

# 連想型多次元シソーラスとその意味解析への適応性

森田陽介

宮崎正弘

新潟大学大学院自然科学研究科

## 1 はじめに

日本語意味解析において、従来の格パターンとシソーラスをベースにした意味解析の限界が指摘されている。従来のシソーラスは語や概念を基本的に上位-下位関係で分類しているが、単語や概念の分類観点には上位-下位関係以外に、反意関係、部分-全体関係、静的属性、動的属性、機能、源泉など多くの観点が存在しており、分類が不明確である。格パターンにおいては、名詞の選択制限記述の網羅性が保証されておらず、選択制限への反例の存在や、類語の微妙な意味やニュアンスの差を記述する問題、メタファーやメトニミーなどの比喩が扱いきれない。

本稿では、より高度な意味解析を実現するため、多くの分類観点を持ち、単語連想機能を持たせた大規模な日本語の連想型多次元シソーラスを提案し、それを用いた日本語意味解析への適応性について論じる。また、それによって格パターンと従来型のシソーラスをベースにした意味解析の技術的行き詰まりを打破する新しい意味解析が可能となることを論じる。

## 2 多次元シソーラス

### 2.1 従来型シソーラスの問題点

従来、語(概念)間の関係を表すシソーラスは、同義・類義関係に基づき、基本的には上位-下位関係(包含関係)という単一の分類観点から分類しており、分類体系全体を木構造やグラフ構造で表現している。しかし、分類観点には上位-下位関係以外に、反意関係、部分-全体関係、静的属性、動的属性、機能、源泉などの様々な観点が存在し、上位-下位関係でも詳しく見ると種々の観点が混在している。例えば、『人工物』を『食料』『衣類』『道具』のように分類する観点は、

通常では上位-下位関係とみなしているが、詳しく見ると機能による分類であると考えられる。さらに、語と語の関係は分類観点によって近さが異なってくる。例えば、「鳥」と「飛行機」は上位-下位関係から見ると『動物』と『乗り物』となり近い関係とは言えないが、「飛ぶ」という動的属性に着目することで極めて近い関係となる。このように、単一の分類観点から分類した場合、分類観点が不明確になり、分類観点による語と語の関係の変化を扱えない。また、分類観点の順序によってシソーラスの構造は異なってくる。例えば、具体物を+liv(生きている)という観点を第一の分類観点とするか、+art(人工の)という観点を第一の分類観点とするかによってシソーラスの構造も大きく変わってしまう。これらの問題点を簡潔にまとめると次のようになる。

#### ● 語と語の間の様々な関係を表現できない

- 1 分類観点が一つ
- 2 分類観点が不明確
- 3 分類順序が固定

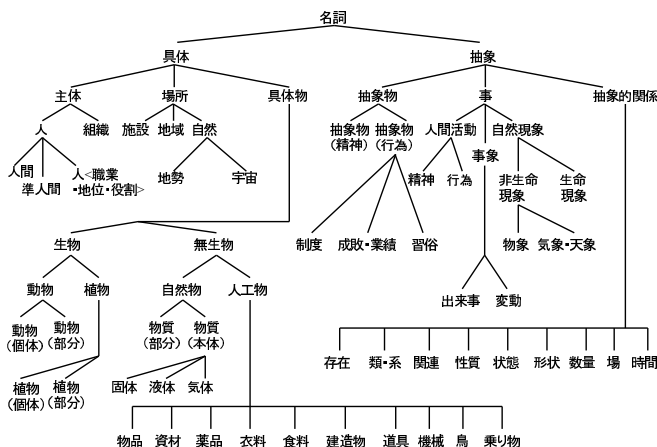


図1: 従来型シソーラスの例 [1]

## 2.2 多次元シソーラスによる問題点の解決

多次元シソーラスは 2.1 で述べたような従来型シソーラスの問題点を解消するために提案された柔構造型のシソーラスである。

本シソーラスは、語と語を表 1 に示すような種々の分類観点から明確に関連付け、分類観点による語と語の関係の変化を扱ことができ、分類順序を排除することによってノードの拡張性を増している。簡単にまとめると次のようになる。

- 語と語の間の様々な関係を表現できない → できる

1 分類観点が一つ → 分類観点が複数

2 分類観点が不明確 → 分類観点が明確

3 分類順序が固定 → 分類順序を排除

表 1：属性一覧表

略号	属性名	例
isa	上位-下位	「鳥」 > 「燕」
ant	対/反意	「上」 < > 「下」
has	全体-部分	「鳥」 > 「翼」
com	構成要素	「会社」 > 「社員」
con	内容	「グラス」 > 「飲み物」
sa	静的属性	「空」 > 「広い」
da	動的属性	「鳥」 > 「飛ぶ」
use	機能/用途	「本」 > 「読む」
cau	源泉	「本」 > 「書く」
mat	原材料	「土器」 > 「土」
loc	場所	事物が生起・存在する空間
tim	時間	事物が生起・存在する時点
seq	連鎖	「出発」 > 「到着」
pal	並列	「売る」 < > 「買う」
mea	手段	「乗る」 > 「行く」

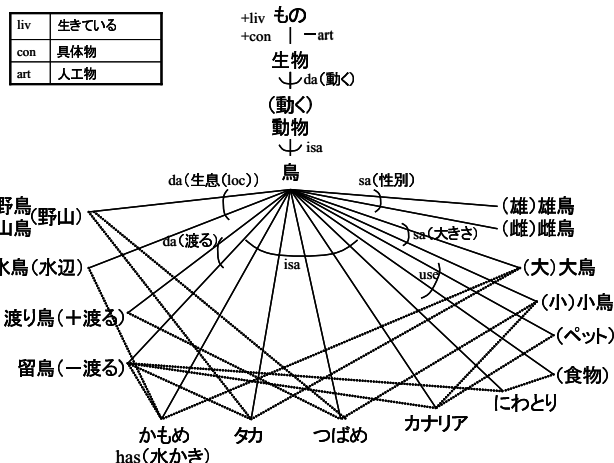


図 2：多次元シソーラスの例 [2]

## 3 連想型データベース

膨大な知識をコンピュータ内に蓄積しても必要な時、必要な知識を効率的に検索できなければ処理の爆発が生じる。そのように問題を解決するものとして、連想に基づく知識の効率検索を行うことができる連想型データベースを提案する。

連想型データベースとは、従来の上位-下位関係を利用した木構造型データベースでは見つけることが出来ない単語間の新たな関連性を見つけることができるデータベースである。

図 3 に連想型データベースの基本的枠組みを示す。

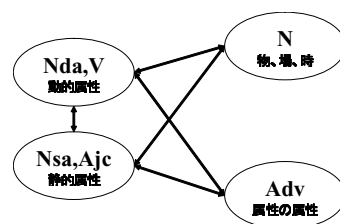


図 3：連想型データベースの枠組み

名詞類、動詞類、形容詞類、副詞類間で連想機能を持つ。連想機能により刺激語から連想語が抽出されるだけでなく、刺激語-連想語対の関連も推定する。

この連想データベース間では名詞間の関連は、動詞や形容詞を介して間接的に関連づける。そこで、isa、has、com、con、mat などの名詞間の直接的な関係は多次元シソーラスによって関連づける。

人間用のシソーラス [3] は基本用言 (V,ajc) から関

連する (N,ad,ajc,v) を検索できる構成となっており、一種の連想機能を備えていると見ることができる。

図4に動詞-名詞間の連想を行う連想型データベースの例を示す。

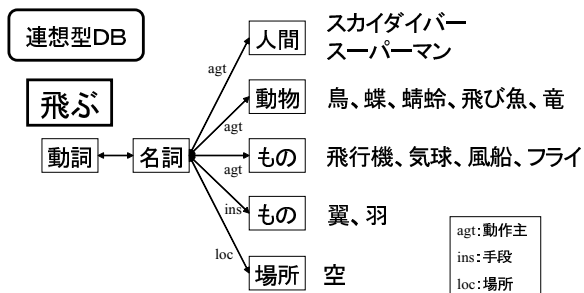


図4：連想型データベースの例

以下に連想型データベースを用いた連想例を示す。

図5に示す連想例では、「鳥」という刺激語から『翼で空を飛ぶ』と『羽ばたく』という動作を連想し、飛ぶ場所としての【空】と飛ぶ手段としての【翼】を推定している。

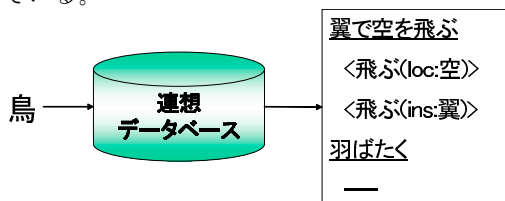


図5：連想型データベース 連想例1

また、図6に示す連想例では、「鳥」と「飛行機」という刺激語から『空を飛ぶ』という共通の動作/事象を連想し、飛ぶ場所としての【空】を推定している。



図6：連想型データベース 連想例2

## 4 連想型多次元シソーラス

連想型多次元シソーラスは、多次元シソーラスを連想型データベースを組み合わせる構成したシソーラスである。図7に連想型多次元シソーラスの具体例を示す。

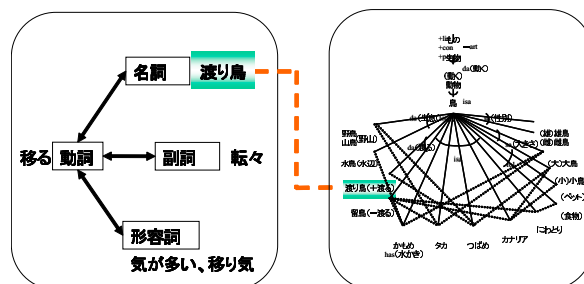


図7：連想型多次元シソーラスの具体例

本シソーラスは、2つのシソーラスの特徴を兼ね備えたことにより、

- 語と語の間の種々の関連を表現
- 単語の連想機能

だけでなく、

- 並列語や複合語の意味解析への応用
- 単語の語義記述への応用
- 各種の多義の絞りこみ

をも可能とする。これにより、連想型多次元シソーラスを用いることによって、従来の格パターンを用いた場合の意味解析の限界を突破することが可能となる。

## 5 意味解析への適応性

### 5.1 単語多義の絞り込み

次に示す2つの図は、格パターンと連想型多次元シソーラスを用いた場合の多義の絞り込みの過程についてそれぞれ示したものである。

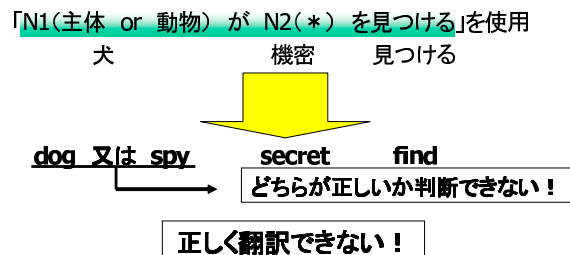


図8：多義の絞り込み (格パターン)

図8の結果が示すように、格パターンを用いた多義の絞り込みの場合、文中に出てくる「犬」が『dog』なのか『spy』なのかを解析することができず、多義が生じてしまう。

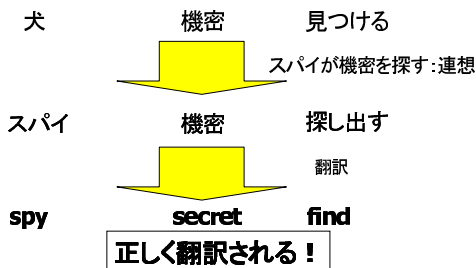


図9：多義の絞り込み (連想型多次元シソーラス)

図9が示すように連想型多次元シソーラスを用いた多義の絞り込みの場合、各々の刺激語・連想語から推定された要素を用いることによって、「犬」が『spy』であることを解析することができるようになる。これは従来の格パターンでは無視していた格要素同士の関連を連想型多次元シソーラスの単語連想機能によって関連づけることにより可能としている。

## 5.2 文の意味解釈の妥当性

「人間は文の統語構造を手掛かりに、文を構成する単語間の意味的結束性を時には前文、前々文などに出現した単語との間の意味的結束性を単語連想機能によって評価しながら、言語を理解している」という考えに基づき、以下に述べるような連想型多次元シソーラスを用いた新しい意味解析が可能となる。

図10は単文中の単語間の意味的結束性を評価し、これに基づき、単文の意味解釈の妥当性を検証した例である。

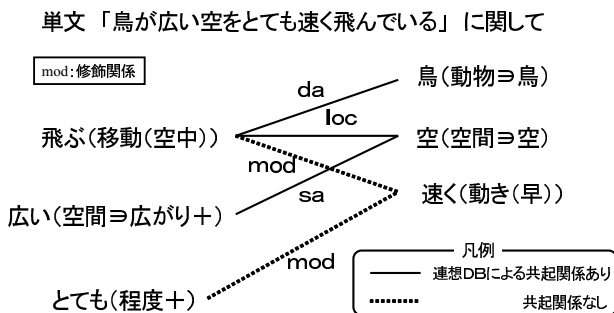


図10：単語間の意味的結束性の例

単文を統語解析した後、連想型多次元シソーラスを用い、係り受け関係のある単語間の共起関係を調べ、意味的結束性を以下のように評価する。

「鳥」と「飛ぶ」は動的属性 (da) による共起関係にあり、「飛ぶ」と「空」は場所 (loc) による共起関係にある。また、「空」と「広い」も静的属性 (sa) による共起関係にあることを示している。これらの単語は連想型多次元シソーラスの共起関係より、意味的結束性があることが確認される。

連想型多次元シソーラスによる共起関係がなくとも、「速く」は動きに関する属性であるので、「飛ぶ」と意味的結束性があることがわかる。「とても」は属性の程度を表す単語であるので、その属性「速く」と意味的結束性があることがわかる。

このように、連想型多次元シソーラスを用いて単語間の意味的結束性を評価することによって、単文の意味解釈の妥当性が検証できる。

## 6 あとがき

従来の格パターンとシソーラスを用いた意味解析の限界を突破し、より高度な意味解析を実現するために「連想型多次元シソーラス」を提案し、その意味解析への適応性について論じた。

現在、連想型多次元シソーラスの構築を進めているが、今後、連想型多次元シソーラスによる新しい意味解析法を実現すると共に同形語や同音語の判別、多義語の語義や統語的曖昧さの絞り込み、複合名詞/名詞句や並列語句の解析などへの適応性を検証していく予定である。

## 参考文献

- [1] 池原、宮崎、ほか6名：日本語語彙大系、岩波書店 (1997)
- [2] 川村、片桐、宮崎：語の種々の観点から分類した多次元シソーラス、電子情報通信学会技術報告、NLC94-48(1995)
- [3] 柴田武、山田進 編：類語大辞典、講談社 (2002)