

SD 式意味モデルにおける自然言語概念の体系化と 認識・理解・解釈動作の定式化

吉原 将太・脇山 正博・河口 英二

The Concept Hierarchy of Natural Language and a
Formalization of Recognition, Understanding and
Interpretation Process in the SD-Form Semantics Model

Shota Yoshihara, Masahiro Wakiyama and Eiji Kawaguchi

Abstract

The SD-Form Semantics Model, developed by the authors, is a framework to deal with the meaning of natural language in a quantitative way. It is associated with a formal language named SD-Form that describes the semantic structure of each language expression.

In the SD-Form semantics model, knowledge data in each world should be organized in a hierarchy. Such hierarchy is a model for compact storage of knowledge data. We give a formal definition of a concept hierarchy according to the elaboration relations among concepts. Formalization of intelligent operations such as recognition, understanding, interpretation, etc., are proposed regarding this concept hierarchy.

The objective of the present paper is to show the algorithm to construct a concept hierarchy as well as to show our formalization of intelligent operations such as recognition, understanding and interpretation. We implemented an experiment system (SDENV-3) by Prolog. The system operation is shown in brief by illustrating an example problem.

1. はじめに

自然言語 (natural language) とは, 相互の意思疎通を行う手段として, 人類の誕生とともに自然発生的に生まれ, 人類の進化とともに発展してきた言語のことをいう^[1]。自然言語は, 意思の疎通だけでなく, 理解・思考・発想・記憶・表現など人間の知的活動すべてに密接に関わっている。このような人間の知的活動についての諸問題を解明し, その結果を実用システムに役立たせることが自然言語処理研究の目指すところである。

自然言語を理解するためには, 言語内知識として, 語彙に関する知識や文法に関する知識が必要である。言語外知識としては, 対象となっている世界に関する知識や常識も必要である。人間の持つ知識の形態は様々であり, 脳の中ではどのような構造になっているのかは十

分には解明されていない。しかし、知識を計算機で利用するためには、知識を形式化して記述しなければならない。知識は一般に分類学的階層構造をしている。さらに、各要素は他の階層構造内の要素ともいろいろな関係を持つため、全体としてネットワーク（概念体系）で表現する必要が生じる。このような概念は、事実だけでなく判断規則にも存在する^[2]。これまでに提案されてきた概念の体系化の方法には、シソーラス^[3]やオントロジーなどがある。これらは、語をその概念で分類し、同意語や関係のある語をまとめて並べ、概念体系としてまとめたものである。これまでの研究では、概念体系を構築する際に用いる概念として、主に単語だけが使われており、文を含めたものは少ない。

近年、著者等の研究グループは自然言語概念の意味表現形式の一つとして SD 式 (Semantic-structure Description Form) を提案している^[4,5,6,7]。SD 式は、自然言語における個々の概念、陳述表現、感情表現、或いはシステムに与える知識データ等を記述するための一種の中間言語であり、その構文は、曖昧さの無い文脈自由文法で規定されている^[7]。自然言語概念を SD 式として捉え、その記述データを基にして意味処理を行おうとするモデルを SD 式意味モデルと言う。SD 式意味モデルの最大の特徴は、従来からの意味記述モデルでは扱い難かった 2 つの概念間の意味的な近さを定量的に扱える点である。そのため、概念上の二項関係として「詳述関係 (elaboration relation)」を定義し、これに基づく意味差を導入している。詳述関係は、SD 式の構文構造に基づくものと知識データに基づくものとを定義している。ここでいう詳述関係とは、2 つの概念 (SD 式) D_i , D_j に関して、 D_i の意味をより具体化したものが D_j であり、逆に D_j の意味をより抽象化したものが D_i であるような関係のことである。

詳述関係は半順序関係である。従って、この詳述関係を用いてシステム内の全ての知識データを体系化することが可能である。SD 式意味モデルでは、システムの知識データは初めから完成されているのではなく、外部から与えられる知識データをシステム自身が逐一内部に取り込み、新たな知識を増やしながら自ら知識の体系化を進める過程を想定している。

本論では、SD 式意味モデルで定義している詳述関係を用いた概念の体系化アルゴリズムについて、また、体系化された概念を利用した認識・理解・解釈という知的処理について述べる。

2 章では、SD 式意味モデルの概要を述べる。3 章では、概念の体系化について、4 章では、認識・理解・解釈動作について定義し、SD 式意味モデル実験システム (SDENV-3) による動作例を示す。最後に、5 章でまとめと今後の課題について述べる。

2. SD 式意味モデルの概要

2.1 SD 式による意味構造の記述例

SD 式は、自然言語における概念の意味構造を記述するための一つの形式言語であり、「SD 式記号」の列である。この記号は(1)概念ラベル、(2)修飾子、(3)規定子、(4)結合子などに類別

される。SD式のそれぞれの構文は、SDG (SD-Form Generative Grammar) と名付けた文法の生成規則^[7]によって定義されている。SD式の例を以下に示す。

〈Ex. 2-1〉

- 靴/新しい : 新しい靴
 [s(美紀), v(である), c(学生/長崎純心大学)] : 美紀は長崎純心大学の学生である。
 [s(相手), v(読む/mood/義務), o(本/当該)] : あなたは、その本を読まなければならない。
 [e(感嘆/賞賛)] : すばらしい!

2.2 知識の記述例^[8]

本論では、知識データを内蔵し自然言語概念を処理するシステムを知識ベース・システムと呼ぶ。知識ベース・システムは、対象世界を記述する「世界データ (World Data)」とその世界に関する「知識データ (Knowledge Data)」を持つものとする。これらは何れも対象世界の事実データ (Fact Data) である。図1にSD式意味モデルにおける事実データの階層を示す。知識データにおける「事実知識 (Fact Knowledge)」とは、対象世界の中で常に真となる個別事実データである。事実知識は、更に「一般事実 (General Fact)」と「個々の事実 (Individual Fact)」に分類される。また、「規則知識 (Rule Knowledge)」とは、その世界での推論に根拠を与える規則データである。規則知識は、更に「一般規則 (General Rule)」と「個別規則 (Specific Rule)」に分類される。「導出概念 (Induced Concept)」とは、システムに与えられた事実知識と規則知識により導出された概念のことである。

知識データは、規定子形式、結合子形式、陳述形式のいずれかの形式で記述する。知識データの記述例を以下に示す。

〈Ex. 2-2〉

[事実知識]

- (爬虫類) *incl*(ワニ) : 爬虫類はワニを含む
 (中学生) *kdof*(学生) : 中学生は学生の一つである
 [s(地球), v(自転)] : 地球は自転する

[規則知識]

- (*assu*([s(X), v(買う), o(Y)])) *caus*([s(X), v(払う), o(代金/Y)])
 : もしXがYを買うならば、XはYの代金を払う

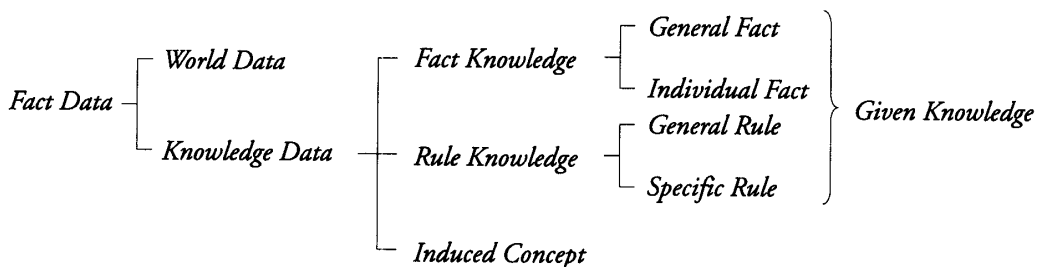


図1 事実データの階層

2.3 SD 式の意味的情報量

SD 式では、記号列の構造で何か固有の概念を表現しようとするだけでなく、その概念の意味量の大小も表すこととしている。そのため、SD 式意味モデルでは、各 SD 式記号に意味素量を設定している。任意の SD 式を D とするとき、その意味量を

$$si(D) = n \text{ [semit]}$$

と表す。また、その単位を *semit* としている。SD 式意味モデル実験システム SDENV-3 の場合は、意味素量を以下のように設定している。

- | | | |
|------------|-------------------------|--------------|
| (1) 変数ラベル | “X, Y, Z, ...” | : 1 [semit] |
| (2) 単純ラベル | “馬, CAT, ...” | : 10 [semit] |
| (3) 修飾子 | “/” | : 1 [semit] |
| (4) 規定子 | “nega, only, assu, ...” | : 2 [semit] |
| (5) 結合子 | “plus, incl, para, ...” | : 1 [semit] |
| (6) 機能項目記号 | “s, v, c, o, ...” | : 1 [semit] |
| (7) 区切り記号 | “[]” | : 1 [semit] |
| (8) 区切り記号 | “()”, “,” | : 0 [semit] |

SD 式全体の意味量は、それらの総和となる。

SD 式意味モデルでは、これらを定義するときの一般的指針は示しているが、意味量の値については、モデルの利用者が独自に定めて良いとしている。

〈Ex. 2-3〉

$$si(\text{靴/新しい}) = 21$$

$$si([s(\text{美紀}), v(\text{である}), c(\text{学生/長崎純心大学})]) = 45$$

$$si([s(\text{相手}), v(\text{読む/mood/義務}), o(\text{本/当該})]) = 67$$

$$si([e(\text{感嘆/賞賛})]) = 23$$

2.4 2 概念間の詳述関係

2 つの概念間の「詳述関係 (elaboration relation)」は、SD 式意味モデルにおける最も基本的な枠組みである^[6]。2 つの概念 D_i, D_j に関して、 D_i の意味をより具体化したものが D_j であり、逆に、 D_j の意味をより抽象化したものが D_i であるとき、 D_i と D_j には詳述関係があるという。このときの D_i を D_j の先祖、 D_j を D_i の子孫と呼んでいる。 D_i から D_j への詳述関係を

$$elab(D_i, D_j) = n \quad \text{または} \quad elab(D_i, D_j, n)$$

と表す。ただし、 n ($0 \leq n < \infty$) は詳述量で、詳述の詳しさの程度を表すものである。

SD 式意味モデルにおける詳述関係には、 D_i, D_j そのものの構造による「構文的詳述関係」とシステムが利用できる知識データに基づく「知識に基づく詳述関係」の 2 種類があり、それぞれ、

$$elab_{synt}(D_i, D_j) = n, \quad elab_{know}(D_i, D_j) = n$$

と表す。一般的には、

$$elab(D_i, D_j) = \min\{elab_{synt}(D_i, D_j), elab_{know}(D_i, D_j)\} = n$$

と定義している。ここで、上式で定義される詳述関係は1段の関係である。そして、この関係が連結して、多段の詳述関係となることもある。詳述関係の例を以下に示す。

〈Ex. 2-4〉

[構文的詳述関係]

$$elab_{synt}(\text{靴}, \text{靴/新しい}) = si(\text{靴/新しい}) - si(\text{靴}) = 21 - 10 = 11$$

$$elab_{synt}([s(\text{美紀}), v(\text{である}), c(\text{学生})], [s(\text{美紀}), v(\text{である}), c(\text{学生/長崎純心大学})]) = 11$$

[知識に基づく詳述関係]

2つのSD式を特別な関係で結びつける知識データがシステムに存在するとき、両者の間には詳述関係が生じる。

次の知識データがシステムに与えられているとする。

(靴) *incl*(サンダル) : 靴はサンダルを含む

(*assu*([*s*(X), *v*(である), *c*(先生/英語)]) *caus*([*s*(X), *v*(話す/上手), *o*(英語)])

: もしXが英語の先生ならば、Xは英語を上手に話す。

このとき、次の知識に基づく詳述関係が成り立つ。

$$elab_{know}(\text{靴}, \text{サンダル}) = 3$$

$$elab_{know}([s(\text{恵子}), v(\text{話す/上手}), o(\text{英語})], [s(\text{恵子}), v(\text{である}), c(\text{先生/英語})]) = 2$$

2.5 spec 関係

SD式意味モデルでは、詳述関係を拡張し、厳密な因果関係だけでなく、自然言語に現れるやや曖昧な因果関係を包含する *spec* 関係 (specialization relation) を導入している^[9]。*spec* 関係に基づいた意味の詳しさの程度を *spec* 量 (単位: *semit*) と呼んでいる。ただし、詳述関係と *spec* 関係を厳密に区別しない場合には、*spec* 関係も詳述関係と呼ぶ場合もある。同様に、*spec* 量についても詳述量と呼ぶ場合がある。

spec 関係の定義は、以下の通りである。

(1) $spec_{synt}(D_i, D_j) = elab_{synt}(D_i, D_j)$

(2) $spec_{know}(D_i, D_j) = elab_{know}(D_i, D_j)$

(3) (*assu*(D_j)) *indu*(D_i) が true のとき、 $spec_{know}(D_i, D_j) = 3$

(4) $elab_{know}(D_i, D_j) = 2$ または、 $elab_{know}(D_i, D_k) = 3$ が成り立つとき

$$spec_{know}(D_j, D_i/MOST) = 4$$

$$spec_{know}(D_k, D_i/MOST) = 4$$

$$spec_{know}(D_i/D, D_i/MOST) = 4$$

spec 関係の例を以下に示す。

〈Ex. 2-5〉

次の知識データがシステムに与えられているとする。

(アメリカ合衆国) *incl*(テキサス州) : アメリカ合衆国はテキサス州を含む

$(assu([s(X), v(\text{行く/場所/終点/海})]))\text{indu}([s(X), v(\text{泳ぐ})])$

：もし X が海に行くならば、 X は泳ぐだろう。

このとき、次の *spec* 関係が成り立つ。

$spec(\text{アメリカ合衆国, テキサス州}) = elab(\text{アメリカ合衆国, テキサス州}) = 3$

$spec([s(\text{三郎}), v(\text{泳ぐ})], [s(\text{三郎}), v(\text{行く/場所/終点/海})]) = 3$

2.6 最近共通先祖と概念間の意味差

与えられた 2 つの概念 D_i, D_j について、

$spec(D, D_i), spec(D, D_j)$

が同時に成立するとき、そのような D を D_i と D_j の「共通先祖」と呼ぶ。 D_i や D_j が有限の記号列であること、知識データの個数が有限であることから共通先祖も有限個であることは明らかである。また、「何か」という意味で用いる変数ラベル“ X ”は全ての概念の先祖となるので、少なくとも 1 個の共通先祖は存在する。特に、

$spec(D_0, D_i) + spec(D_0, D_j) = \min_0 \{spec(D, D_i) + spec(D, D_j)\}$

であるような D_0 を D_i と D_j の「最近共通先祖」と定義している^[10]。この関係を

$ncoa(D_i, D_0, D_j, n_i, n_0, n_j)$ 、または、 $ncoa(D_i, D_0, D_j) = n_0$

と表す。また、このときの n_0 を D_i と D_j の「意味差」といい、

$diff(D_i, D_j) = n_0$

と表す。すなわち、意味差は、与えられた D_i, D_j について最近共通先祖を探索することにより求められる。

最近共通先祖の導出及び意味差の例を以下に示す。

<Ex. 2-6>

次の概念 D_1, D_2 の最近共通先祖 D_0 、及び意味差 n_0 を求める。

$D_1 = [s(\text{明子}), v(\text{育てる/手段/粉ミルク}), o(\text{健太})]$: 明子は健太を粉ミルクで育てる。

$D_2 = [s(\text{智子}), v(\text{育てる/手段/母乳}), o(\text{裕二})]$: 智子は裕二を母乳で育てる。

ここでは次の知識データがシステムに与えられているものとする。

$F_1 = (\text{女性})\text{incl}(\text{明子})$: 女性は明子を含む

$F_2 = (\text{女性})\text{incl}(\text{智子})$: 女性は智子を含む

$F_3 = (\text{乳児})\text{incl}(\text{健太})$: 乳児は健太を含む

$F_4 = (\text{乳児})\text{incl}(\text{裕二})$: 乳児は裕二を含む

$F_5 = (\text{飲物})\text{incl}(\text{粉ミルク})$: 飲物は粉ミルクを含む

$F_6 = (\text{飲物})\text{incl}(\text{母乳})$: 飲物は母乳を含む

動作を分かり易く示すために、システムは $F_1 \sim F_4$ 以外の知識データを持たないものとする。これらの知識データより、概念 D_1, D_2 の最近共通先祖 D_0 は、

$D_0 = [s(\text{女性/SOME}), v(\text{育てる/手段/飲物/SOME}), o(\text{乳児/SOME})]$

：ある女性は、ある乳児をある飲物で育てる。

となる。D₀導出における詳述関係を図2に示す。図2における数字は詳述量である。概念D₁とD₂の意味差 n₀は、図2における詳述量の総和で、

$$\begin{aligned} n_0 &= \text{diff}(D_1, D_2) \\ &= \text{elab}(D_0, D_1) + \text{elab}(D_0, D_2) \\ &= (1+1+1) + (1+1+1) \\ &= 6 \end{aligned}$$

となる。

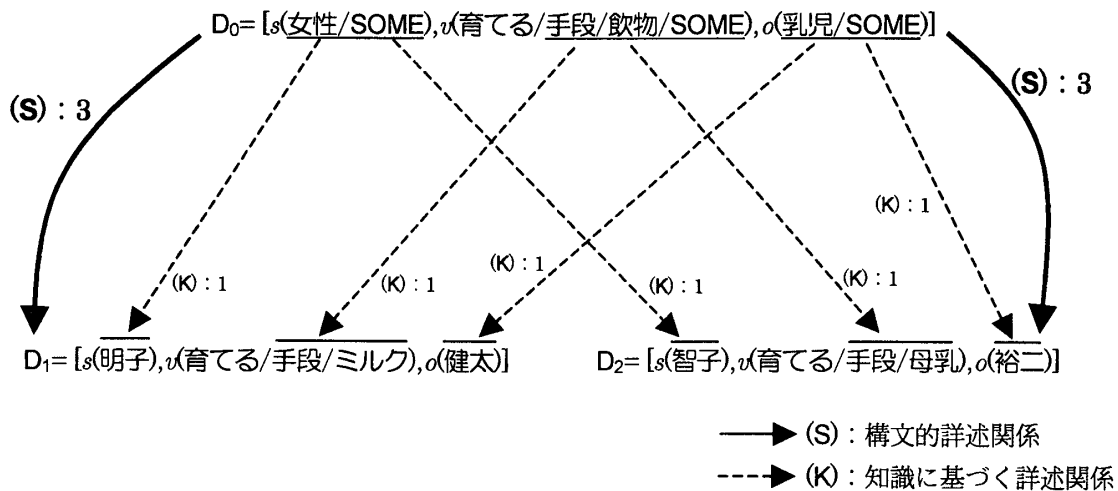


図2 D₁、D₂の最近共通先祖D₀の導出における詳述関係

3. SD 式意味モデルにおける概念体系

知識システム内の知識データは、一つ一つの概念として孤立して蓄えるよりも、全体を有機的に関連付けて蓄積しておく方が効率的である。詳述関係および spec 関係は半順序関係であり、この関係を用いてシステム内の全ての知識データを体系化することが可能である。

3.1 概念の体系化

概念体系 $\{\Omega/\Omega_0, \Psi\}$ とは、概念集合 Ω と詳述量の集合 Ψ からなる意味ネットワークである。ここで概念集合 Ω は、変数 X と与えられた事実データの集合 Ω_0 を含むものである。概念集合 Ω のそれぞれの概念を節 (node) とし、詳述量を枝 (branch) とすることで概念の体系を表すことができる。体系化された概念の集合とは、以下の条件を満たすものである。

- (1) Ω_0 は Ω の部分集合である。
- (2) Ω_0 における任意の二つの節点 D_i と D_j に対して最近共通先祖 D_0 が存在し、 D_0 は Ω に属する。
- (3) Ψ は一段の詳述関係からなる。 Ω において、全ての節点間の詳述量は、 Ψ の詳述量を累積することによって与えられる。

(4) 差集合 $\Omega - \Omega_0$ に属する任意の節点を除けば、(2) または (3) が満たされなくなる。

(5) Ψ に属する任意の枝を除くと (3) が満たされなくなる。

以上のような概念体系を用いると知識システムにおける認識動作、解釈動作、理解動作の形式化が容易となる (4 章参照)。

概念体系における「総事実記憶量(TM)」を個々の知識データの意味量の総和と定義する。また、「実知識量(NK)」を、知識データを体系化した際の概念間の詳述量の総和と定義する。これらの尺度は、知識システムにおける知識データの記憶効率を定量的に評価するための指標となる。

3.2 概念の体系化のアルゴリズム

以下に概念の体系化アルゴリズムを示す。

(1) 与えられた事実データ $\Omega_0 = \{D_1, D_2, \dots, D_i, \dots, D_k\}$ に対して、まず、 $\Omega = \Omega_0$ と初期設定する。そして、 D_1 と D_2 , D_1 と D_3 , \dots , D_i と D_k のように全ての組み合わせに対して、1 段の詳述関係による最近共通先祖 D_{ik} をそれぞれ求める。

$$\begin{aligned} & ncoa(D_1, D_{12}, D_2, n_{121}, _, n_{122}), ncoa(D_1, D_{13}, D_3, n_{131}, _, n_{132}), \dots, \\ & ncoa(D_i, D_{ik}, D_k, n_{ik1}, _, n_{ik2}), \dots \end{aligned} \quad (4-1)$$

(2) (1) で求めた最近共通先祖 D_{ik} が X (変数) または Ω_0 の概念と同一で無いならば、 D_{ik} を Ω_0 に加える。更に

$$elab(D_{ik}, D_i, n_{ik1}), elab(D_{ik}, D_k, n_{ik2}) \quad (4-2)$$

のデータを蓄積しておく。

(3) (2) の処理において Ω に加える最近共通先祖 D_{ik} が無くなるまで (1), (2) を繰り返す。

(4) X (変数) と D_i ($i=1, 2, \dots, k$) との詳述関係をそれぞれ求め、データを蓄積する。

$$elab(X, D_i, n_i) \quad (i=1, 2, \dots, k) \quad (4-3)$$

(5) (4-2) 式及び (4-3) 式において

$$elab(D_q, D_r, n_{qr}) \quad (4-4)$$

$$elab(D_r, D_s, n_{rs}) \quad (4-5)$$

$$elab(D_q, D_s, n_{qs}) \quad (4-6)$$

$$(q, r, s=1, 2, \dots, k)$$

の 3 つの関係が存在し、

$$n_{qs} \geq n_{qr} + n_{rs} \quad (4-7)$$

が成り立つならば (4-6) 式のデータを削除する。

(4-7) 式の条件を満たすものが存在する限りこの処理を繰り返す。この例を図 3 に示す。

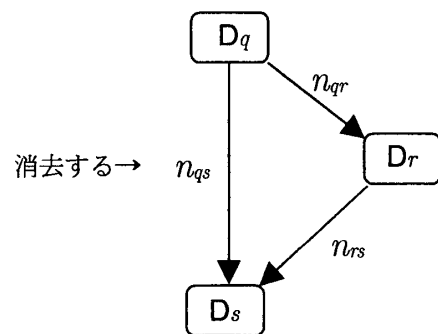


図 3 $n_{qs} \geq n_{qr} + n_{rs}$ の状態

(6) 最終的に残った蓄積データ $elab(D_x, D_y, n_{xy})$ について、 D_x, D_y を節とし、 n_{xy} を枝とした順序関係を作り出す。

3.3 概念の体系化の例

SD 式意味モデルが有効であるか否かは、実際の SD 式データを用いて具体的な処理を試す必要がある。そこで、著者等は SD 式意味モデルの実験システムとして SDENV-3(SD-form ENVironment system version 3)を試作中であり、この SDENV-3に概念の体系化の処理を組み込み、体系化アルゴリズムを確かめた。その動作例を以下に示す。

〈Ex. 3-1〉

次の知識データをシステムに与えて実験した。但し、GF は一般事実(*General Fact*), IF は個々の事実(*Individual Fact*), GR は一般規則(*General Rule*)である(2.2節参照)。

- GF₁ = (女性) *incl* (明子) : 女性は明子を含む
- GF₂ = (女性) *incl* (智子) : 女性は智子を含む
- GF₃ = (乳児) *incl* (健太) : 乳児は健太を含む
- GF₄ = (乳児) *incl* (裕二) : 乳児は裕二を含む
- GF₅ = (飲物) *incl* (粉ミルク) : 飲物は粉ミルクを含む
- GF₆ = (飲物) *incl* (母乳) : 飲物は母乳を含む
- IF₁ = [*s*(明子), *v*(育てる/手段/粉ミルク), *o*(健太)] : 明子は健太を粉ミルクで育てる。
- IF₂ = [*s*(智子), *v*(育てる/手段/母乳), *o*(裕二)] : 智子は裕二を母乳で育てる。
- GR₁ = (*assu* [*s*(X), *v*(産む/過去), *o*(Y)]) *caus* [*s*(X), *v*(育てる/手段/母乳), *o*(Y)]
: もし X が Y を産んだならば, X は Y を母乳で育てる。
- GR₂ = (*assu* [*s*(X), *v*(である), *c*(親/Y)]) *indu* [*s*(X), *v*(育てる/手段/_), *o*(Y)]
: もし X が Y の親ならば, X は Y を何かで育てるだろう。

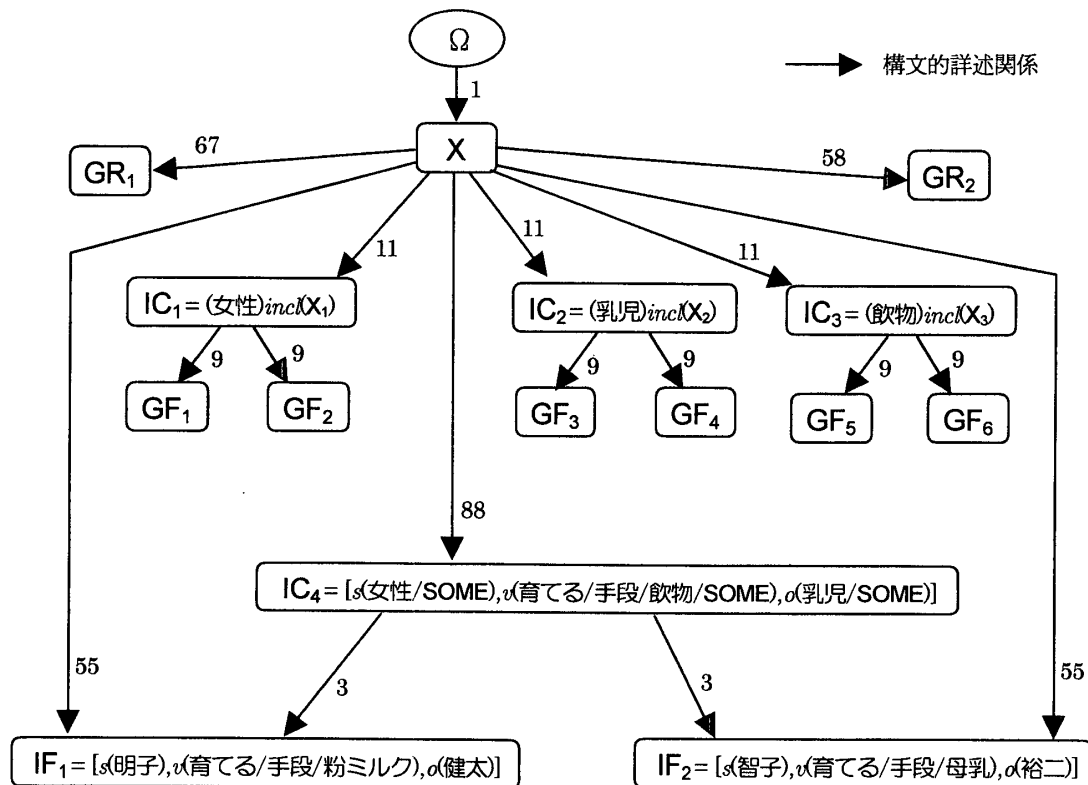


図4 概念の体系

このときの概念体系図を図 4 に示す。体系化の過程で次の新たな概念 (導出概念: *Induced Concept*) が導出されている。

$IC_1 = (\text{女性})\text{incl}(X_1)$: 女性は X_1 (変数) を含む

$IC_2 = (\text{乳児})\text{incl}(X_2)$: 乳児は X_2 を含む

$IC_3 = (\text{飲物})\text{incl}(X_3)$: 飲物は X_3 を含む

$IC_4 = [s(\text{女性}/\text{SOME}), v(\text{育てる}/\text{手段}/\text{飲物}/\text{SOME}), o(\text{乳児}/\text{SOME})]$

: ある女性は, ある乳児をある飲物で育てる。

本例題における総事実記憶量(TM)と実知識量(NK)は, 次のようになった。

総事実記憶量 (TM) = 490

実知識量 (NK) = 417

4. 認識・理解・解釈の枠組み

3章で述べた概念体系を用いて, 外部から入力された概念について, その認識 (Recognition), 理解 (understanding), 解釈 (Interpretation) 等の知的動作を定式化することができる。

以下では, 認識, 理解, 解釈について述べる。

4.1 認識 (Recognition)

知識システムにおいて, 外部入力概念 D_{in} を「認識」するとは, 知識システム内の概念体系 $\{\Omega/\Omega_0, \Psi\}$ において D_{in} に最も近い先祖 D_{rec} を見出すことである (Ω : 概念集合, Ω_0 : 変数 X と $\Omega, \Psi: \Omega$ と詳述量の集合)。つまり, D_{in} を含む知識システム内の最も近い事実データ D_{rec} を見出すことである。このことは, 認識とは日本語で「ああこういうことだな」という表現に対応するように, 与えられた概念を自分の概念体系内と照合して, どれかに分類することである。何も無ければ X (未知の概念) となる。新しい概念やより詳しい概念が与えられた場合に相当する。

概念 D_{in} に対して

$$spec(D_{rec}, D_{in}) = \min_D \{spec(D, D_{in}) \mid D \in \Omega\}$$

を満たす D_{rec} を見出すことを「認識」と定義し,

$$recog(D_{in}, D_{rec})$$

と表記する。

認識の例を以下に示す。

<Ex. 4-1>

Ex. 3-1の知識データが知識システムに与えられているとする。このとき, 次のような外部入力

$D_{in} = [s(\text{明子}), v(\text{である}), c(\text{親/健太})]$

: 明子は健太の親である。

が与えられたとき、システムは、

$D_{rec} = [s(\text{明子}), v(\text{育てる/手段/粉ミルク}), o(\text{健太})]$

: 明子は、健太を粉ミルクで育てる。

と認識する ($recog(D_{in}, D_{rec})$)。

D_{in} が知識システムに入力されたときの概念体系図を図5に示す。

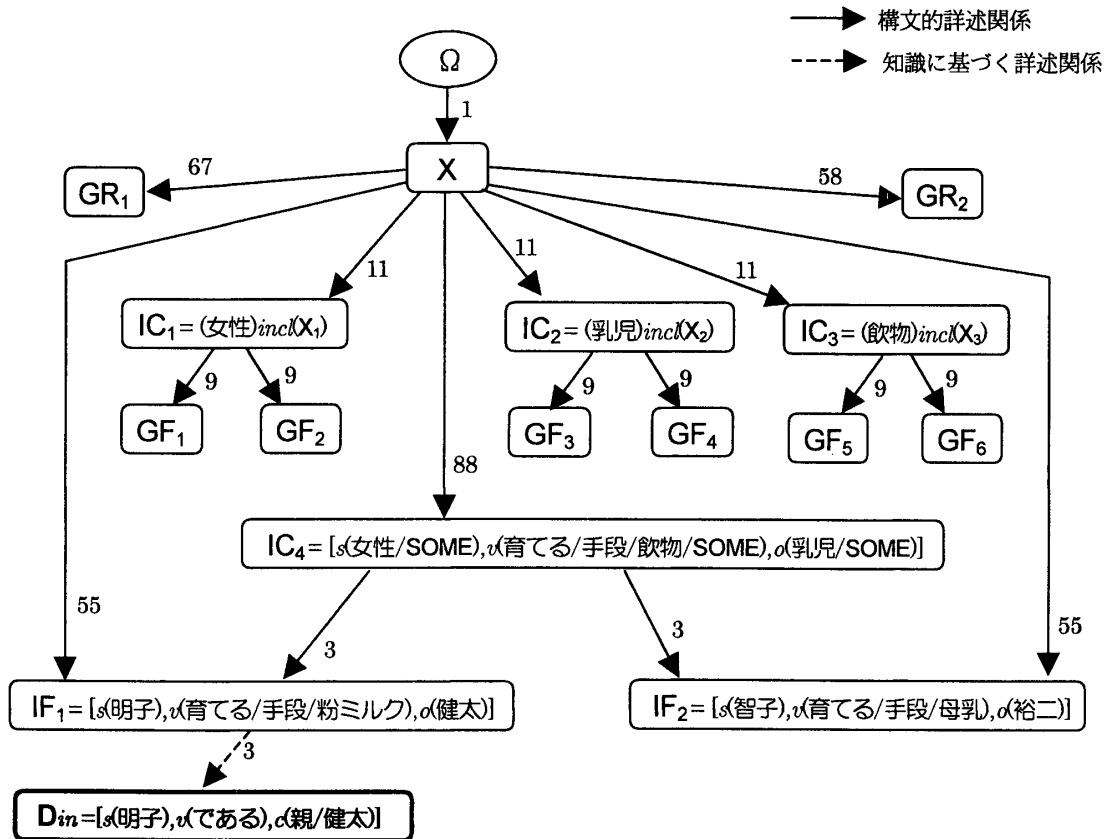


図5 $D_{in} = [s(\text{明子}), v(\text{である}), c(\text{親/健太})]$ が知識システムに入力されたときの概念体系図

4.2 理解(Understanding)

入力概念 D_{in} の理解動作とは、知識システム内の概念体系において $D \in \Omega$ なる任意の D に対して $D_{in} \supseteq D$ を確認することである。つまり、 D_{in} が真であることを確認すること ($true(D_{in})$)^[9]、または、 D_{in} と知識データからの帰結が真であることを確認することである。このことは、理解は、日本語で「ああそうですね」と言う表現に対応するように自分の概念体系内に照合して TRUE (真) と確認することで、理解しているとは、全てを知っていること (概念が全て真) に対応するようにしている。概念体系内から真となれば理解していることになる。

概念 D_{in} に対して

$elab(D, D_{in}), recog(D, D_{in}), D \in \Omega$

を満たす D_{un} を見出すことを「理解」と定義し、

$$\text{underst}(D_{in}, D_{un})$$

と表記する。但し、 D_{un} は概念体系の節 (node) でなければならない。

理解の例を以下に示す。

〈Ex. 4-2〉

Ex. 3-1の知識データが知識システムに与えられているとする。このとき、次のような外部入力

$$D_{in} = [s(\text{智子}), v(\text{産む/過去}), c(\text{裕二})]$$

：智子は裕二を産んだ。

が与えられたとき、システムは、

$$D_{un} = [s(\text{智子}), v(\text{育てる/手段/母乳}), c(\text{裕二})]$$

：智子は裕二を母乳で育てる。

を見出し、システムは D_{in} を理解できたことになる。

D_{in} が知識システムに入力されたときの概念体系図を図 6 に示す。

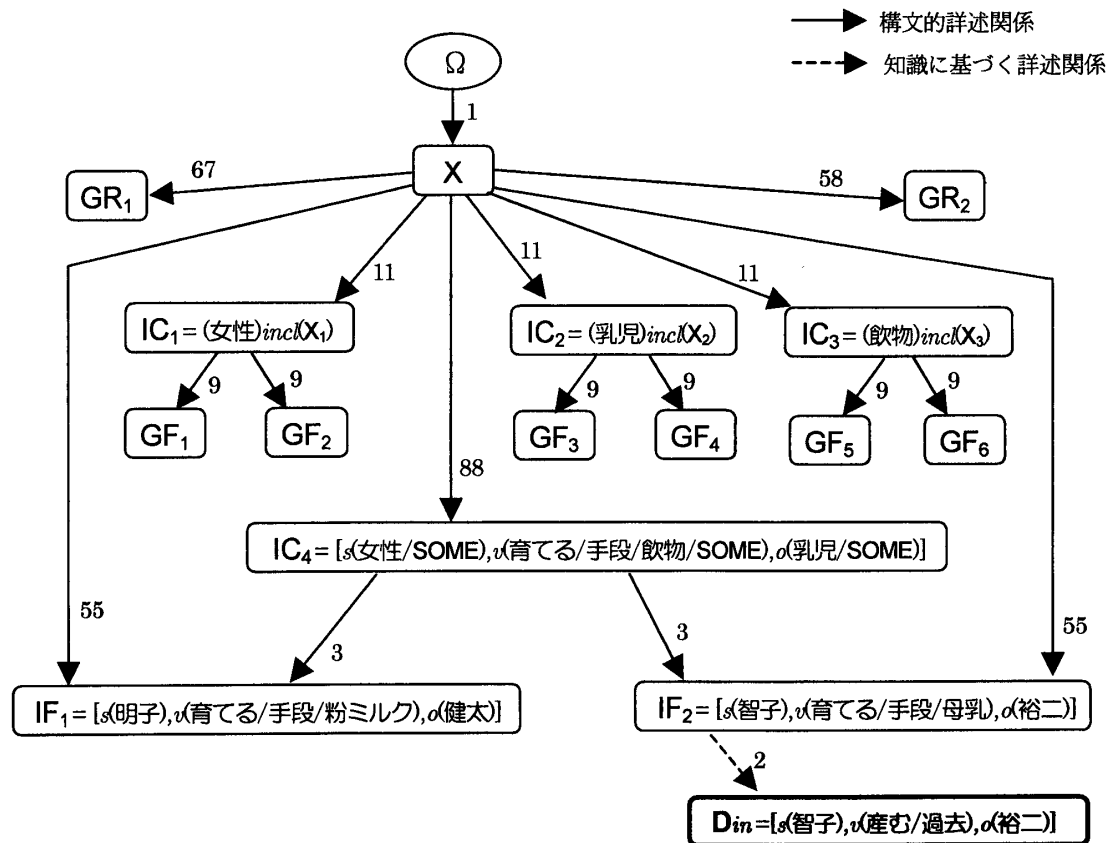


図 6 $D_{in} = [s(\text{智子}), v(\text{産む/過去}), o(\text{裕二})]$ が知識システムに入力されたときの概念体系図

4.3 解釈 (Interpretation)

D_{in} を外部入力とするとき、以下のような関係(1)~(3)を満たす D_{int} を「解釈」と定義し、 $\text{interp}(D_{in}, D_{int})$ と表記する。但し、 D_{int} は概念体系の節 (node) でなければならない。

(1) $spec(D_{in}, D_{int})$ もしくは $spec(D_{int}, D_{in})$ が満たされること。但し, D_{int} は, 概念体系に属する。

(2) $ncoa(D_{in}, D_0, D_{int})$ が満たされること。但し, $D_0 \neq X$ (X : 変数) である。

(3) D_{in} と D_{int} は同じ構文の SD 式であり, D_{in} と D_{int} が対応する部分 $D_{in}(j)$, $D_{int}(j)$ について

$$interp(D_{in}(j), D_{int}(j), n_j), n = \sum n_j$$

が満たされること。

解釈は, 入力概念を SD 式の生成規則 (SDG) [7] から生成される部分木となる多くの概念が該当する。解釈の例を以下に示す。

<Ex. 4-3>

Ex. 3-1の知識データが知識システムに与えられているとする。このとき, 次のような外部入力

$$D_{in} = [s(\text{女性/SOME}), v(\text{育てる/手段/飲物/SOME}), o(\text{乳児/SOME})]$$

: ある女性は, ある乳児をある飲物で育てる。

が与えられたとする。以下は解釈の例である。

$$interp(D_{in}, [s(\text{明子}), v(\text{育てる/手段/粉ミルク}), o(\text{健太})], 3)$$

$$interp(D_{in}, [s(\text{智子}), v(\text{育てる/手段/母乳}), o(\text{裕二})], 3)$$

5. おわりに

SD 式意味モデルでは, 知識システムに対して, 与えられた事実データ (自然言語概念および知識データ) を自ら体系化 (概念ネットワークの構築) できる能力を持たせることが可能である。知識システム内の知識データは, 一つ一つの概念として孤立して蓄えるよりも, 全体を有機的に関連付けて蓄積しておく方が効率的である。SD 式意味モデルでは, 詳述関係および $spec$ 関係を定義し, 概念相互間の意味差の尺度を導入している。詳述関係および $spec$ 関係は半順序関係であるので, これを用いて概念を体系化できる。このような概念体系を用いると知識システムにおける認識, 理解, 解釈等の知的動作の定式化が容易に, そして高速に処理できる。

本論では, SD 式意味モデルで定義している詳述関係を用いて, SD 式で記述された自然言語概念を体系化するアルゴリズムを示した。また, 人間の知的動作を計算機上に実現するために, 認識, 理解, 解釈の知的動作についてモデル化を行った。それらの動作の有効性を SD 式意味モデル実験システム SDENV-3 で確認した。

今後の課題としては,

- 体系化された概念ネットワークに対して, 事実データを知識獲得 (追加・修正・削除) した場合の効率的な再体系化
- 概念の体系化処理の高速化

- ・認識, 理解, 解釈動作以外の知的処理の定式化 (e.g. 学習)
- ・規模の大きい対象世界での動作実験
- ・SD 式-自然言語間のインターフェース

などが挙げられる。

参 考 文 献

- [1] 田中穂積, 辻井潤一 共編: “自然言語理解”, オーム社 (1988)
- [2] 上野晴樹, 石塚満 共編: “知識の表現と利用”, オーム社 (1987)
- [3] 内藤衛亮, 中倉良夫, 影浦峯 他訳: “シソーラス構築法”, 丸善株式会社 (1989)
- [4] Kawaguchi,E., Wakiyama,M. and Nozaki,K.: “A Semantic Structure Description Model of General Concepts in Natural Language World”, Proc. of PRICAI, pp.298-303 (1990)
- [5] Kawaguchi,E., Kamata,S. and Wakiyama,M.: “Elaboration Relation and the Nearest Common Ancestor of a Concept Pair in the SD-Form Semantics Model”, Proc. of 2nd PRICAI, pp.426-432 (1992)
- [6] Kawaguchi,E., Kamata,S. and Wakiyama,M.: “The Semantic Metric Computation Scheme in the SD-Form Semantics Model”, Proc. of 3rd PRICAI, pp.623-629 (1994)
- [7] Wakiyama,M., Shao,G. and Kawaguchi,E.: “The Toward Generalization of the Semantics Metric in the SD-Form Semantics Model”, Proc. of 4th PRICAI'96 Poster, pp.61-68 (1996)
- [8] 野崎剛一, 脇山正博, 河口英二: “自然言語概念の意味構造モデルと会話文データベースへの応用”, 情報処理学会自然言語処理研究会報告81-8 (1999)
- [9] 吉原将太, 脇山正博, 河口英二: “SD 式意味モデル実験システムの試作”, 情報処理学会九州支部研究会報告, pp.247-254 (1999)
- [10] 吉原将太, 脇山正博, 河口英二: “SD 式意味モデルにおける 3 概念からの帰納推論システムの試作”, 純心女子短期大学「紀要」第37集, pp.23-39 (2000)
- [11] Shouta Yoshihara, Masahiro Wakiyama and Eiji Kawaguchi: “An Experiment on Japanese-Sentence Generation from SD-Formed Semantic Data”, In: E.Kawaguchi et al. (Eds.), INFORMATION MODELLING AND KNOWLEDGE BASES XI, IOS Press & Ohmsha, pp.205-221 (2000)
- [12] Masahiro Wakiyama, Shouta Yoshihara, Hideki Noda, Koichi Nozaki and Eiji Kawaguchi: “A Discussion on the Property of the Semantic Space in the SD-Form Semantics Model”, Transactions of Information Processing Society of Japan Vol.41 No.2, pp.509-512 (2000)
- [13] Masahiro Wakiyama, Shouta Yoshihara, Hideki Noda, Koichi Nozaki and Eiji Kawaguchi: “An Extended SD-Form Semantics Model”, Transactions of Information Processing Society of Japan Vol.41 No.2, pp.513-516 (2000)
- [14] Shota Yoshihara, Masahiro Wakiyama and Eiji Kawaguchi: “An Experimental System of Inductive Inference from a Concept-Triple in SD-Form Semantics Model” The 10th European-Japanese Conference on Information Modelling and Knowledge Bases, pp.132-139 (2000)