

高度な意味処理向けの新しいシソーラスの提案 ー連想型多次元シソーラスー

宮崎 正弘

株式会社ラングテック

NPO 言語研究アソシエーション (ALR)

あらまし より高度な日本語意味処理を実現するため、多くの分類観点を持ち、単語連想機能を持った新しい日本語シソーラス(連想型多次元シソーラス)を提案し、その意味処理への適用性について論じる。連想型多次元シソーラスによって格パターンと上位-下位関係に基づく従来型シソーラスをベースにした意味処理の技術的行き詰まりを打破し、新しい意味処理が可能となることを論じる。

キーワード 日本語処理、意味処理、格パターン、シソーラス、単語連想機能

A Multi-Dimensional Thesaurus with Word Associations Function for More advanced Japanese semantic processing

Masahiro MIYAZAKI

LangueTech Corporation

NPO the Association for Language Research

Abstract Aiming at a breakthrough to overcome the limitations of Japanese semantic processing using existing case patterns and thesauri based on one semantic relation such as is-a relation, this paper proposes a new Japanese thesaurus, a multi-dimensional thesaurus with word associations function, and shows the validity of it for Japanese semantic processing. A multi-dimensional thesaurus with word associations function can represent various relations between words (eg. is-a, has-a, function, or other relations) and will make it possible to realize more advanced Japanese semantic processing using new case patterns described by recognition primitives which this paper proposes.

Keywords Japanese processing, semantic processing, case pattern, thesaurus, word associations function

1. はじめに

日本語意味処理において、格パターンと上位-下位関係に基づく従来型シソーラスをベースにした意味処理の限界が指摘されている。従来のシソーラスは単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、全体-部分関係、静的属性、動的属性、機能/役割/用途、原因/源泉など多くの観点が存在し、分類観点が不明確である。従来の格パターンでは、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切れない。

本稿では、より高度な意味処理を実現するため、多くの分類観点を持ち、単語連想機能を持った新しい日本語シソーラスである連想型多次元シソーラスを提案し、その意味処理への適用性について論じる。連想型多次元シソーラスによって

格パターンと上位-下位関係に基づく従来型シソーラスをベースにした意味処理の技術的行き詰まりを打破し、新しい意味処理が可能となることを論じる。

2. 新しい意味処理の探求

2. 1. 従来型意味処理の限界

従来、日本語処理では、言語表現の中心要素である動詞にどんな意味属性(概念)を持った名詞がどのような格要素となるかなどを記述した格パターン[1][2]と上位-下位関係を中心に単語や概念を体系的に分類したシソーラス[1][3]を用いることが多い。しかし、格パターンにおける名詞の選択制限は、ノード数の多いシソーラスでもその意味属性だけでは十分に記述しきれず、単語そのものを利用することもあるが、その網羅性は十分に保証されていない。また、「いる」と「ある」との格パターンの差は、一般に「が格にくる名詞の意味素性が有性であるか否か」とされるが、「彼には2人の子供が

ある」「～に反対する人がある」「家の前に車がいる」などこの原則に反する例はいくらでもある。さらに、「上がる」「登る」などのような類語の微妙な意味やニュアンスの差については、現状では名詞の選択制限として記述しきれないので、「山{に/を}上がる」より「山{に/を}登る」の方が自然な表現であることを決定するには意味処理によってではなく、用例や統計・確率的手法によらざるをえないのが現状である。「食べ物」「食料」のような名詞の類語でも「食べ物を食べる」と言えるが、「食料を食べる」とは言わないというような類語の使い分けは自然な日本語文を生成する場合には問題となる。メタファーやメトニミーなどの比喩表現は、名詞の選択制限に違反して扱えないか、名詞の選択制限は一応満たしているが、意味的に正しい解釈が得られない。例えば、「風呂があふれる」「黒板を消す」などは「風呂の水や湯があふれる」「黒板に書かれている文字や図などの情報を消す」が自然な解釈であり、従来の格パターンでは扱い切れない。しかも、このような表現は、日常の言語表現の中では比喩とは意識せずに使われることも多い。

文の意味解釈の妥当性の検証、類語の使い分けを含む統語的・意味的に正しい自然な日本文生成、文脈を利用した文の解析・生成など意味処理の基本的な課題は、今後の日本語処理の大きな課題となっている。構文・意味解析における統語的・意味的曖昧さの解消、複合名詞・名詞句や並列語句の統語的・意味的曖昧さの解消、多義語の意味的曖昧さの解消などにはまだ残された課題も多い。

2. 2. 新しい意味処理の基本的枠組み

2.1 で述べた格パターンとシソーラスをベースとした従来型意味処理の行き詰まりを打破するには、難問である意味の問題に真正面から取り組み、汎用性のある意味処理を探求する必要がある。新しい意味処理の探求はブレークスルーを探求する正攻法の正面突破作戦であり、統語処理に比べて言語学的知見やセンスが要求されると同時に多大の時間と労力を要する質の良い言語データの収集・分析、および体系化が必要である。

これに対して、難問である意味の問題に真正面から取り組み、これを迂回する取り組みが、統計・確率的手法を導入した用例ベースの処理である。新しい意味処理の探求は質を重視した取り組みであるのに対して、用例ベースの処理は量を重視した工学者向きの取り組みであり、近年のコンピュータの処理能力の著しい向上と人工知能分野における深層学習技術の進展によって世の中の趨勢となっている。しかし、単に用例だけを増やすだけでは解決できない本質的限界も指摘されている。まず何よりも質の良い大量の用例が必要であるが、このような用例をどのように収集するかという問題がある。また、文脈情報や様々な言語知識をどのように活用して様々な曖昧性を解消するかという困難な問題もある。

文脈や様々な言語知識を活用した高度な意味処理には、難問である意味の問題に真正面から取り組む新しい意味処理の探求が必要である。新しい意味処理の基本的枠組みは、以下

の三つの項目より構成される(図1参照)。

(1) 三浦の言語モデルと意味論

従来、コンピュータによる言語の意味処理の研究が様々行われているが、これらの研究の多くが言語の意味を明確に定義せずに行われている。言語の意味を明確に定義し、それに基づき言語理解のモデルを構築することによって、初めてコンピュータで言葉の意味をどのように扱ったらよいか明らかにできる。従来、言語の意味に関しては様々な議論がされてきた。意味論は言語学や言語哲学の大きな課題の一つとなっている。ここでは、意味は実体や機能ではなく、表現自体が持っている客観的關係であるという三浦つとむの關係意味論[4][5][6]をベースとなる意味論として採用する。三浦の意味論によれば、言語の意味は表現に結び付き固定された対象と話者の認識の關係であるが、対象は直接表現に結合されるのではなく、話者の目を通して得られた認識が表現に結合されるのであるから、より限定的に捉え、表現と認識の關係を意味と考える。コンピュータでは、關係はポインターで表現されるため、工学的には表現と話者の認識を対応づけるポインターが意味であり、このようなポインターを張ることが意味処理であると言える。

(2) 認知的プリミティブによる新しい格パターン

より高度な意味処理を実現することを目指して、日本語の基本的な動詞・形容詞の語義や用法を、主体が事物を捉える際に着目する様々な認知的プリミティブなどによって、コア概念(基本義から派生し認知の仕方に共通性がある概念)として形式的に記述した新しい格パターンを構築する。これにより、従来の格パターンでは十分に記述しきれない類語の動詞や形容詞の間の意味・用法の差や、意味属性や単語そのものでは記述しきれない格要素となる名詞の選択制限などを形式的に記述する。また、基本義やコア概念を、人間にとって直感的に理解しやすい、外界に存在して人間が五感で認知できる具体的な事物・属性として設定し、名詞の選択制限を記述する。ここで、このような選択制限を満たさない場合、比喩表現と見なして解釈できる条件、すなわち、選択制限を満たす事物・属性への何らかの見なしが行われる条件を記述する。これにより、基本義やコア概念に比喩表現を適用する一般的枠組みを与えることを可能とし、用例や従来の格パターン規則ではカバーしにくい、基本動詞・形容詞という高頻度語の低頻度用法(比喩表現など)をカバーするものとして期待できる。さらに、基本動詞・形容詞の種々の語義や用法をコア文型として集約・抽象化することにより、既存の格パターンに比べて語義数を大幅に減らすことができ、意味処理における多義の爆発の抑止や多義の解消処理の軽減なども期待できる。認知的プリミティブによる新しい格パターンの詳細については3章で述べる。

(3) 連想型多次元シソーラス

上位-下位關係を中心に単語や概念を体系的に分類した、従来のシソーラスの問題点を解決するものとして、多くの分類観点を持ち、分類観点を明確化した多次元シソーラスを構築

する。多次元シソーラスでは、意味属性(概念)を上位の意味属性(概念)と様々な分類観点、および認知的プリミティブを用いて形式的に記述する。これによって、(2)で述べた認知的プリミティブによる新しい格パターンを機能させることが可能となる。また、言語理解において単語連想機能は重要な役割を果たしていることに着目して、連想による知識の効率的検索を行うことができる、単語連想機能を持った連想型データベースを構築する。多次元シソーラスと連想型データベースを組み合わせて構成したものが連想型多次元シソーラスである。連想型多次元シソーラスの詳細は4章で述べる。

3. 認知的プリミティブによる新しい格パターン

3. 1. 多義の扱いとコア概念

一般に多くの基本語は多義語であり、中核となる基本的な語義(基本義)から様々な語義が派生して多義語を構成しているものと考えられている。田中のコア理論[7]によれば、「単語には、文脈に依存しない多義を包摂する抽象度の高い語義(コア)が存在していて、単語がある文脈や状況においてある文で実際に使われると、抽象的なコアがその場面に応じて具体化・特殊化し、種々の意味が生み出される」としている。また、国広は、語を多義のあり方に関して概念語と現象語に二分している。国広[8]によれば、「概念語が指すものは脳中の概念であり、五感でとらえることができない。一般に助詞・助動詞・接続詞などの文法的な語がここに属する。現象語はその指すものが外界に存在し、五感でとらえることの出来るものである。物、動き、出来事、性質などを指す語がここに属する。この種の語が指すものを“現象素”と呼ぶとすると、人間は言語で表現する以前にその現象素を感覚的・心理的に認知するのだと考えられる。その時、同一の現象素でも認知のしかたが異なればそれに応じて異なった意味(認知的多義)を生じるのだと考えられる。こうして生じた多義は、単一概念が文脈の影響で生じた多義とまったく性質が異なったものとして区別されなければならない。」としている。動詞「とる」の現象素は「把握、獲得、除去」である。

コア理論では、一つの単語には一つのコアが存在するとしているが、種々の語義を持つ多義語を一つの語義でカバーしようとする、語義の抽象度が高過ぎ、認知のしかたが異なればそれに応じて異なった意味を生じる過程をコンピュータ上に実現することは容易ではない。そこで、発想を転換し、原則として外界に存在して人間が五感で認知できる具体的事物・属性(国広の提唱する現象素をより具体化したもの)を、多義語のすべての語義の基礎となる基本義として設定する。次に、基本義から派生する様々な語義を導出するため、設定された基本義に認知のしかたが異なることによって生じる語義の違いに応じてコア概念を設定する。なお、人間の脳中に観念として存在して人間が五感で認知できない事物・属性については、「人間は、頭の中に観念としてしか存在しない抽象的な事物・属性について表現する場合、抽象的な事物・属性をそれらとなんらかの類似性や関連がある五感で認知できる

具体的な事物・属性と対応付ける一種のの見なしを行うことにより、人間にとって直感的に理解しやすい表現とすることができる。」との考えに基づき、可能な限り具体的な事物・属性の基本義のコア概念と対応付け、その基本義やコア概念の枠組みで解釈するようにする。このような考えは、「語に関する深い意味表現と意味に関する生成的な操作を考えることにより、従来多義としてとえられていた言語現象を単一の意味記述からの派生として説明しようとする」「生成語彙」[9]の考えと通じるところがある。

以下に、動詞「あふれる」の基本義とコア概念の設定例を示す。

- ・基本義：「液体が容器状の物に一杯となり、その外に出る」
- ・コア概念1(視点：容器状の物の外)「液体が容器状の物に入りきらず、その外に出る>>物理的移動」

用例1：水が水槽からあふれる。水槽があふれる。

「液体」や「容器状の物」へのなんらかの見なしがない。

用例2：砂がバケツからあふれる。

「砂」(粉体)が「液体」と見なされている。

- ・コア概念2(視点：容器状の物の内)「液体が容器状の物に一杯だ>>状態」

用例1：{水が水槽に/水槽が水で}あふれている。

「液体」や「容器状の物」へのなんらかの見なしがない。

用例2：会場に聴衆があふれる。顔に喜びがあふれる。

頭にアイデアがあふれる。港が活気にあふれる。

店が商品であふれる。彼は自信にあふれている。

「液体」や「容器状の物」へのなんらかの見なしがある。

3. 2. 認知的プリミティブ

単語の多義が認知のしかたの相違によって生じることに着目し、従来の格パターンでは十分に記述しきれない類語の動詞や形容詞の間の意味・用法の差や、意味属性や単語そのものでは記述しきれない格パターンにおける名詞の選択制限などを記述するものとして認知的プリミティブをとりあげる。

認知的プリミティブは、主体が事物を捉える際に着目する様々な観点である。我々は具体物を捉える際、その視覚的特徴である形状にまず着目する。また、それをどのように用いるかという観点にも着目する。特に、人間がある目的のために作った人工物ではその機能、自然物ではその用途に着目する。視覚的に捉えることができない抽象物については、それらとなんらかの類似性や関連のある五感で認知できる具体物に対応付ける見なしを行う。以下、形状に関する認知的プリミティブ[10][11]を中心に考察を試みる。

形状には、物理的形状と認知的形状がある。物理的形状とは、「球」「円」に代表される直接視覚的に捉えられる形状の情報であり、単純な形状から複雑な形状まで数限りなく存在する。複雑な物理的形状の情報は、単純な物理的形状に近似させることによって名詞の選択制限のための情報として利用できる。認知的形状とは、物理的形状に付随して存在する情報であり、形状を概念的に抽象化した情報である。例えば、上部に開口部がある入れ物のような物は認知的形状の属性と

しては、「穴がある」「中に物を入れる空間がある」などの情報が得られる。この例において、「穴」がたとえどのような形状をしていても、「穴がある」ことは変わらない。このように、物理的形状の情報を排除し抽象化することによって得られる情報が認知的形状である。「穴がある」は「どのように見えるか」といった捉え方(物の見え)であり、「中に物を入れる空間がある」は「どのように存在するか」といった捉え方(物のあり方)である。

認知的形状には、動詞概念が対応している。例えば、認知的形状「球状」には動詞概念「\$転がる」(単語の直前に付加される「\$」は概念であることを表す識別子)などが対応し、認知的形状「穴」には動詞概念「\$入れる」「\$出す」「\$保存する」などが対応する。ここで、「ボールが転がる」という文に対し、格パターンの名詞の選択制限として、名詞の字面「ボール」を直接記述することによって動詞「転がる」と名詞「ボール」を対応付けておくのではなく、名詞「ボール」に「球状」という認知的形状を与えておき、「ボールは球状であるから転がる」と推論し、格パターンにおける名詞の選択制限を満たしていると判断するようにする。「球状」「穴状」のような認知的形状を形状に関する認知的プリミティブと呼ぶ。形状に関する認知的プリミティブとして以下のものを設定した。○形状(物の見え)に着目した認知的プリミティブ

- 空間の次元：
0D；0次元(点)、1D；1次元(線)、
2D- α ；2次元(面){ α =top(上面)/bottom(底面)/side(側面)/end(端)/--}、3D；3次元(立体)<デフォルト値>
- 空間の領域：空間(1~3次元)が開いているか否かの表示
開；空いた空間(空など)
半開；ある方向に閉じ、他の方向に開いた空間
閉；閉じた空間<デフォルト値>
- 通過性：面の通過可能性を表示
通過可；「穴」「水面」などを通じて面を通過可能
+通過可；「ドア」など面の通過可能／通過不能を切り替え可能な部分を持つ
-通過可；「壁」など面を通過不能<デフォルト値>
- 障害：具体物の移動の障害となることを表示
障害(高さ)；「山」「塀」など高さによる移動の障害
障害(深さ)；「谷」「凹地」など深さによる移動の障害
障害(水)；「海」「川」「湖」など水による移動の障害
障害(心理)；「道路」「踏切」など心理的な移動の障害
- 傾斜の度合い：空間(1~3次元)の傾きを表示
水平／緩い斜め／斜め／急な斜め／垂直
- 曲がり：線の曲がりの度合い
曲がり／-曲がり、丸み／-丸み
- 凹凸の度合い：面の凹凸の度合い
凹凸／凹／凸／-凹凸
- 特徴的形状(単体)：
点状 | 粒状
線状 | 紐状、棒状、経路状

面状 |

带状；隙間状

円状；環状、楕円状、半円状

方形状；三角形、四角形、多角形 星形 布状

網状；ネットワーク状

領域状；

立体状 |

塊状；断片状

球状；卵状、楕円球体、半球状

ドーナツ状；リング状

円盤状；車輪状

管状；棒状

容器状；箱状、籠状、袋状、穴状、溝状

方体状；立方体、直方体、多面体、板状、本状、台状

柱状；三角柱状、四角柱状、円柱状

錐状；三角錐状 四角錐状 円錐状

衣服状；コート状、ワンピース状、着物状、シャツ状、ズボン状、スカート状

動物状；人状、獣状、海獣状、鳥状、トカゲ状、亀状

蛙状、魚状、海老状、蟹状、蛸状、イカ状、

クラゲ状、二枚貝状、巻貝状、蝶状、トンボ状、

昆虫状、蜘蛛状

植物状；樹木状、草状、苔状、茸状

不定形；気体、霧状体、液体、粘液体、粉体、粘体

・特徴的形状(集合体)：

列状(単体の形状)、束状(単体の形状)、群状(単体の形状)

○形状(物のあり方)に着目した認知的プリミティブ

・保持可能性：物を空間内部に入れておける空洞の有無

保持可(空間状態)；物を空間内部に入れておける空洞あり

-保持可；空洞なし<デフォルト値>

・空間状態：空間内部の状態(空虚／液体／固体／・・・)を表示

空虚；空間が空虚(特殊な場合として気体)<デフォルト値>

液体；空間が液体(特殊な場合として水)

固体；空間が固体

粉体；空間が粉体(砂、灰、粉など細かい粒状の固体の集合<連続体と見なせる>)

粘体；空間が粘体(飴、糊、ゼリー等固体と液体の中間状態)

粘液体；空間が粘液体(汁、ペンキ等粘り気がある液体)

霧状体；空間が霧状体(霧、雲、煙など細かい粒状の固体・液体の集合<気体とみなせる>)

・変形：空間自体の変形のしかたを表示

弾性体；弾力性のある物(バネ、ゴムなど)

軟体；簡単に変形する柔軟な物(紙、布、縄など)

硬体；簡単に変形しない硬い物(木材、金属、コンクリートなど)<デフォルト値>

形状に関する認知的プリミティブを用いた名詞概念の形状 sha についての形式的記述例を以下に示す。

- ・\$トンネル=\$交通路{sha(3D(管状, 保持可), 2D-end(通過可, 垂直, 半円状)*2}.

中に空洞のある管状の立体で、両端に通過可能な半円状の垂直面を持つ交通路(*2は2D-endが2つあることを表示)

- \$山=\$地形{sha(3D(障害(高さ), 鍾状), 2D-side(緩い斜め,斜め,急な斜め, 凹凸))}.
中に空洞のない錐状の立体(高さによる移動の障害となる)で、側面に凹凸のある(緩い～急な)斜面を持つ地形
- \$川=\$地形{sha(3D(障害(水), 溝状, 保持可(水)), 2D-top (通過可, 障害(水), 帯状, 水平))}.
中の空洞が水で満たされた溝状の立体(水による移動の障害となる)で、上面に通過可能な帯状の水平面を持つ地形
- \$茶碗=\$食器{sha(3D(容器状, 保持可), 2D-top(通過可, 水平, 円状))}.
中に空洞のある容器状の立体で、上面に通過可能な円状の水平面を持つ食器
- \$家=\$建造物{sha(3D(容器状, 保持可), 2D-side(+通過, 垂直, 方形状))}.
中に空洞のある容器状の立体で、側面に通過可能／通過不能を切り替え可能な(ドアや窓などのような)部分を含む方形状の垂直面を持つ建造物
- \$風船=\$玩具{sha(球状, 保持可(気体), 軟体).
中に気体のある球状で、簡単に変形する玩具

3. 3. 基本動詞のコア概念の記述

動詞「渡る」「越える」の基本的用法を以下のように表層格パターンで記述してみる。

- N1(\$人,\$動物,\$乗り物)/が N2(\$場所)/を 渡る
- N1(\$人,\$動物,\$乗り物)/が N2(\$場所)/を 越える

名詞 N2 の意味制約は「\$場所」となっているため、「場所」となりうる名詞すべてが N2 の意味制約を満たしてしまう。しかし、各パターンにおいて実際に N2 となりうる名詞は、それぞれ以下のようなものであろう。

- N2：川、海、谷、道路、踏切、橋、綱、世の中、・・・
- N2：山、堀、国境、・・・

「\$場所」は制約としては不完全で、何らかの制約を加えなければならない。「\$場所」より細かい意味属性を意味制約として記述したり、単語そのもので制約を記述したとしても、その網羅性を十分には保証しきれない。

そこで、3.2 で述べた認知的プリミティブによって意味制約を記述してみる。「渡る」は「二地点を隔てている移動の障害となるような場所を移動する」ことを表現する動詞である。「川、海」は「水による移動の障害」、「谷」は「深さによる移動の障害」、「道路、踏切」は「心理的な移動の障害」となるような二地点を隔てている場所である。「橋」は「水や深さによる移動の障害がある場所で移動の障害を回避するために設置されたものである」ため、「綱」は「元々、移動の障害となる場所である」ため、「世の中」は「世の中で苦労しながら暮らしていくことを、世の中という移動が困難な場所を移動する」ことにたとえた表現であるため「渡る」で表現されることが考えられる。「越える」は「高さによる移動の障害となるような場所を移動する」ことを表現する動詞である。「山、堀」

は「高さによる移動の障害」となるような場所であり、「国境」は「高さによる移動の障害を感じる堀のようなもの」であるため「越える」で表現されることが考えられる。

以上の考察に基づき、格パターン中の N2 の意味制約として「\$場所」に移動の障害となることを表示する認知的プリミティブによる制約を付加する。なお、「川、海、---」や「山、堀、---」などの名詞に移動の障害となる場所であることを表示する認知的プリミティブを設定しておく必要がある。

多義語に存在する種々の語義の一つ一つに対して個別に語義を記述するのではなく、すべての語義の基本となる基本義を設定すると共に、基本義から派生し認知の仕方に共通性がある概念をコア概念として設定し、その形式的記述を行う。動詞「あふれる」のコア概念の形式的記述例を以下に示す。

\$物理的移動(流出){(SUB N1)(LFR N2)(LTO N3)};

あふれる(溢れる){(N1/が N2/から (N3/に)) < (N2/が)};

N1{\$con{sha(液体)}}

<\$con{sha(粉体,粘体,粘液体,群状(立体状))}, ---(1a)

N2{\$con{sha(容器状, 保持可)}<\$spa}, ---(1b)

N3{外(N2)}. ---(1c)

\$状態(充满){(SUB N1)(LOC N2)};

あふれる(溢れる){(N1/が N2/に)<(N2/が N1/で/に)};

N1{\$con{sha(液体)}}

<\$con{sha(粉体,粘体,粘液体,群状(立体状))},

,\$phe,\$pes,\$men,\$inf},--- (2a)

N2{\$con{sha(容器状, 保持可)}<\$hum,\$anip,\$men,\$inf,\$spa}.

---(2b)

ここで、N1, N2, N3 は格要素となる名詞をパターン中の出現順に番号をつけて表したもの、SUB, LFR, LTO, LOC は、それぞれ、主体、場所・始点、場所・終点、場所を表す深層格ラベル([12]の日本語格ラベルを一部、修正・拡張)である。また、\$con, \$spa, \$phe, \$pes, \$inf, \$hum, \$anip, \$men は、それぞれ、概念、\$具体物、\$ところ、\$現象、\$知覚物、\$情報、\$人、\$動物の部分、\$精神を表す。動詞「あふれる」は、「液体が容器状の物に入りきらず外に流出する」(\$物理的移動)と「液体が容器状のものに充滿している」(\$状態)という二つのコア概念を持つ。\$物理的移動---;、\$状態---; は各コア概念の深層格パターンを記述したものである。あふれる(溢れる)---: は各コア概念の表層格パターンを記述したものであり、< の左側がコア文型、< の右側が派生文型を表している。: 以下の記述は、格要素となる名詞の認知的プリミティブを用いた選択制限を示したものであり、< の左側が標準的選択制限、< の右側が非標準的選択制限となっている。非標準的選択制限は標準的選択表現を満たさない場合に適用される選択制限である。

表層格パターン中の{(N1/が N2/から (N3/に)) < (N2/が)}において、(N3/に)は省略されることが多い要素であることを示している。また、派生文型(N2/が)を「N1/が N2/から」に文型変換可能であることを示している。なお、N1 の標準的選択制限が\$con{sha(液体)}であることから、N1 は液体であると推測でき、「バケツがあふれる」という文を「液体がバケツから

あふれる」と解釈できる。格要素となる名詞の選択制限(1a)において、標準的選択制限を満たさない場合、「粉体,粘体,粘液体,群状(立体状)」を「液体」とみなして N1 の選択制限のチェックをやり直すことを示している。これによって、「砂がバケツからあふれる」のような文では、「砂」(紛体)が「液体」と見なされて選択制限がチェックされる。「会場に聴衆があふれる」のような文では、以下の選択制限(2a)(2b)に基づき、「聴衆」(多くの人の集まりを群状(立体状)と見なす)が「液体」、「会場」(spa)が「容器状のもの」と見なされて選択制限がチェックされる。これによって、比喻表現を扱うことが可能となる。

4. 連想型多次元シソーラス

4. 1. 多次元シソーラス

(1) 従来型シソーラスの問題点

従来のシソーラスは、同義・類義関係に基づき、単語や概念を基本的に上位-下位関係(包含関係)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、全体-部分関係、静的属性、動的属性、機能/役割/用途、原因/源泉など多くの観点が存在し、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」「\$衣料」「\$道具」「\$乗り物」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、詳しくみると機能による分類と考えられる。

また、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ときわ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、分類観点が不明確となり、分類観点による語と語の関係の変化を扱えない。さらに、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)という観点で分類するか、art(人工の)という観点で分類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。

日本語シソーラスには、分類語彙表[13]、角川類語新辞典[14]、講談社・類語大辞典[15]、三省堂類語新辞典[16]、日本語語彙大系・意味属性体系[3]など階層的な木構造(上位ノードが一つ)で構成されたものが多いが、EDR 概念辞書[17]、日本語 WordNet[18]のように上位ノードが複数のネットワーク構造のものもある。シソーラスのノードにあたるノード名(意味属性名、分類項目名)は概念化の視点を表すのに適切な単語や補助的な記号を組み合わせて表現されている。通常、単語は複数の語義を持つので、意味属性を表す単語は原則として基本義とし、一語一義で用いている。しかし、本来曖昧性のある単語では意味属性名を十分に表現しきれない。自然言語処理用のシソーラスはコンピュータで扱えるような曖昧性のない明確な形式で意味属性名を記述し、シソーラスのノード間の関連をより明確にする必要がある。

(2) 分類の多観点化と明確化

従来型シソーラスの問題点を解消するため、分類の多観点

化と明確化を図った柔構造型のシソーラスが多次元シソーラス[19]である。多次元シソーラスの構成例を図2に示す。多次元シソーラスでは、名詞の分類観点として11のおおまかな分類観点(分類観点グループ)を設定し、各分類観点グループ内に詳細な分類観点を設けることで、語間の様々な関係を明確に表現できる。また、分類観点の順序性を極力排除するため、認知的プリミティブを素性とする集合で表す。例えば、「\$人工物」は con(具体)、per(認知可)、art(人工の)、-liv(生きている)という四つの素性を持つ「もの」として表現する。

多次元シソーラスの名詞の分類観点を以下に示す。

- isa 上位-下位：種類による分類を典型とする包摂関係など
例：\$鳥.isa=>\$燕,\$鳩,\$鷹,---,\$樹木.isa=>\$松,\$桜,\$梅,---
- ant 対/反意：対語や反意語関係
例：上.ant<=>下、売り.ant<=>買い
- hasa 全体-部分
pof 部分：場所の部分、抽象的領域の部分、情報の部分、組織の部分、期間の部分など
例：東京都.pof=>千代田区, ---, 三浦郡.pof=>葉山町、
\$競技場.pof=>\$グランド,\$スタンド,---,
\$化学.pof=>\$有機化学,\$無機化学,---,
\$文書.pof=>\$文,\$図,\$表,---,\$学部.pof=>\$学科,---,
\$部.pof=>\$課,---,\$季節.pof=>春, 夏, 秋, 冬
com 要素/成分：生物の器官・組織、自然物の要素・成分、人工物(衣服・料理・施設・道具・機器・乗り物・システム、薬品など)の構成要素・成分、組織の構成員など
例：\$植物.com=>\$茎,\$葉,\$根,---,\$分子.com=>\$原子、
太陽系.com=>太陽,水星,金星,地球,火星,---,
\$家.com=>\$屋根,\$壁,\$土台,\$部屋,---,
\$自転車.com=>\$ハンドル,\$タイヤ,\$サドル,---,
\$会社.com=>\$役員,\$社員,---
- cot 中身/内容：容器の中身、情報の内容など
例：\$グラス.cot=>\$飲み物,\$言葉.cot=>\$意味
- sa 静的属性
qua 数量・存在：事物の数量、多少・有無など
例：\$コンビ.qua=>2人,\$トリオ.qua=>3人
sha 形状：物や空間の形・構造など
例：\$グラス.sha=>容器状,\$飯.sha=>群状(粒状)
pro 性質：事物・空間・時間の性質など
例：\$空.pro(広さ)=>広い,\$海.pro(広さ)=>広い
sta 状態：事物・空間・時間の状態など
例：\$空.sta(色)=>青い,---,\$海.sta(色)=>青い,---
rel 関連：事物・空間・時間の関係・程度など
例：\$兄.rel=>親族関係,\$友人.rel=>対人関係
- da 動的属性：特徴的な動的属性など
例：\$鳥.da=>\$飛ぶ((LOC \$空圏)(INS \$翼))
\$水鳥.da=>\$生息する(LOC \$水辺)
- use 機能/役割/用途：人工物の機能、主体の役割、自然物の用途など
例：\$本.use=>\$読む((SUB \$人)(OBJ *))

\$教師.use=>\$教育する((OBJ \$生徒,\$学生)(LOC \$学校))
\$馬.use=>\$移動する((SUB \$人)(LOC \$陸圏)(INS *))

・sor 根源

cau 原因/源泉：物の源泉となる動作・行為・事象、事の原因となる事象など

例：\$本.cau=>\$書く((SUB \$人)(OBJ *)),
\$利子.cau=>\$預金する((SUB \$人)(REC \$金融機関)),
\$足音.cau=>\$歩く((SUB \$人)(LOC \$陸圏)),
\$到着.cau=>\$出発

mat 原材料：物の原料・材料など

例：\$土器.mat=>\$土、\$布.mat=>\$糸、\$パン.mat=>\$粉

・sp 空間：事物が生起・存在する空間上の位置・方向など

例：\$風.sp(方向)=>東,西,---、\$空.sp(位置)=>高い,---

・ti 時間：事物が生起・存在する時点・期間など

例：\$風.ti(季節)=>春,秋,---、\$空.ti(時刻)=>朝,夕,---

・evl 評価：「快/不快」の意味・感情を含む

例：\$暖かさ.evl=>快、\$暑さ.evl=>不快

・pha 位相：日常語/文章語(古語・雅語を含む)/口頭語(俗語・隠語・方言・男性語・女性語・若者語・幼児語・老人語を含む)/専門用語(分野)/固有語(地名,人名,組織名,施設名,---)

ここで、分類観点の記述例において、「\$教師.use=>\$教育する((OBJ \$生徒,\$学生)(LOC \$学校))」は「教師の役割 use は、”(教師が)生徒や学生を学校で教育する”ことである」を表し、分類観点 use の対象である「\$教師」を指し示す「*」によって、(SUB *)と陽に記述すべきところを省略している。「\$生徒,\$学生」は「\$生徒 or \$学生」を表し、「\$生徒,\$学生」は「\$生徒 and \$学生」を表す。OBJ,OBJ',LOC,REC,INS はそれぞれ、作用対象、作用結果、場所、受け手、手段・方法を表す深層格ラベルである。(SUB *)は必ず省略するが、(OBJ *)など SUB 以外の場合には省略しない。

(3) 意味属性(概念)の形式的記述

センサーのノード間の関連を明確にするため、意味属性名(分類項目名)はコンピュータで扱えるような曖昧性のない明確な形式で記述する必要がある。また、類語間の語義や用法の差異を弁別できるようにするため、類語の語義(概念)の形式的記述が必要となる。ここでは、意味属性(概念)を上位の意味属性(概念)と(2)で導入した多次元センサーの分類観点をを用いて形式的に記述することを考える。これは、意味属性(概念)が素性の束であるという考え方による。分類観点の属性値を 3.2 で導入した認知的プリミティブなどによって記述したものを素性とし、その素性の集合によって意味属性(概念)を定義する。その際、意味属性の下位の意味属性に共通の素性のみ記述する。ただし、下位の意味属性に共通な素性を見つけ出すのは困難で、いくつかの例外が存在することがある。例えば、「\$鳥」はほとんどは「空を飛ぶ」が、「\$ペンギン」や「\$ダチョウ」のように「空を飛ばない鳥」も存在する。このような場合、意味属性「\$鳥」には分類観点 da として「\$飛ぶ((LOC \$空圏)(INS \$翼))」と記述して、「\$鳥」の下位概念である「\$ペンギン」や「\$ダチョウ」には改めて「空

を飛ばない」ことを da として「-\$飛ぶ((LOC \$空圏)(INS \$翼))」のように記述する。意味属性(概念)の定義としては、下位の意味属性すべてに共通の素性ではなく、典型的な素性、いわゆるプロトタイプの素性を記述する。下位の意味属性は、上位の意味属性の定義が、そのまま継承されるため、定義を繰り返さないで、上位の意味属性を記述するだけでよい。

「鳥」や「飛行機」に関連する意味属性(概念)の形式的記述例を以下に示す。

\$生物=\$もの{da(liv), pro(con, per, -art)}.

\$動物=\$生物{da(\$動く,\$食べる(OBJ \$食物))}.

\$鳥=\$動物{da(\$飛ぶ((LOC \$空圏)(INS \$翼)), \$歩く((LOC \$陸圏)(INS \$足)), sha(鳥状), com(\$翼(NUM 2), \$羽毛,\$足(NUM 2), \$胴体,\$首,\$頭{com(\$嘴,\$目(NUM 2))})}.

\$具体物=\$もの{da(-liv), pro(con, per)}

\$人工物=\$具体物{pro(art)}.

\$乗り物=\$人工物{da(\$移動, use(\$移動((SUB \$人)(INS *)))}.

\$乗り物(空圏)=\$乗り物{da(\$移動(LOC \$空圏), use(\$移動((SUB \$人)(LOC \$空圏)(INS *)))}.

\$飛行機=\$乗り物(空圏){da(\$滑走する,\$走る((LOC \$陸圏)(INS \$車輪)), com(\$主翼(NUM 2), \$水平尾翼(NUM 2), \$垂直尾翼,\$車輪,\$エンジン,\$プロペラ?, \$胴体{com(\$操縦室,\$客室?, \$貨物室?))}}.

ここで、「?」は必須の要素でないことを表す。例えば、\$飛行機の形式的記述項目中の「\$プロペラ?」は「プロペラがない飛行機(ジェット機など)」もあることを示している。NUM は名詞と名詞の関係(数)を表す深層格ラベルで、「\$足(NUM 2)」は「2本の足」を表す。このようなものにはその他に QUN,UNI,POS,ADJ 等があり、それぞれ、量、単位、所有関係、形容詞類による連体修飾を表す。

4. 2. 連想型データベース

人間はあることを行う時に、それに伴って起こる環境の変化が行為に関連しているかという問題(フレーム問題)で迷わない。行為に関連しないことは意識せずに見捨てるからである。人間は連想(ある種の決めつけや思い込み)に基づく知識の効率的検索を行っているものと考えられる。膨大な知識をコンピュータ内に蓄積しても、必要な時に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。そのような問題を解決するものとして、連想に基づく知識の効率的検索を行うことができるものが連想型データベース[20]である。

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のセンサーでは見つけることができない単語間の新たな関係を見つけていることができるデータベースである。連想型データベースは、名詞類、動詞類、形容詞類、副詞類の間で連想機能を持っており、連想機能によって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。この連想型データベースでは、名詞間の関連は、動詞や形容詞を介して間接的に関連付ける。そこで、名詞間の直接的な関係は多次元センサーによって関連付ける。

連想型データベースを用いた連想例を図3、図4に示す。

図3に示す連想例では、「鳥」という刺激語から「翼で空を飛ぶ」や「羽ばたく」という動作を連想し、「飛ぶ」場所としての「空」や飛ぶ手段としての「翼」を推定している。また、図4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。

4. 3. 連想型多次元シソーラスとその効用

連想型多次元シソーラスは、多次元シソーラスと連想型データベースを組み合わせて構成したシソーラス[20]である。図5に連想型多次元シソーラスの構成を示す。本シソーラスは、分類観点の多観点化と明確化により語と語の間の種々の関連を表現するだけでなく、単語連想機能も持ったシソーラスである。

連想型多次元シソーラスによって、3章で述べた認知的プリミティブによる新しい格パターンを機能させることが可能となる。これによって、従来の格パターンとシソーラスを用いた意味処理では取り扱えなかった、3.3で詳しく述べた格パターンにおける名詞の選択制限のよりの確な記述やメタファー・メトニミーなどの比喩表現に対する統一的な扱いを可能となる。その他、以下に述べるような、より高度な意味処理の実現が可能となる。

(1) 文の意味解釈の妥当性検証

人間は文の統語的構造を手掛かりに、文を構成する単語間の意味的整合性を表す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考えられる。この考えに基づいて、連想型多次元シソーラスを用いた文の意味解釈の妥当性が検証できるようになる。図6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。単文「鳥が広い空をととても速く飛んでいる」を統語解析した後、連想型多次元シソーラスを用いて、係り受け関係に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」は空間 **sp** による共起関係にある。また、「空」と「広い」は静的属性 **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により、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ても、「速く」は動きに関する属性「速さ」の属性値であり、「とても」は属性の程度を表す副詞であるので、「速く」と意味的結束性があることが分る。

(2) 単語の多義の絞り込み

図7は、連想型多次元シソーラスを用いた単語の多義の絞り込みの過程について示したものである。従来の格パターン「N1(\$主体,\$動物)が N2()を 見つける」(「」は意味制約なしを表示)とシソーラスを用いた多義の絞り込み場合、「犬が機密を見つける」という文中に出てくる「犬」が「dog」なのか「spy」なのかを解析できず、多義が生じてしまう。これに対して、図7に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

定される要素を用いることによって、「犬」が「spy」であると決定することができる。これは、従来の格パターンでは無視していた格要素間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付けることによって可能としている。

(3) 類語の使い分け

自然な日本語文を生成する場合、類語の使い分けが問題となる。例えば、「食べ物」「食料」のような名詞の類語でも「食べ物を食べる」と言えるが、「食料を食べる」とは言わない。「食べ物」は「料理されていてすぐに食べられるもの」を表すのに対して、「食料」は「米、野菜、肉などの料理の材料(食材)を総称的に表す」ことに着目すると、概念としての「食べ物」「食料」は、以下のように形式的に記述される。
\$食べ物=\$人工物{use(\$食べる((SUB \$人)(OBJ *))), cau(\$料理する((SUB \$人)(OBJ \$食材)))?}.

\$食料=\$人工物{use(\$料理する((SUB \$人)(OBJ *)))?}.

「\$食べ物」は「食べるためのもの」であるのに対して、「\$食料」は「料理するためのもの」であるから、「食べ物を食べる」が、自然な表現だと分る。

(4) 格パターンにおける名詞の選択制限の反例への対応

日本語には、存在を表す基本動詞として「いる」と「ある」がある。一般に、「いる」は「が格」にくる名詞の意味素性が「有生」であり、「ある」は「が格」にくる名詞の意味素性が「-有生」であるとされている。しかし、以下のようにこの選択制限を満たさない例がいくらかでもある。「有生」でなくても「乗り物」「台風」などのように「自律的に移動すると見なせるもの」は、「家の前に車が{ある/いる}」「房総沖に台風が{ある/いる}」のように「ある」の他に「いる」も使える。この場合、両方とも「存在」を表すが、「ある」を使った場合に比べ「いる」を使った場合、「車」や「台風」に「生き物」のような動きを感じられる。逆に、「-有生」でなくても、「昔々、おじいさんとおばあさんが{い/あり}ました」「お降りの方は、{い/あり}ませんか」「原発建設に反対する人が{いる/ある}」「彼は最高権力の座に{いる/ある}」などのように「いる」の他に「ある」も使える。この場合、両方とも「存在」を表すが、「いる」を使った場合、には「{偶然/臨時的に/仮に}その場に存在する」という「動きを感じさせる」ニュアンスを表すのに対して、「ある」を使った場合には「{必然的に/不変的に}その場に存在する」という「動きを感じさせない」ニュアンスを表す。以上の分析に基づき、3.3で述べた認知的プリミティブによる新しい格パターン中に反例に適用するような非標準的選択制限として記述することによって対応することが可能となる。存在を表す基本動詞「いる」「ある」について記述した格パターンを以下に示す。

- ・\$存在(ある){(SUB N1)(LOC N2)};
ある(有る){N1/に N2/が};
N1{\$obj{da(-ani)}<\$shum,\$abs,\$sa,\$da},
N2{\$spa<\$shum,\$org}.
- ・\$存在(いる){(SUB N1)(LOC N2)};
いる(居る){N1/に N2/が};

N1{\$Obj{da(ani)}<\$Obj{da(ptr(-int))}},

N2{\$spa<\$hum,\$org}.

意味属性(概念)／ \$hum:\$人、\$org:\$組織、\$abs:\$抽象物、
\$obj:\$もの、\$da:\$こと、\$sa:\$さま、\$spa:\$ところ
認知的プリミティブ／ani:有生、 ptr(-int):物理的移動（意志なし）、＜以下は非標準的選択制限を表す。

5. おわりに

文脈や様々な言語知識を活用した高度な意味処理を実現するため、分類観点の多観点化と明確化により語と語の間の種々の関連を表現する多次元シソーラスと単語連想機能を持つ連想データベースを組み合わせて構成する連想型多次元シソーラスを提案し、その意味処理への適用性について論じた。

連想型多次元シソーラスによって、本稿で提案した認知的プリミティブによる新しい格パターンを機能させることを可能とし、これによって、格パターンにおける名詞の選択制限のよりの確な記述、名詞の選択制限の反例への対応、メタファー・メトニミーなどの比喩表現に対する統一的な扱いが可能となることを示した。また、文の意味解釈の妥当性の検証、単語の多義の絞り込み、類語の使い分けなど、日本文の解析や生成の基本的手法として有効であることにもふれた。

今後、多次元シソーラスの構築を手始めに、連想型データベース、認知的プリミティブによる新しい格パターンなどを順次構築し、より高度な日本語意味処理の実現を目指した研究を進めるとともに、多次元シソーラスや連想データベースなどを公開していく予定である。

文 献

- [1] 池原悟，宮崎正弘，白井諭，横尾昭男，中岩浩己，小倉健太郎，大山芳史，林良彦，”日本語語彙大系”，全5巻，岩波書店，Sep.1997.
- [2] 池原悟，宮崎正弘，横尾昭男，”日英機械翻訳のための意味解析用知識とその分解能”，情報処理学会論文誌，vol.34,no.8,pp.1692-1704, Feb.1993.
- [3] 宮崎正弘，”日英機械翻訳のための意味属性体系”，第5回国立国語研究所国際シンポジウム，第1専門部会「言語研究とシソーラス」，pp.163-171, Aug.1997.
- [4] 三浦つとむ，”日本語とはどういう言語か”，講談社学術文庫，Jun.1976.
- [5] 池原悟，”非線形モデルによる自然言語処理 基礎と応用”，岩波書店，Jun.2009.
- [6] 宮崎正弘，”言語過程説に基づく日本語解析の試み”，（佐良木昌・編「言語過程説の探求 第三巻 自然言語処理への展開」），明石書店，pp.21-107, Sep.2017.
- [7] 田中茂範(編著)，”基本動詞の意味論 コアとプロトタイプ”，三友出版，Oct.1987.
- [8] 国広哲弥，”理想の国語辞典”，大修館書店，Nov.1997.
- [9] Putejovsky，”The Generative Lexicon”，MIT press,1995.
- [10] 中田淳一，宮崎正弘，”認知的な形状-機能対を仮設した意義素に関する考察”，情報処理学会第49回全国大

会，3G-4, Sep.1994.

- [11] 中田淳一，宮崎正弘，”形状プリミティブによる格パターン辞書の拡張”，情報処理学会第51回全国大会，3H-10, Sep.1995.
- [12] 長尾真，”画像と言語の認識工学”，コロナ社，Jan.1989.
- [13] 国立国語研究所(編)，”分類語彙表 増補改訂版”，大日本図書，Jan.2004.
- [14] 大野晋，浜西正人，”角川類語新辞典”，角川書店，Jan.1981.
- [15] 柴田武，山田進(編)，”類語大辞典”，講談社，Nov.2002.
- [16] 中村明，芳賀綏，森田良行(編)，”三省堂類語新辞典”，三省堂，Nov.2005.
- [17] 荻野孝野，小林正博，”日本電子化辞書研究所における概念体系”，第5回国立国語研究所国際シンポジウム，第1専門部会「言語研究とシソーラス」，pp.172-181, Aug.1997.
- [18] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama and kyoko Kanzaki ”Development of Japanese WordNet”，LREC-2008, Marrakech, Jul.2008.
- [19] 川村和美，片桐康裕，宮崎正弘，”語を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, pp.33-40, Mar.1995.
- [20] 森田陽介，宮崎正弘，”連想型多次元シソーラスとその意味解析への適用性”，言語処理学会第12回年次大会発表論文集，pp.727-730, Mar.2006.

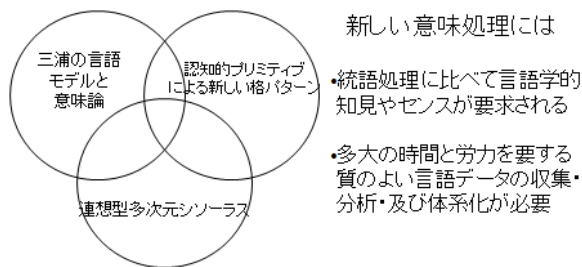


図1. 新しい意味処理の基本的枠組み

単語連想機能をもつ
連想型データベース

多くの分類観点をもった
名詞の多次元シソーラス

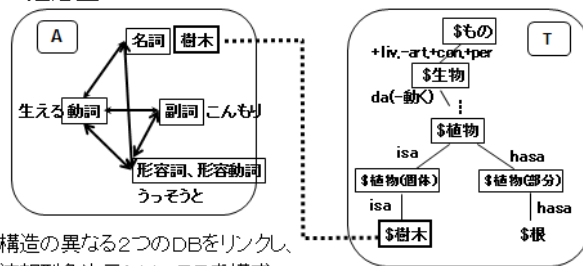
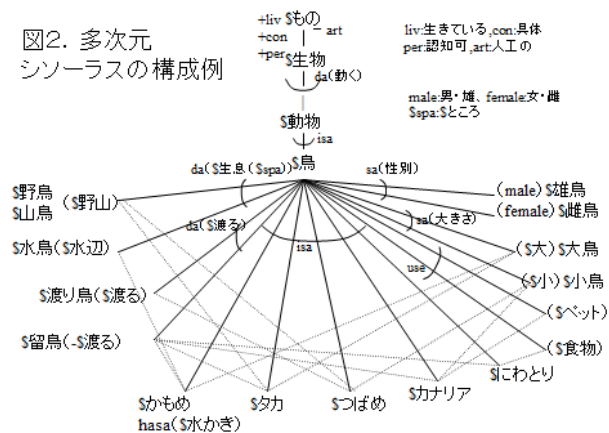


図5. 連想型多次元シソーラスの構成

図2. 多次元
シソーラスの構成例



単文「鳥が広い空をととても速く飛んでいる」

(mod 修飾関係)

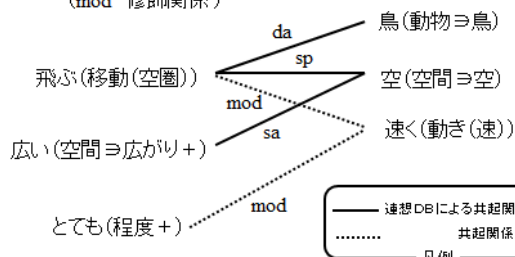


図6. 単語間の意味的結束性の例

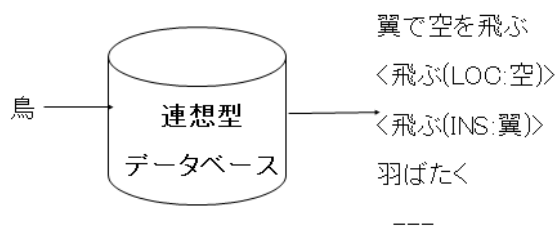


図3. 連想型データベースによる連想例1

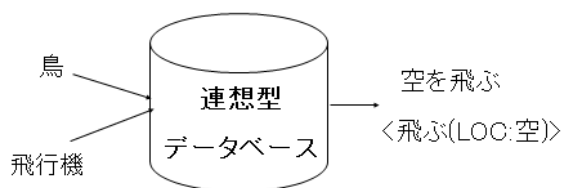


図4. 連想型データベースによる連想例2

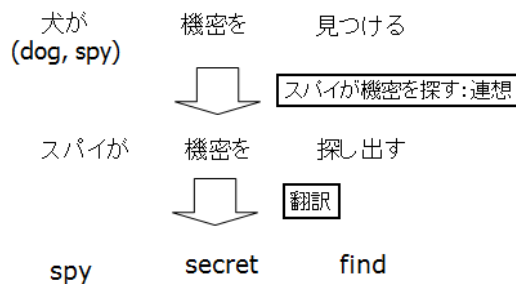


図7. 連想型多次元シソーラスによる
単語の多義の絞り込み例