題の提示と深い関係があって、「の」の後ろに動詞がある場合には主題を提示しているからで、「の」が格を体現していないからかもしれません。

「の」そして、関連語「こと」も微妙な働きをします。そして、文の中で機能的に重要な働きを担っています。

- (例文6) 彼が来るのが遅れた。 cf: 彼が来るのを、彼女は問題視した。
 cf: 彼が来ることを、彼女は問題視した。
 cf: x 彼が来ることが遅れた。

【説明】

・「の」も「こと」も句とか文などを名詞化して後に続けて発話するのに利用されます。ただ、「の」が適切な文と「こと」が適切な文とがあるようです。「の」は事実を表し、「こと」は状況/事象を表すという、「こと」という単語の意味的な要素が原因でしょう。「こと」だけでなく、どんな単語も、機能語になっていても、その他の状況下での使われ方を引きずって、文法として、いつも現れてくるものです。

2.2 係り結び/並置機能

「しか」は否定と呼応します。

- (例文7) 彼はコーヒーしか飲まない。

【説明】

・このように言い方の呼応関係があります。これを実現するには、単語が本来持っている意味を、さらにズバリかたる単語で受けるという規則を言語システムが持っていることが必要です。この場合には、単語「しか」は内面的に否定要素を持っていることを、言語システムが認知する必要があるわけです。それは、「しか」といっしょに何時も否定「ない」が現れるといった言語体験の裏打ちからくるのでしょうか。強化学習ですね。そんなものがある。もし、文が否定的意味を持たないのであれば、「だけ」という単語を使っていくというような、選択肢を認知できる意味土台も必要です。

並置関係として、このような係り結びの言い方が沢山あります。

(例文8) 私が言ったのと君が言ったのとでは問題が大いに違ってくる。

【説明】

・並置される単語や句がこのように対となる語によって明示されることがおおい。
・ちなみに、並置語には文レベルを接続するもの(「しかし」, 「ところが」, 「さて」, 「ところで」), 語レベルを接続するもの(「および」, 「ならびに」, 「ないし」) その両方が可能なもの(「つまり」, 「すなわち」, 「そして」)とがあります。語レベルの接続詞としては、並列助詞(「と」, 「とか」, 「たり」)もあります。文の名詞化とか、入れ子された句を用いて動詞句にするとかすれば、語レベルの接続詞も文を並置できるわけです。この辺の性質は、文内の並置構造を解析するときの手がかりとなります。

2.3 数量と程度

数量と程度の違いは微妙で、すぐには分からないことがあります。

(例文9) もっと速く歩け。 cf: x もっと少しく下さい。

cf: もう少しく下さい。

cf: x もう速く歩け。

【説明】

・「速く」は程度の副詞ですから、「もっと」で修飾できます。「もう」は数量を修飾する語です。だから、「速く」には掛かりません。

「もっと」という単語を覚えるとき、スケールな動詞(「もっとください」)についていることが多いので、そういうことが言語の中に定着したのでしょうか。「もう」は「もういません」のように量を表す語をうける動詞について利用されることが定着したものでしょう。

2.4 テンス・アスペクト・モダリティ、文末要素

助動詞、終助詞およびその相当語を考えていきます。これらは文末や動詞句の末尾にある要素です。

これらは、文末にくることから、文全体の状況というか、文全体に掛かる意味を込める語群です。

また、日本語には「て」でつながった連結動詞として働くものが多数あります。この辺の議論は「ここからはじまる日本語文法」(森山卓郎著：ひつじ書房)に詳しいので、割愛します。ただ、連結詞としての文法および意味は、各単語が担っている意味要素(隠れている場合が多く、発話者も意味を気づいていない)の組み合わせで決まってくるようであると述べておきましょう。

2.5 語の省略

日本語は様々な語の省略が可能です。談話では省略がないことは希でしょう。

(例文10) 昨日、彼女は街に出た。街で着物を買おうというのだ。それで、バスに乗った。……

【説明】

1番目の文に主題提示の「は」があって、以下の文の全ての主語が「彼女」であることを示しています。主語がないのは日本語では当たり前になっています。

これは、知識(文脈知識)の中に、格フレームがあって、デフォルト値がそのスロットに設定されていて、それを使うのが経済的であるからでしょう。言わんとしていることが明確なら発話しないという原則です。

(例文11) 私行く。(「が」が落ちている) cf: 街行く。(「へ」欠落)
cf: リンゴ食う。(「を」欠落)
cf: これは私のだ。(「もの」欠落)

【説明】

格の欠落も、機能語(「もの」)の欠落も、あるいは虫食いのある文も、単純な文では、推測が可能な場合があります。場所を表す名詞(「街」)、食べられるもの(「りんご」)、行為の主格になりうるもの(「私」)などの情報が単語に付加されて管理されているのでしょう。文法を学習するときには、そうした意味情報を元に行っているわけですから、当然のことです。文法学習と文解析とは互いに逆問題になっているわけですね。

第3章 文法を決めていくもの

意味は、オブジェクト（もの、事象、行動、操作）と属性、言語機能（格、修飾・被修飾、入れ子（開始、終了）、並置、順次配置）に分散して保持され、鍵と鍵穴の関係で、マッチしたものが有効に認知されていくものです。

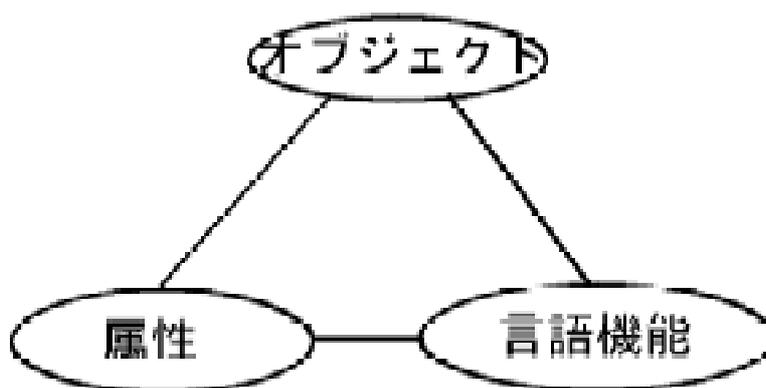


図3.1 意味のネット

単語の配置などの物理的に決まる項目以外の意味的な文法項目としては次のものがあって、それぞれ学習が行われて、脳に設定されるはずですが。

(1) 慣用語化（文法化）

例：不幸なことに

「不幸な xxxxx ことに」と「不幸な」と「ことに」を分離できない。

「これにたいして」とか、「行ってくる」、「行ってみる」なんかも文法化の例です。

(2) 語の共起性

例：・・・しか・・・ない

「しか」は「ない」でうけます。また、「もう少し」とはいえるけれども、「もっと少し」とは言えません。

共起性として、活用形もあります。特定の語に接続する動詞や形容詞、形容動詞は特定に活用語尾を持ちます。

(3) 表層格と深層格の関係

てにおはの表層格が実際にどのような深層の主格、目的格、間接目的格、場所格、時間格、is-a、has-a・・・、などにマップされるかを判定できなくてはなりません。

(4) テンス、アスペクト、モダリティ

単に、現在、過去、創造、継続といって、割り切れないような微妙な意味合いを、テンス、アスペクト、モダリティは持っています。これも文法に反映されます。

(5) 機能語の働き

助詞、助動詞、接続詞、および、それらの相当語句（慣用句）は文法の要です。

まないということです。少なくとも、物理的な文法は基本文の提示が学習には必須なはず
です。それ以上の意味的文法では、強化学習で、意味の断片と、文法の断片を対で切り取
り、プロトタイプを推定していき、学習が可能かもしれません。曖昧模糊とした言語認識
がだんだんに明確な輪郭を形作っていくような、戦略もあるかも知れません。ですが、や
はり簡単な基本文によって、プロトタイプを学習していくのが効率的でしょうし、学習シ
ステムを構築しやすい。プログラマーだから、そんなことを考えます。

前の第2章で述べたことは、すべて意味要素の台があったと仮定して、すべて説明つけ
られていくものです。意味要素の台の無いところには文法はできません。文法を検討して
いて、そう思いました。

文法を決める規則はローカルなもので、異なる意味合いの文法が、互いに重複した言語
表現になることもあり得ます。そうした場合は曖昧性として、文脈の支持の強い方が選択
されます。規則は厳密なものでなくていいのです。だから、複雑な言語というものが自己
組織的に学習できていくのでしょ

3.1 慣用語化

慣用句は、成分の単語がそれぞれ持っている意味が合成されて、その意味パターンが何
時もまとまって現れる時、学習されます。

「の中に」は、名詞を受ける助詞「の」に「中：in」と場所格「に」が何時もまとまっ
て現れることから、学習されます。もう一つは、英語で前置詞一つ「in」で表されるよう
に、概念がまとまって意識される単位「の中に」になっているともいえます。

3.2 語の共起性

語に共起性があるということは、双対単語の間に、鍵と鍵穴のような関係のある意味要
素を持つということです。

「もっと速く」では、単語「速く」が鍵穴の意味である「スケーラブルな様態」という
意味要素を持つとし、単語「もっと」が鍵である「スケーラブル」を持つというわけです。
鍵と鍵穴は意味要素をあらかじめプリミティブとして生得的に持っている必要があり、そ
れを言語現象から選択していくという学習方略が働くのでしょ

3.3 表層格と深層格の関係

主体、目的格体、間接目的格体というフォーカスの対象を選択する能力があらかじめ脳
の中にあるのだと思います。

(文1) 金槌で釘を打つ。

(文2) 電車で長野へ行く。

(文3) 海原を行く。

(文4) 電車でものを運ぶ。

文1はハンマーが釘をめがけて移動するイメージが浮かびます。一方で、文2は電車が長野をめがけて移動するイメージが浮かびます。しかし、文1では釘は目的格で、文2では長野は場所格（行き着く場所）です。この違いをどう説明したらいいでしょうか。

文1はハンマーにフォーカスが当たっていて、文2では長野にフォーカスが当たっていない、つまり、「釘」には目的格体が添付されて学習されると考えるべきだということです。実際、文3のように、「行く」は「を」格を持つ表現がありますが、この「を」格を目的格と見る人は少数派でしょう。「を」格をとるから、目的格であるとするのは、機械的すぎます。

同じ移動を表す「運ぶ」では、明確に「を」格のものにフォーカスが当たっています。だから目的格になるのです。

3.4 テンス、アスペクト、モダリティ

この学習は体験しかない。文全体を修飾するので、文末に置くことをまず学習します。これは、文法学習者が本能的、感覚的に分かるというものだと思います。文頭、文末効果（記憶に残りやすかったり、フォーカスを引いたりする）というのはよく知られた事です。英語では、動詞についていますけれども、動詞の性質として学習する体系もあるということですね。

その他の細かい意味は、言語体験によってつかんでいくわけです。

「する」は普通の表し方で、眼前に見えるような表現として学習されます。「た」は過去のことという体験ですね。「している」も体験。意味はプリミティブとして生得的なものです。そして、これらの組み合わせ方も体験によって学習していく。

「私は、昨日長野へ行った。にぎやかだった。行列がくる。人垣ができる。それはもう楽しかった。」などと、「た」と「する」が混合していますが、言わんとするところはいかに伝わってきます。非文ではありません。英語では許されないことです。全て、体験学習です。

3.5 機能語の働き

「とき」と「しかし」について考えてみましょう。助詞については他でふれていますし、助動詞はテンス、アスペクト、モダリティなどで語られています。

例：山へ登るとき、リックを持って行きなさい。

この文では、「時」が条件節を起こしています。時間感覚の転用でしょう。比喻みたいなものです。言語では転用（比喻）は重要な機能です。実際、類推というのは脳の活動の根本にあるものです。

例：学校へ急いで行った。しかし、遅刻した。

「しかし」は論理の働きで学習されます。論理的ベースの上に言語体験があって、「しかし」が論理的な意味要素とコネクティングしていくのでしょうか。

3.6 省略、欠落

これは、言語を獲得していることから可能となる能力です。言語は、前後関係を連想によって想起されるところが大です。知っている曲だと、次々に歌詞が浮かんできます。あ

んな学習が言語でも起きるのです。省略や欠落に耐えるというのは、連想記憶の問題なのです。

3.7 指示代名詞

言語体験によるところ大です。基本は、生得的な能力である、自己とか相手とか、第三者という概念が操れることです。

それと、語の共起性との関係があります。

例：私は、机にあったリンゴを食べた。それが学校で問題となった。

「それ」は「私は、机にあったリンゴを食べた」という文全体の事象を指しています。

例：私は、机にあったリンゴを食べた。そのリンゴが学校で問題となった。

「そのリンゴ」は「机にあったリンゴ」を指しています。

このように、指示代名詞は前後の語の意味によって、何を指すかが変わってくるのです。指示代名詞は前後どこかの作業品詞を指すわけで、この解析は微妙な意味処理を必要とします。語の共起も同じでしょう。単語間の共起関係を文という大きな枠の中に拡大投影したものが指示代名詞の働きであるといえます。

3.8 論理

指示代名詞と同じく、論理は生得的な能力です。それが言語現象とコネクして、言語の中に組み入れられていくのです。基本的な構文での学習が重要です。あとは徐々に複雑な論理を組み合わせから自動学習していきます。

3.9 数量

論理と同じです。

3.10 様態、属性

様態、属性はクオリア体験が言語とコネクしていくことで学習されます。だから、基本構文での学習が重要です。

形容詞か形容動詞かは、やまとことばか漢語（属性名詞）かの違いだと思います。

例：なんと美しい。 : 形容詞表現

なんとという美しさだ。 : 形容動詞表現

3.11 時間格

これは、生得的なものです。これが場所格のイメージで把握されたものです。

3.12 場所格

生得的なもので、言語と場所の意味体系がコネクされていき、学習が成ります。

3.13 文法学習のアウトライン

コーパス（文を沢山集めたデータ）の中に、未定義の単語があるとき、その品詞を推測するというソフトウェアがあるそうです。英語の世界ですけれど、分かっている単語の配置などからかなり高い確率で未定義単語の品詞が同定できるとか。日本語は動詞や形容詞、形容動詞は活用語尾が決め手になるし、名詞だと助詞が付着しますから、前後の単語の品詞情報からだけでなく、多くの手がかりがあります。

品詞と活用が決まれば、修飾・被修飾関係はほぼ決定され、文内での語の共起関係も解析できていくのではないのでしょうか。対と非対をコーパスのなかで件数分析すればいいのです。

こういう風に、計算機学的な学習方法も大いに追求していくべきでしょう。でも、一応ロボットが学習するとしたらということで、解析的な方法も考えてみたく思います。

(1) 連体詞の学習

(例) 修飾語 被修飾語

例のように語が並んでいて、被修飾語が名詞だとします。修飾語は連体詞と判断します。そしてそれは、名詞の属性を表現しているだろうと判断します。

(2) 副詞(連用詞)の学習

(例) 修飾語 被修飾語

例のように語が並んでいて、被修飾語が動詞だとします。修飾語は副詞と判断します。そしてそれは、動詞の属性を表現しているだろうと判断します。

(3) 形容詞

連体詞と副詞で語幹が同じもの、かつ、名詞的働きがないものを形容詞とします。

(4) 形容動詞

連体詞と副詞で語幹が同じもの、かつ、名詞的働きがあるものを形容動詞とします。

(5) 格助詞の働き

(例) パパは6時に帰ると言ってたわ。

「パパは6時に帰る」という文と「言う」という動詞の意味を知っていたとして、それら2つを結びつけるものが「と」だと学習します。それは単なる連想です。動詞と格というものが形作る構造が頭に入っていたとしてですが。それは、「言う」の学習で行われることでしょう。

こうして、学習が進むとして、文や句といった作業品詞のデータ表現はどうあるべきでしょうか。柔軟な処理を実現するためには、作業品詞のなかに意味を展開し、保持しておくべきでしょう。表層表現と深層表現(フレーム)の両方で、意味を表現しておくことになると思います。

第4章 分散問題解決と品詞

日本語文を解析して、意味を捉えて行く処理を作ろうとしたとき、その文法の微妙さ、曖昧さが問題を複雑にしているのに気づきます。言語現象を全て捕捉しているわけでないし、非常に困難を感じます。でも、プログラム開発に着手しなくてはならない。ましてや、文法を学習するシステムをや・・・であります。

プログラム開発は反復方式(XP開発)で、少しずつ核心に迫っていくとして、一生ものの開発だなとつい思ってしまいます。そうであったとして、そうでなかったとして、プログラムは柔軟な作りであって欲しい。モジュールとして、新しい機能を簡単追加で対応できるようにでありたい。そんなとき、昔のHearsay-の黒板システムに飛びつきたくなります。これは、分散問題解決の一つのインプリメントであって、複数のプロセスが協調して、曖昧性のある自然言語を解析していくシステムです。対象の言語データとか、作業データとかは、共有メモリである、黒板に保持して、これを並行処理で書き換えていき、最後に最適解にいたるという技術です。

著者が「意味処理の考察」で提案した動画システムを黒板に置いて、さまざまな解析処理をし結果を出すプロセス群を協調させて、順次曖昧性を取り除いていくという方法がありそうです。その場合問題になるのは、各プロセスの作業結果をどう集約して、次のステップへ利用していくかというデータ同期のことです。

その対策を考えていて、文というものが多くの入れ子構造をしていることに注目しました。時枝文法を勉強すると、この入れ子構造に力を込めて、理論化されているのを見ます。その他の文法理論も同じです。私は入れ子を作業品詞と見ました。次の例文を見てください。

(例文1) 彼女は、かわいい顔をした人形を持っていた。

この文は次のような入れ子構造をしていることがわかります。

(NP:(ND:彼女) XX:は)(VP:(NP:(VP:(NP:(ND:かわいい顔) XX:を) VB:した)(ND:人形) XX:を) VB:持っていた))

ここで、NDは名詞を、XXは助詞を、VBは動詞という品詞を表します。単語のレベルではこの品詞をつかいカテゴリー分けします。作業中には入れ子などがありますから、入れ子をカテゴリー分けして、作業品詞として、NPとかVPとかをたてます。でも、物理的な文法においては、NPはNDとして振る舞いますし、VPはVBとして振る舞います。

このように品詞や作業品詞を導入すれば、文は綺麗に処理単位ごとに分離できるようになります。分離した単位ごとに、文解析プロセス群は並列処理して、局所的な操作で解析結果を集約していけばいいのです。並列度を上げることができます。

分散問題解決では並行処理が重要になりますが、それは曖昧性への対処としても有効になります。複数の解釈が可能であれば、それぞれの解釈に対して、後続処理を発生させて、平行で解析を進めます。そして、有効性をその発生された処理の中で判定して、有効であれば、値を呼び出し元に返し、無効だったら死滅します。こうゆう、適者生存戦略が、黒板システムでは有効です。

いつの間にか、学習システムが文解析システムの話になってしまいました。学習システムも同様なことが行われるはず。学習の過程では文解析が行われるはずで、その上に、カテゴリーの構築プロセスが動くのです。その過程も黒板システム、すなわち分散問題解決プロセスなのです。

この作業品詞（入れ子）はフレーム問題の解決とか、プランニングの問題再起構造と深く関係しています。フレーム問題とは、ロボットが問題を解決するために参照するデータ（レストランとか会社というものを記述するデータ）が膨大（無限に考えられる）であるため、その取捨選択、参照や解析に時間が掛かり、ロボットは結局なにもできないというものです。ロボットが自然環境で生きていくために、どんなデータをフレームに設定しておくべきかも、あれこれ考えられるので、定まらないという問題でもあります。

これを解決するには、フレームを入れ子構造の階層構造にして、大まかな記述を上位のフレームに置いて、必要な時に詳細のフレームを参照する、心の中のイメージを解析して得る（内的パターン認識）というようにする必要があります。しかも、このフレーム構造はダイナミックに自動で組み替え可能とする。元々は、構造のない世界を記述するデータがあって、カテゴリー化システムを独立に設けて、フレームの組織化を行っていくのですね。

プランニングもロボットが行動するとき、計画を練りますが、その計画を練ってできるデータには大まかなものから、その詳細なものへと階層構造を持たせなくてはならないというものです。入れ子構造（再起構造）をしているデータが作られるのです。

入れ子構造はこのように重要なものです。それはカテゴリー化を担うプログラムが担当するものでしょう。

第5章 曖昧性の管理

曖昧性は作業品詞として管理すべし、というのがもっとも新しい私の考えです。文脈もフレーム構造を内部エンティティに持っている作業品詞と考えるのです。そうすれば、全ての言語処理はカテゴリー分類の問題をベースに議論ができるでしょう。学習とはカテゴリーを得ることなのです。

曖昧性は全てなくすべきというものでは無いのかもしれませんが。機械翻訳をするのであれば、曖昧性は天敵かもしれませんが、日本語理解では、曖昧性は曖昧性そのまま、味があると見るべきではないでしょうか。

（例文1）信子は明雄が好き。

【解説】

この文は、次の2通りの解釈があります。

（1）信子は明雄のことが好きだ。

（2）明雄は信子のことが好きだ。

この2つの意味はこのまま詮索せずに、必要に応じて、利用していくべきでしょう。文の前後関係から（文脈から）どちらかが分かれば、その旨を作業品詞に付していけばいいのです。これをするのは、結構、負担の多い作業ですけれども。

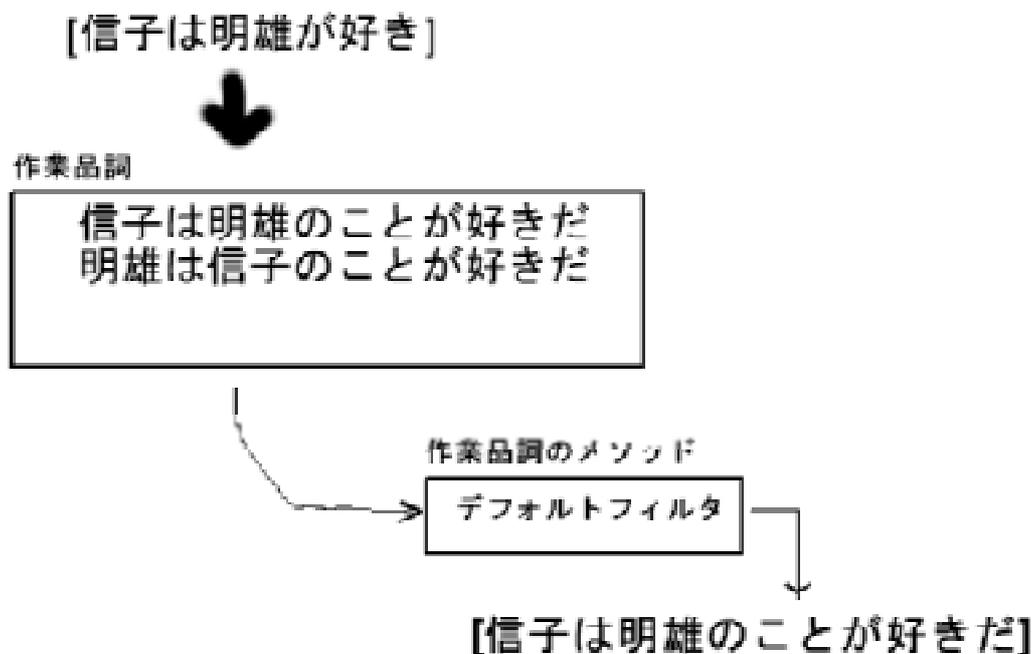


図5.1 曖昧性管理の作業品詞

（例文2）私はウナギだ。

【説明】

この文は、いくつにも解釈できます。

（1）I am an eel.

（2）私はウナギが食べたい。

（3）私はウナギを買う。

などなど。文脈を考慮しなくてはなりません。こうゆう曖昧性をもっていることは、談話の場合には普通のことです。談話からも言語は学習しなくてはなりません。そうした曖昧性を作業品詞を立てて管理していくのです。どんな文脈ではどんな事を言っているのかという、コーパスを作って、文法として学習していく必要があります。まさに、カテゴリー分けの醍醐味であります。

カテゴリーは多くの下部組織を持ったエンティティなのです。

第6章 文法と意味の表現

今までの章であげてきた事柄から、文法と意味の表現としてどんなことが重要か考えていきたいと思います。

6.1 記号とイメージ

人間が学習していくとき、物事のイメージを明確にしようとしています。そして、そのイメージ群をあれこれいじり回します。それは思考という過程になり、新しい発見へと成長していきます。このように、自ら内省しても、イメージの重要さは推し量れるところです。だから、付録に示しますような意味プリミティブプリミティブを記号で表しても、その背後にはボクセルイメージとか、イメージ生成プロセス(メソッド)とかがあべきものなのです。

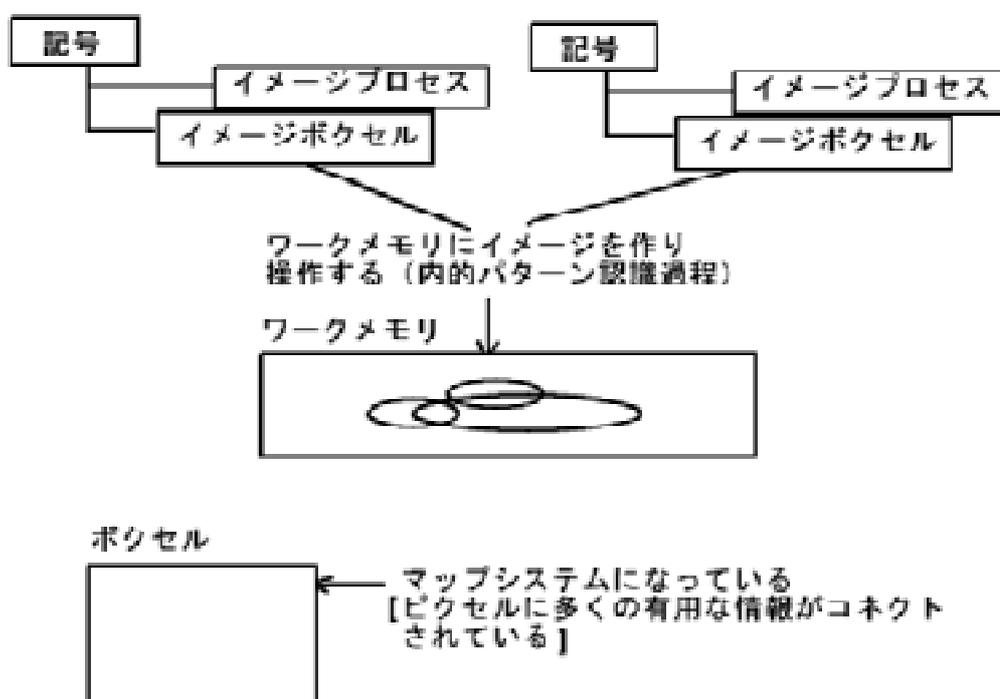


図6.1 記号とイメージ

次に主なイメージの形式を考察していきます。

(1) 集合論理

集合論理は「だけ」とか、「の中の1つ」とか、言語機能の重要な要素です。これはしかも、推論を実行していく際に土台となる意味要素ですから、是非実現しなくてはなりません。

基本的にマップシステムです。マップの各ピクセルにどんな情報を付加するかが応用問題となります。

(2) ことあげ詞

「も」、「しか」、「だけ」などのことあげ詞は集合論理に話者の感情が盛り込まれたものです。だから、マップシステムで表現できます。

(3) 時間格、場所格

次のことを考慮しなくてはなりません。

- ・ある時点、ある一つの場所を表す表現が必要
- ・ある有限な範囲を表す表現が必要
- ・半直無限遠を表現することが必要
- ・繰り返しパターンを表現することが必要
- ・これらの組み合わせも表現できなくてはならない

これらのことは、マップシステムで表現できます。マップのピクセルにマップを接続することも必要になってきますが。

(4) 動詞

操作や運動などは、プロセスで利用時にイメージを描画するというようなことで仮想的な記号でイメージを表現しなくてはなりません。格については単なる構造ですから、マップシステムで表現できるでしょう。

(5) その他

このほかに、類推機構もなければなりません。比喻を行える機能は言語処理には必須になるからです。

類推は、内容からのワイルドカード検索ということで、記号、イメージの集合をキーにしたワイルドカードデータ検索になります。これは、結局インデックスシステムをいかに作り込むかという問題になります。ですが、なんでも作ればいいたらうという議論にはなりません。知識とは膨大なデータになりますから、全件データに内容検索用のインデックスを張るのは不可能でしょう。となれば、文脈機能に注目して、今必要なものだけをインデックスシステムに登録しておく、という方向になります。

これは、実は脳の中でも行われていることです。プライミング効果という脳の働きです。ある単語を提示されると、関連語の検索が速く行えるようになるという、脳の自動機能です。前頭葉にこの機能があるとか。それを参考にして、工夫していけば、効率的な類推が行えるシステムが作れそうです。

類推が内容検索であるとは次のようなことから言えます。

野球のインニングは攻撃と守りを交互に繰り返して行きます。このことはコインと類推がなされることですね。それは、インニングが2状態からなるということがパターン認識(イメージベースでのパターン認識とか、「交互」という記号の意味からの推論)で得られます。これで、2状態を持つものという内容検索から、コインが表裏2状態を持つことから「コイン」が検索されるでしょう。

第7章 まとめ

曖昧性を許容しながら、カテゴリーの順次精密化を図る過程が学習システムです。いろんな実現方向があると思います。その中で、人間にもっとも近い学習効果があるシステムが生き残っていくでしょう。これから、その試行錯誤が始まるのです。

本論では、日本語文法の現象を意味の問題からざっと見てきました。それを文法学習という面から捉え直し、さらに学習システムのポイントを考察してきました。最後に、システムのインプリメントを意識して、文法の内部表現に至っています。

ところで、学習は

- ・データの蓄積とカテゴリー化
- ・プロセスの作成と改善

で行われます。

プロセスでの学習はコンパイラを考えると大変ですが、プログラムデザインパターンのコマンドパターンを利用すれば、コマンド要素のセットというデータ列に帰着されます。データベースのデータと違って、データの生起順序が重要になりますが、それは自然言語とて同じです。語順は重要です。順序を持ったデータとして総合して捉えて、その獲得と改変、カテゴリー化を考えていけばいいのだと思います。

もう少しで、学習システムの開発に着手できるところまでいけますでしょうか。どうでしょうか。夢が膨らみます。

以上