

考察：「カテゴリーシステム」

前考察「日本語文学習システム」を、カテゴリー管理に注目して更にブレイクダウンして、考えていこうと思います。

カテゴリーはどのような形式をしているべきかということを先ず考え、その中で、文脈空間という、マップシステムを提案します。全ての有用な意味を記号に展開してもっていることは不可能で、イメージで持っていて、必要に応じてパターン認識を行い、記号化していく、そんな考えです。これはもう5年くらい前に考えていたことですが、どうもシステムとして複雑になり嵩張ったものとなりそうで、なにかエレガントな解はないかと棚上げにしていました。それで、1年前の「動画システム」に至ったのですが、やはりフレーム問題はクリアできていません。それで、ここにこの「文脈空間」案を公開しようとしたわけです。この考察のトピックスとなります。

そして、作業品詞について、前考察「日本語文学習システム」に続いて機能の考察を深めていきます。内容検索についても触れます。文脈というものにも触れます。プロセスの学習というものも考えてみます。そうして、作業品詞の生成、構造化、といったことを追求して、カテゴリーシステムの基本的な機能を明らかにしたいと思います。その中で、確信度という確率に代わる考え方を導入したいとおもいます。

最後に、和文解析を例にしてカテゴリーシステムの使われ方をおさらいしていきます。

この考察は以下の本を読みながら行ってきたものです。

- ・「認知心理学」 東京大学出版会
- ・「グラフィック認知心理学」 森敏昭、井上毅、松井孝雄著 サイエンス社
- ・「現代心理学入門：認知心理学」 守一雄著 岩波書店
- ・「ダイナミックメモリ」 シャンク著 近代科学社
- ・「情報処理心理学入門」 リンゼイ、ノーマン著 サイエンス社
- ・「子どもはことばをからだで覚える」 正高信男著 中公新書

第1章 はじめに

日本語解析をしていて、最初にカテゴリーというものの威力に触れたのは、品詞というものでした。これによって、文というものの構造が明晰に捉えられ、解析ができるようになる。ただの単語の並びではなく、計算機でも解析できる、法則を持ったものにかわる。そして、品詞とは数えるほどしかないものでした。無数の単語がそんな品詞というもので分類されること、それは、混沌としているように見える言葉に、美しい規則性、法則性があることを示唆します。そして、意味を計算機で扱うことへの大きな励ましとなってくれます。

もう一つ明確なものは入れ子構造です。単語だけでなく、文の一部の単語列(句)が単語相当の品詞(名詞句 = 名詞、副詞句 = 副詞、動詞句 = 動詞というように)として振る舞うことです。

ただ、現在のところ、明確となっているのは、単語と品詞と入れ子構造というこの3者だけです。単語の持つ意味のレベルにはいると、話は混沌としてきます。なにか規則のようなものがあるのですが、焦点が結べない。多くの曖昧性が覆っています。

意味とは本来曖昧性のあるものです。「山」といっても、頂上付近をイメージできますし、中腹から上をイメージしても正解です。起伏のある一帯を山とも呼ぶことがあります。アフォーダンス理論というものもあります。コップは水を飲むもので、乗っかったりしたら壊れます。しかし、これは人間の視点で観たもの。蟻にしてみれば、コップは乗っかるものです。つまり、意味は視点によって異なってくるのです。そんな言葉の世界の中で、単語、品詞、入れ子構造という、情報処理の基盤は実に明晰なものとして驚異に映ります。そして、意味処理を成功裏に行うにはこの3者を手本に考えてシステムを構築するべきではないか、・・・と思わせます。その手本とは、

- (1) 曖昧性を作業品詞として代表させてしまう。
- (2) 作業品詞を分類していく。意味は無限にあるかもしれないので、作業品詞は無限個あるかもしれない。そのことを考慮した分類であること。
- (3) 入れ子構造を中核の機構としてもつ体系とすること。
- (4) 作業品詞は意味データを持つだけでなく、機能(データ処理能力)も持つものである。オブジェクト指向がぴったりと思えます。

本考察は以下の順序で論を展開していきます。

- (1) カテゴリーシステムで管理するデータとカテゴリーシステムの動作概略
- (2) 確信度によるカテゴリー管理
- (3) 文脈空間についてのアイデア紹介
- (4) 作業品詞の表現についてアイデア紹介
- (5) カテゴリーシステムの管理手法
- (6) カテゴリーシステムを支える技術(アイデア紹介)
- (7) 日本語処理への応用例

第2章 カテゴリーシステム概要

作業品詞は状況に応じて、次のデータを管理していかなければなりません。

- (1) 意味記号
- (2) イメージ
- (3) 管理情報
- (4) 作業プロセス
- (5) 集積データ
- (6) 関連データ

作業品詞が単語そのものである場合、単語の意味定義を持たねばなりません。「りんご」ですと、

「りんご」: = 概念: 「real-object」
属性: 「shape=round color=red/green weight=light . . .」
記事: 「human eat (this)」
.

「りんご」の全ての属性を書き表すことはできません。記事もです。どうしても、イメージが必要です。ボクセルマップのイメージでりんごデータを持ち、必要に応じて、内部パターン認識を行って、必要な情報を得るようにするのです。このパターン認識プロセスは(4)の作業プロセスの一部でもあります。

管理情報としては、作業品詞の分類コードとか、何を管理する作業品詞なのかとか、何を目的にしているのかとか、いろいろ考えられます。ここは、必要に応じて追加されるものなので、固定エントリー数にはしないで、いつでも追加更新できるようにしておくべきです。単語、品詞、句、曖昧句管理、多義語管理とか、あげていけばきりがありません。方針としては、「なんでも新しいカテゴリーには作業品詞を立てましょう」というのを取ろうと思います。あとで、便利さを考慮して、大きなカテゴリーを設け、それでくくればいいのです。入れ子構造という便利な考え方もありますし。

作業プロセスとして、今あげた、パターン認識の他に、デフォルト抽出とか、文脈によるバイアス処理とか、確信度決定とか、分類処理(カテゴリー管理機能)とかあります。カテゴリーシステムのフレームワークでこういうものを持つべきかもしれませんが、システムを漸次高度化していくという方針なら、作業品詞にローカルな処理を担わせた方がいいように思います。バランスの問題でしょうが。

集積データとして念頭にあるのはコーパスです。この作業品詞が関連する言語現象を逐一スタックしていき、作業プロセスでCPUが空いたときに分析していく。そのカテゴリー再編のための情報を集めるためのネタとなるデータです。

関連データはコネクションされた連想データです。文脈処理などに使います。

2.1 カテゴリーシステムの位置づけ

カテゴリーシステムは知識ベースの根幹です。知識がカテゴリーで管理されないことはないのです。運動系もカテゴリーシステムを持つことは、プランニング技術で語るところです。従って、人工知能でのカテゴリーシステムを位置づけるとすれば、図2.1のようになります。

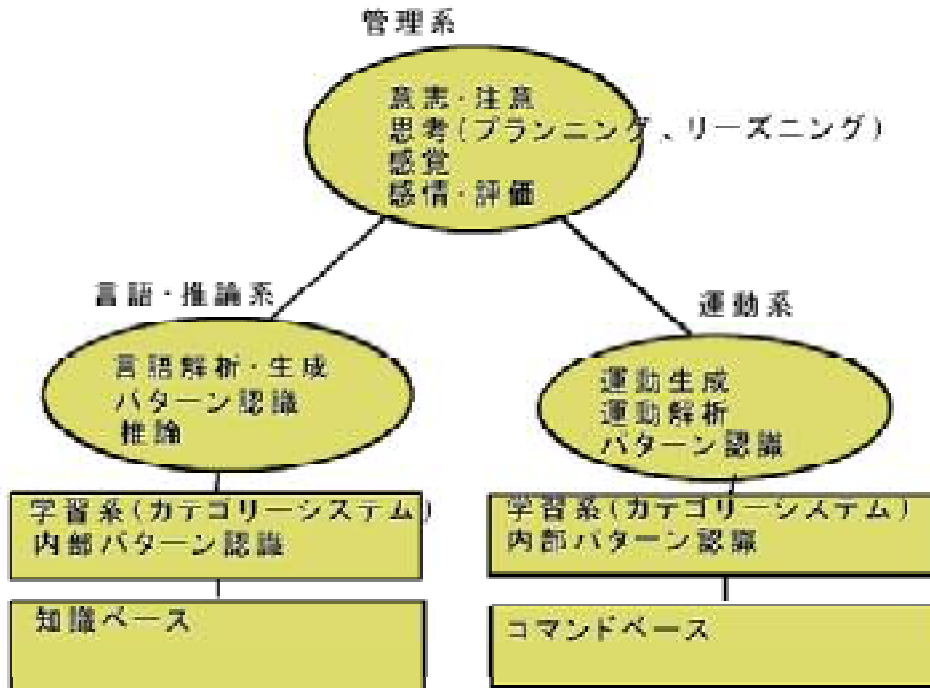


図2.1 カテゴリーシステムの位置づけ

意志がなければ、システムは存在する意味がありません。計算機システムだって、自動車だって、人間が意志系になっているのです。自律型の人工知能を目指すとなれば、意志系を実現しなくてはなりません。ものの価値は感情系、外部入力に感覚系、外部出力は運動系です。

思考は毎日の生活の中で、常に作動しているものです。「なぜ?」、「なにが?」、「いつ、どこで?」なんて、いつも自分に問いかけ、知識ベースを更新し、最新な情報の維持につとめています。特に、物事の因果関係の把握には力が入っていて、外界への確に反応できるように、常時、不測の事態に備えています。オートノミクス実現にはこの辺の技術が重要になるとおもいます。

そして、言語系があって、その下に学習系として、カテゴリーシステムが位置して、知識ベースを紡いでいる。そんな図面が描けると思います。人工知能は巨大なシステムですね。言語系だけでも生きていく内になにか結果だせるか自信ないですが、それくらい大きなシステムですが、人工知能としてみれば、わずかな部分でしかないのです。

2.2 カテゴリーシステムで管理するデータ

以下、言語・推論系の学習系のカテゴリーシステムを考えていきます。

カテゴリーシステムが作動するのは次のような状況でしょう。

(1) 曖昧性管理

単語の多義性、句や文の多義性をカテゴリーとして隠蔽し管理します。

曖昧性の種類ごとにカテゴリーを立てて、その配下に意味を担う記号列やイメージを配置していきます。一つの意味にはキーホールとなるデータエリア (Java のオブジェクト相当) を添付して、特定のキーによって、この意味が選択できるようにします。

(2) 文脈管理

曖昧性管理を含みますが、他に、アフォーダンスとかの視点によって変わってくる状況を管理します。また、プライミング効果を実現し、内容検索や曖昧性低減の効率化を可能とします。

文脈は単語や句、文といった知識ベースの基本エントリのアドレスのセットです。文脈の中に、こういった基盤データを置くのでは、組み替えとかが負担となりますし、複数の文脈で共通の基盤データを持つなどの無駄なことをしなくてはなりません。アドレス集合を表す情報だけをもつという、軽いインプリメントがよろしいと思えます。

(3) 知識の組織化

生物分類や人体の構造 (体、頭、胴、手、足・・・) とかを参考にすれば、知識を構造化して管理すると便利なのが分かります。この知識の構造化を柔軟に実現しようというものです。

これも、文脈と同じで、構造情報を担うカテゴリーを立てて、構造をフレームで表現したデータエリアを保持させます。そのフレームのスロットに基盤データのアドレスを登録する、というふうに行います。構造の連鎖も構造カテゴリーを基盤データと見なして、同様に管理していくことができます。

(4) コーパス分析

文を無数に集めて分析し、確固とした意味関係、知識を構築するための機能です。

プロセスを管理することになります。のべつ動作しているプロセス群を管理していきます。プロセスをも学習によって構築していくべきかどうかは、作り込んでいく中で見えてくるものだと思います。

(5) イメージから意味を抽出するパターン認識

イメージでものや状況を持っていなければ、フレーム問題に対応できないと申しました。しかし、イメージは生では利用できません。いったん意味記号を抽出して、それで、計算機処理 (記号処理) することになります。このようなイメージを意味記号に変換する処理がパターン認識機構です。

プロットタイプのイメージ間の関係を解析するのが基本ですから、感覚情報のパターン認識よりは易しい課題です。この機構として、文脈空間というものを案出しました。

カテゴリーシステムが基本的に持っているべきカテゴリーは次のものが考えられます。

大分類	基本カテゴリー
語	名詞 動詞 属性詞（形容詞、形容動詞、副詞） 機能詞（助詞、助動詞、接続詞）
構成	文脈 構造 コンポジション 集約
オブジェクト	記号 イメージ 論理 集合 メトリクス 数値 時間 場所、枠組み
表現	コネクション フレーム ボクセル（文脈空間） 意味列
データ	テキストコーパス イメージコーパス 意味列コーパス フレーム表現コーパス

表 2 . 1 基本カテゴリー

2.3 カテゴリーシステムの動作概略

一番最初に、そして最も重要な処理は、新規の事象が提示されたのか、もう既に知っていることが提示されたのかを判断することです。人間にとって、簡単な作業ですが、人工知能として実現するにはどうでしょう、かなり難しい課題だと感じます。しかし、なんとでも、ここは突破したいところ。ところで、新規性とはどんなものがあるでしょう。

(1) 形の新規性

(2) 音声パターンの新規性

(3) 構造の新規性

(4) カテゴリーとしての新規性

と、4つのレベルが考えられます。

(1) (2) はパターン認識の中でも基本的な、テンプレートマッチング課題でしょう。つまり、心的イメージ(知識となっているイメージパターン)との完全マッチング(少しは誤差を許すとは思いますが)の問題です。同じ人物でも、服装が変われば、新鮮に見えます。(4) のカテゴリーの新規性が無くても、イメージレベルは新規なのです。

(2) の構造の新規性は、新規イメージとして弁別されてから、イメージの構造が解析されます。別人でも、人は人の構造をしています。人を見たときには、このレベルでは新規性を失うわけです。これも、構造を管理するカテゴリーの全件検索をもって行う作業でしょう。そう、全件検索をしなくてはならないのです。だから、検索効率のよい、メモリ構造にしておかねばなりません。

最後に(4) のカテゴリー毎の新規性が判定されます。これも全件検索となります。並列処理を駆使して、しかもメモリ構造、インデックス構造を工夫していかねばなりません。

既存ならばそれで、カテゴリーシステムの処理は終わります。新規の場合、カテゴリーシステムに新規データを追加しなくてはなりません。(1) から(4) までの処理を行うカテゴリーシステムのサブシステムの配下にそれぞれ、コーパスとして登録していくのが素直です。分散して、重複して登録してもかまわないはず。必要ないところでは、順次死滅させればいいのです。適者生存戦略が一番エレガントな解決手段となります。登録されたところで、重要な位置を占めれば、生き残って、永久に役に立つというわけです。

此処までが第一段階です。次に、カテゴリー生成です。教示によってカテゴリーを起こします。教示は人が言葉や身振りで教えることもそうですが、一般に区切り情報が外界から得られた場合へ拡張して考えます。エレベーターの中に入ったら、そこは外界から分離されている空間ですから、そこには新しい文脈カテゴリーが生じます。この区切りは生得的なものです。また、文章でも「空白」とかはそうですね。でも、学習によって区切りを表す情報を豊かにしてゆくこともあります。カンマとかピリオドとか、そうしたものです。そうした区切りによるきっかけで、カテゴリーを作ってゆく。

また、カテゴリーシステムの中で、新規性の規則、パターンを発見したら、カテゴリーを設立するという、内的な機構も重要です。

カテゴリーができたら、その内容の充実化を図っていくこととなります。口口という犬を飼ったとします。初めは、同じ犬でも、特別なものということで、犬とは独立にカテ

ゴリー、口口をつくります。そうして、生活するうちに、性格発見とか、懐かしいエピソードとかで、口口のカテゴリーを豊かにしていきます。家族の他の成員とかとの関係もカテゴリーシステムの中に分散して持ったりして、カテゴリーシステムも成長していきます。そうした、成長するものが学習で、知識ベースの根本機構として重要な地位を占めるのです。

カテゴリーシステムは無秩序ではありません。生得的な制約が備わっています。それは、学習を効率よく行うための戦術みたいなものです。例えば、マークマンの制約条件というものがあります。赤ちゃんが言葉を獲得していくときに使うカテゴリー分けの方略として指摘されたものです。

(1) 事物全体制約

象の絵の一部分を指して「象」と教えられたときも、赤ちゃんは絵全体を象と認識します。

(2) 相互排他性

象を覚えた赤ちゃんに鼻といって、特定の部署を指すと、象が「鼻」のではなく、指された部分に局限して、「鼻」と認識します。命名(カテゴリー)がだぶることはないという制約です。

(3) 類制約

「コップ」といって、特定の水を飲む容器を教えられると、その特定のコップではなく、水を飲むもの全部をコップと認識します。

さまざまなカテゴリー学習の方略があると思います。場合に応じて、使い分けられて、利用されていくのでしょう。

第3章 確信度によるカテゴリー管理

曖昧性があると、曖昧要素を選択して利用していかねばなりません。選択の優先順位、それを、確信度として表現して、利用していきましょう。ファジー性ですね。とにかく、総和が1になる正数という条件をはずしたファジー性を確信度としています。確率みたいに総和が1の正数というものを導入すれば、ベイズ推論とか、多くの数学的な定理が使えるのですが、そうすると欠点もあります。確率の小さい時は誤差のなかにデータが埋もれてしまうとか、負の信念が扱えないとか、いろいろ・・・。

確信度では、負の確信度も持とうとおもいます。この事象があれば、こうゆう事はないとか、いいたいですものね。

確信度では、数学的な扱いは得意とせず、パターン解析で物事を処理していきたいと考えます。「Aの確信度が0.5で、Bの確信度が、0.8で、Cの確信度が0.1だから、B Xと判断しよう。」というように確信度を使います。ものごと、それで十分です。

第4章 文脈空間

イメージを用いた、カテゴリー分析、意味処理を考えてみたいとおもいます。

イメージは様々な情報を担っているのです、基本的にマップシステムです。このマップシステムを操作、解析、検索する API を作り込んだものが文脈空間と筆写がよんでいるものです。

人間のイメージを思い浮かべてみれば分かりますように、頭とか、胴体とか、また幾つかのイメージに連想（コネクト）していきます。文脈空間は複数のマップシステムで構成されるのが普通です。そして、コネクションはマップ間に張られるものと、一つのマップのなかのピクセルが他のマップに張られるものと、他のマップの複数のピクセルに張られるものがあります。

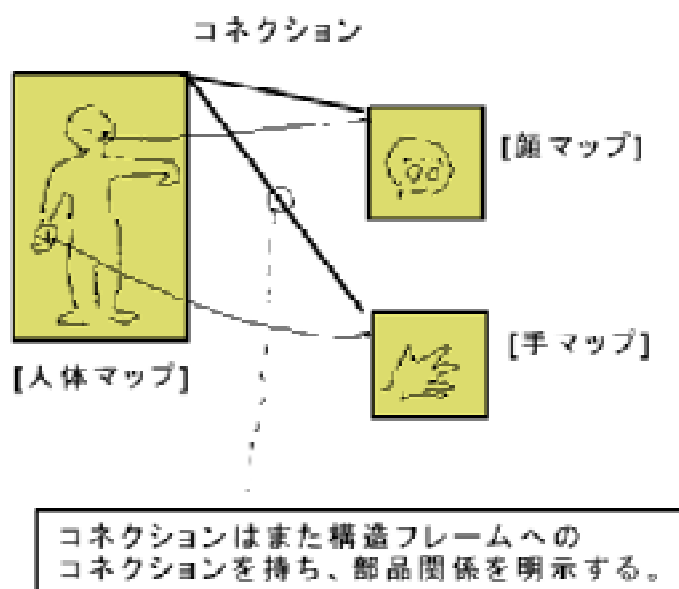


図4.1 文脈空間の要素

文脈空間はこのように物を表現するだけでなく、推論のための作業空間でもあります。例えば、「太郎が花子にボールを投げる。」では、イメージから、ボールの移動、花子のボールのキャッチとか、太郎の投げる時間は花子の受け取る時間の前であるとかを推論しなくなります。そんなとき、作業用の文脈空間に太郎と花子、ボールのイメージを作って、それをパターン認識して行います。

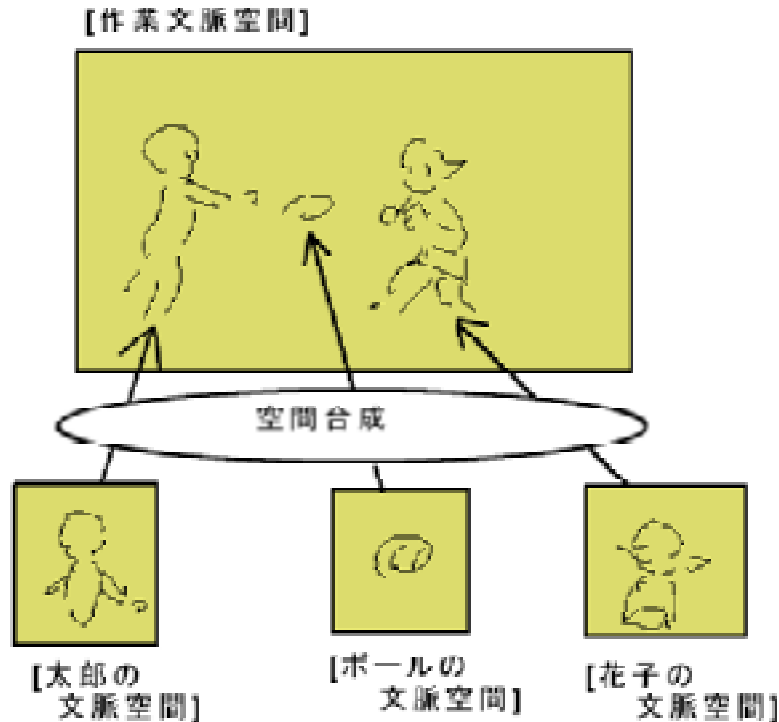


図4.2 文脈空間の合成

文脈空間にはパターン認識のための様々な API が付属しています。その例を上げましょう。

```

MapSystem ms=new MapSystem();    . . . . . マップシステムの作成
ms.connect(Ctype ct,Object o_a); . . . . . マップシステムへのオブジェクトの接続。Ctype は接続の型
ms.setInformation(Information information_a); . . . . . 情報の付加
ms.getInformation();              . . . . . 情報の取得
ms.setActionRelation(Object o_a,Object o_b,Action action); . . . . . オブジェクト o_a とオブジェクト o_b の間に行動 (Action) を設定する。
ms.getActionRelation(Object o_a,Object o_b); . . . . . 行動を得る。

```

大体、こんな API を沢山作っていくことになります。行動を開始させたり、停止させたり、同期を定義したり、といった細かい作業も文脈空間には必要です。行動のキーフレームも用意し、動画認識も可能としなければなりません。ものすごく大きなプログラムとなります。筆写が一度は断念した理由がおわかりでしょうか。

でも、やらねばならないような思いがこの頃しています。

第5章 作業品詞の表現

曖昧性はキーホール定義情報と意味が決定的の作業品詞の組を複数持つことで表現します。

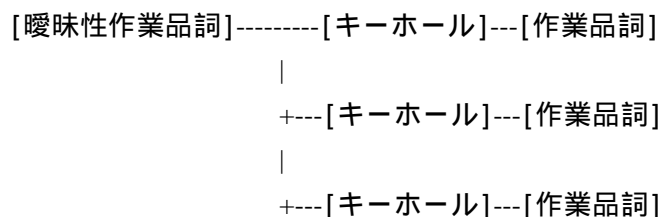


図5.1 曖昧性作業品詞の表現

あとは決定性の作業品詞の表現を定義していけばいいだけです。

作業品詞は利用できる関連データをデータの型を指定して、なんでもくくっていくことで成り立ちます。記号もイメージも記号列（文とか文章とか）も、コマンド列もフレームも、なんでも組み込んでいくのが良いようです。作業品詞に組み込むプロセスが判断して利用していけばいいのです。カテゴリーシステムとしては、なんでも作業品詞の下に取り込んでいく。この作業品詞に関係ない物をフィルターするだけです。

学習によって、物事を知っていくことが、人工知能において重要だと語りました。学習はどう行われるでしょう。人工知能も人間の世界で成長していくいじょう、人間と同じ学習手段が通用しなくてはなりません。とすると、主な手段は、

- (1) 物の提示
- (2) 映像の提示
- (3) 辞書、事典による言葉による説明

でしょう。(1)と(2)はイメージの取り込みとパターン認識によるデータ収集で、(3)は自然言語解析によるデータ収集です。イメージデータも自然言語データも解析されたものはフレームとして、動画システム（著者の第1論文参照）に落とし込んで利用することになります。しかし、作業品詞としては、表層表現として、イメージとか、自然言語を保持しているべきです。深層表現であるフレームは、ある程度の情報がもれている物だからです。システムとして漸次成長していくことを目指すならば、表層表現をずっと保持すべきものなのです。

(2)の映像の提示では、動画を解析しなければなりません。動きとか、操作が生じます。これはコマンド列で表現することになるでしょう。認知の統合理論では、ACT*モデルとか、Soar モデルとかがあるそうです（現代心理学入門1 認知心理学；守一雄著 岩波書店）。これにはプロダクションシステムが利用されています。プロダクションシステムというのは、「if・・・then アクション」という、ルールによってコマンドを制御していく方式です。この方式は、一世を風靡した、エキスパートシステムにも応用されています。

第6章 カテゴリーシステムの管理手法

カテゴリーシステムの動作概要は第2章で述べました。ここでは、考えられるデータマイニング手法を考察していこうと思います。データを分類する方法として、次のものがあります。

- (1) KJ法
 - (2) 距離を導入した分類手法
 - (3) 線形多次空間の点にマップして、線形分離する手法(サポートベクター・マシン)
- 著者が知っているのはこんなところです。(2)と(3)は、データ間の距離を使うという点で、大体同じ手法とっていいですね。

KJ法(川喜多二郎氏の発想法)は、次のステップを踏んで、データを把握する方法です。

- (1) テーマ決定
- (2) 情報を取得し、単位データ化
- (3) データのラベル化
- (4) 主観的に類似したラベルどうしを集める
- (5) ラベル群に内容を凝縮したラベルを与える
- (6) ラベルを更に上位群に集め、ラベルを与える。
- (*) 得心いくまで繰り返す。

図6.1 KJ法

(4)の過程が注目点です。「主観的に類似した」というのを、意味論的にどう表現するか。それは、意味記号(列)に、何かあるメトリックを持った属性(感情とか価値、嗜好)を対応させて、その強さを確信度とします。確信度が高くて、まだ作業品詞として独立していないならば、作業品詞を設立します。同じ意味記号(列)の確信度の高い作業品詞たちは、主観的類似度が高いと評価します。作業品詞に類似度が低い複数の意味記号(列)があって、どちらの記号列も確信度が高ければ、多義語と判断するという考えもできます。

確信度も類似度も、距離のある世界の話になりますから、カテゴリー化は皆、「いかに位相(距離)を導入するかが重要である」というふうに捉えることができそうです。

「ことがある」という言葉の文法化（連語として助動詞相当語になる）を例にあげ、作業品詞の創成を考えてみたい。



図 6 . 2 コーパスの収集

「こと」の下に「ことがある」を含む表現をスタックしていく。「ある」の下にも同様に表現をスタックしていく。

「ことがある」という連続が多数あれば、「ことがある」という作業品詞を設けて、「ことがある」の関連表現をコーパスとしてコピーしてくる。そうして、この組み合わせの間に他の語が来ることがないことを分析するプロセスを実行する。「ことがある」が決まった言い方ならば、晴れて、文法化の認定を行う。

「もっと」が様態の形容詞に掛かるが、数量形容詞には掛からないことの解析を次に見てみよう。

(例文) もっと大きい
 × もっと少し

「大きい」は意味として、[AN,metrix]を持ち、「少し」は意味として、[AN,number]を持つとする。



図 6 . 3 コーパスの収集

「もっと」のコーパスを見て、AN(形容詞)には接続しているが、metrixのみで、numberと接続していないのを知ります。そこで、新しい作業品詞を設けて、「もっと」を普通の副詞(AV)とは別格なものとしします。

第7章 カテゴリーシステムを支える技術

推論、とくに類推は言語処理や、プランニング、リーズニングといった、人工知能の骨格となる処理で重要な位置を占めます。類推には内容検索が基本処理としてあると、「日本語文学習システム」で述べました。そして、それは脳のプライミング効果とも関連してくるのだとも推測しています。

内容検索、それはインデックス管理の問題でもあります。作業品詞を全件検索して、所定の内容を持っているものを探すというのは、時間が掛かって実用的ではありません。内容をキーにして作業品詞を特定していくことが重要です。つまり、意味記号をキーにしたインデックスを張るわけです。その最も効率的と思える方法の概念図を図7.1に示します。

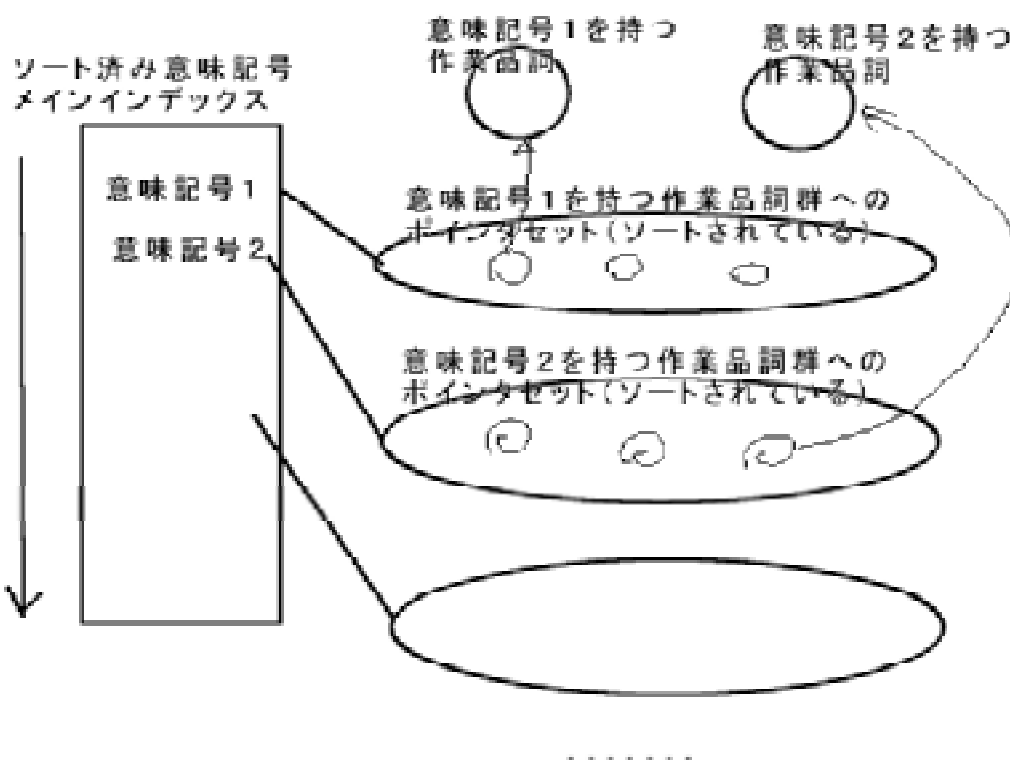


図7.1 内容検索用インデックスシステム

インデックスをソートしておけば、二分探索ができるので、効率です。更に、多層なインデックスシステムにされることを推奨します。並列探索も可能にできれば、それも推奨です。

知識ベースでは、インデックスを張って管理する単位という物がある、島のように、浮かんでいる。それが文脈では、そんなイメージが浮かびます。メインメモリに文脈をロードして、インデックスとして活性化するのに手間取ってはいけません。文脈単位に再配置可能となっているべきでしょう。それはインプリメント時にいかようにでも工夫がなれます。そうして、最新の文脈の優先度を最も高くするという管理もしなくてはなり

ません。それには、超整理法で提唱されている方法が良いかもしれません。文脈が書類に対応するわけですね。

第8章 日本語処理への応用

日本語文解析を考えてみたい。日本語文解析は図8.1のようなステップを踏むと考えられます。

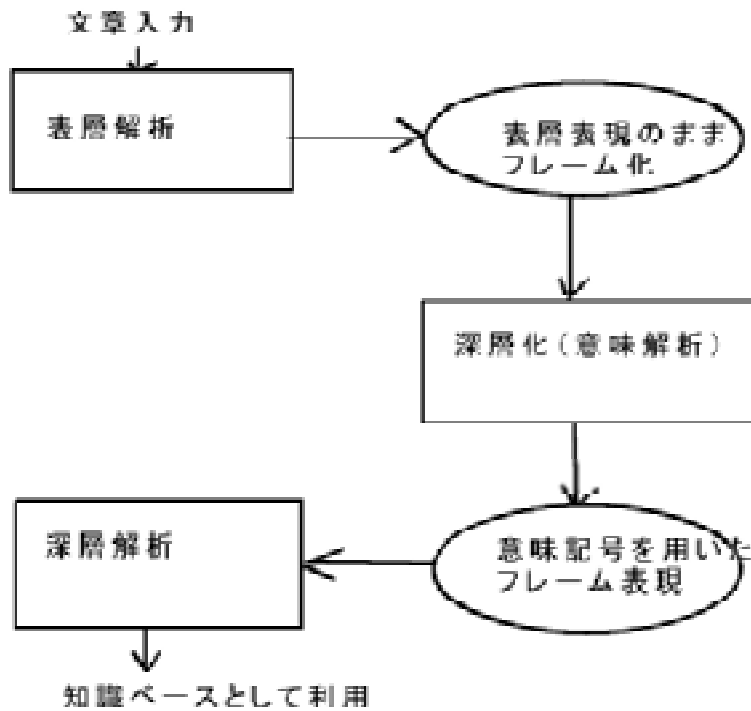


図8.1 日本語文解析の概略

(例文) 私は赤いリンゴが好きだ。

この文は、次のようなフレームに先ず落とすべきでしょう。

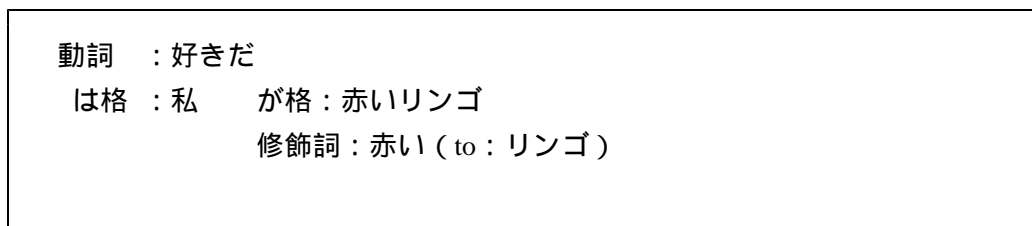


図8.2 表層フレーム

これが、「(は)(が)(好き)」パターン（作業品詞として知識ベースに用意する）によって、次のように解釈されます。

（解釈 1）I LIKE (RED)APPLE. 主題：I

（解釈 2）(RED)APPLE LIKE I 主題：I

しかし、APPLE の属性で（APPLE の作業品詞に記述）「好き」という感情はもたないことを得て、解釈 2 を落とします。

そして、深層フレームをつくります。

動詞	: LIKE
主格	: I
目的格	: (RED)APPLE
テーマ	: I

図 8 . 3 深層フレーム

このようにして、作業品詞は利用していきます。

次に、意味の等価性解析について考えてみます。

（例文 A）太郎は花子を説得した。

（例文 B）太郎の話を聞いて、花子は考えを変えた。

これら 2 つの文が等価であると、どう判断したらよろしいでしょうか。とにかく、「説教した」の意味が的確に設定されていなければなりません。

「説教した」の意味： X say something to Y. Y change Y's opinion.

というのがあって、

（例文 A）は、「Taro say something to Hanako. Hanako change Hanako's opinion」

（例文 B）は、「Taro's saying change Hanako's idea」

そして、「opinion」と「idea」が類似意味という知識があって、「saying」が「say something」と等価という知識があれば、結論として、（例文 A）と（例文 B）が等価であると結論づけられるのではないのでしょうか。

こうしたアイデアは過設計（過度に課題に特化した解法になっていて、応用ができないこと）に陥りやすいので、地道な研究活動が必要です。

第9章 まとめ

「日本語文学習システム」の考察を更に深め、インプリメントに向け大部前進したと思います。ただ、目標が具体化を深めると、そのシステムの巨大さ、複雑さが度を増してきます。ここまで、考えてきたことを更に整理して、エレガントな解法がないか探すことも有意義なことでしょう。特に文脈空間が大きなシステムです。CG とか、パターン認識とかと交差します。びびらずに、勇気を持って突き詰めていけば、あるいは新しい天地が開けるかもしれません。文脈空間の考察には5年かけましたから。更に5年くらい……。

これまで述べてきたカテゴリーシステムは、まとめてみれば、帰納的推論のプロセスの応用ということになります。つまり、次のステップを踏んで作業品詞を確定していくことです。

(1) 事例獲得

- ・コーパス収集

(2) 仮説形成

- ・作業品詞を設けて、そこにコーパスを集積していく。
- ・利用可能ヒューリスティック（発見法）を実施する。
- ・一般化を目指す。

(3) 仮説検証

- ・使わない作業品詞は死滅させる（削除とか、マスクをかけるとか）。

本論が何かのお役に立つことを願っています。

以上