

英語文意味理解のための認識構造モデル

川辺 諭

宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

時枝誠記の言語過程説を継承、発展させた三浦つとむの日本語文法 [1, 2, 3] では、言語表現を、客観的な対象世界のあり方を述べた“客体表現”と、発話者の加えた主観的な判断である“主体表現”の 2 つに分けて取り扱い、それらは“対象-認識-表現”といった過程において運用されるとしている。このような人間の言語活動のあり方に着目して本稿では英語を対象言語として、言語表現の持つ本質的な意味を理解した自然言語処理を実現するために、言語活動の際に人間の内部で構築されている意味認識構造のモデルである“MUSW”のプロトタイプを、論理型言語 Prolog を用いて計算機内に実装した。MUSW を用いることにより、動作主体の意志の変化などを理解したより本質的な自然言語意味理解が可能となる。

2 MUSW を構成する要素

2.1 プリミティブ

実体を構成する属性(最も原初的な要素)であるプリミティブとして、以下のものを準備する。

1. 知覚プリミティブ: sen

五感などから感性的に知覚される物理的な情報である。また、大脳内のテンプレートへの型はめにより判断されてる、形状や音色といった“知覚情報のプロトタイプ”も準備する。表 1 にプリミティブの詳細、表 2 にプロトタイプの例を示す。

MUSW (Meaning Understanding Structure World) for English Sentence Understanding
Satoru KAWABE, Masahiro MIYAZAKI
Niigata University

表 1: 知覚プリミティブ

感性情報	内容
視覚情報	距離、方向、色、形状、大きさ
聴覚情報	距離、方向、音色、強弱、高さ
触覚情報	温度感覚、圧覚(手触り)、痛覚
味覚情報	～の様な味(プロトタイプ)
嗅覚情報	～の様な臭い()
有機情報	飢え、渇き、疲れ、平衡感覚…

表 2: プロトタイプの例

視覚情報	… \$form(ball).
聴覚情報	… \$tone(piano).

2. 空間: sp

空間と時間(後述)は、感覚器官から得られる物理情報である知覚プリミティブとは違い、人間が大脳の機能を用いてその存在を推測しているものである。そのような空間として 3 次元空間、抽象空間を準備する。

3. 時間: tm

時間も空間と同様に、大脳の機能を用いてその存在が推測されるプリミティブである。

4. 関係: rel

関係として、以下のものを準備する。

(a) 集合

集合は属性間の関係の記述の中でも、最も基本的なものの一つである。

(b) 抽象的關係

日本語の言語規範では、視覚情報プリミティブである赤と青が対義の關係を持つものとして扱われることがある。このような抽象度の高い關係の記述である抽

象的關係として、“類義關係”、“對義關係”、“上下關係”を準備する。

(c) 空間的位置關係

具體物、空間といった實體(後述)同士の間に成立する關係である。“離れている”、“接している”、“重なっている”、“包含關係”の4通りを用意している。

2.2 實體

實體は1つ以上の属性(プリミティブ)の集合として定義する。また、構成された實體は再帰的に他の實體の属性となり得る。MUSWでは實體として以下のものを取り扱う。

```
/* クラス
   :class(class 名, 属性). */
class(c_man, [sex(male)]).
class(c_woman, [sex(female)]).
class(c_boy, [age(young)]).
class(c_girl, [age(young)]).
class(c_animate, [dattr(live)]).
/* インスタンス
   :instance(class 名, instance). */
instance(c_boy, john).
instance(c_girl, mary).
/* 性質継承
   :supsub(上位 class, 下位 class). */
supsub(c_man, c_boy).
supsub(c_woman, c_girl).
supsub(c_human, c_man).
supsub(c_human, c_woman).
supsub(c_animate, c_human).
```

図1:具體物の記述例

1. 具體物:Obj

プリミティブの集合として記述される實體である。MUSWではオブジェクト指向の考え方をういて具體物を記述している。図1にその様子を示す。まず、予め實體とそれ自身が持つ特性とまとめてclassとして記述しておき、実行中はclassやinstanceとして取り扱う。また、各々のclassの間で性質継承の考え方をういている。

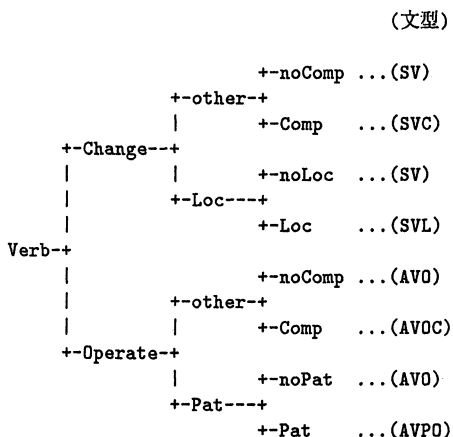
2. 世界

時間変化の概念を扱っておらず、複数の實體の集合として構成される世界を靜的世界(實體SW)とする。また、変化、移動、取り込

みといった“時間変化”を伴う概念は、動的な世界(實體DW)で表す。動的な世界は靜的世界を構成要素とし、それらを時間軸上に配置することにより表現される。

2.3 英語基本動詞コア

英語において言語表現は、話者が着目している實體の属性を動詞を用いて表現することによりなされる。それらの動詞は、實體の属性(Complementとして表記される)や位置(Location)の変化を表現するもの、動作主(Agent)と目的物(Object)や受益者(Patient)との關係を表現するものなど、多様である。このような観点に基づきMUSWでは、英語基本動詞を、(1)實體の変化を扱う/動作主の目的物への操作を扱う(2)変化する属性が實體の位置である/それ以外の属性である—といった点から、4種に大別している(CORETYPE1~4とする)。さらに、(3)変化した属性を明示的に表現する/表現しない—といった観点から8種に分類した(図2参照)。

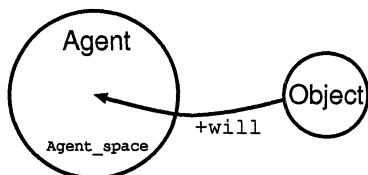


文型略号 S:substance V:verb C:Complement
L:location A:agent O:object P:patient

図2:MUSW 英語基本動詞体系

また、動詞の記述はコア理論[4, 5]で提唱されている英語基本動詞のコア概念を用いた。コア概念とは、対象世界の概念化の際に用いる枠組(言語規範)である英語基本動詞のコアを用いることにより、文脈に依存せず、多義を包摂した、抽象度の高い語義記述が可能となる。図3に英語基本

動詞 “take” のコアイメージと、Prolog によるコア概念記述の様子を示す。



```
verb(take,SL):-
  SL=[111,A,0,P], %111はCORETYPEを表す
  A=[agt,LA1,LA2],
  LA1=[mind(will),mind(will)],
  LA2=[space(agt),space(agt)],
  O=[obj,L01],
  L01=[loc(space(agt)),loc(space(pat))],
  P=[pat,LP1],
  LP1=[space(pat),space(pat)].
```

図 3: “take” のコアイメージと計算機上での記述

3 MUSW の実行結果

4 種に大別した英語基本動詞に関して、MUSW の実行結果を示す。現在言語データとして英語基本動詞のコア概念 18 種類、名詞として実体 32 語を準備している。

ユーザは MUSW に動作主 AGENT、動詞 VERB、目的物 OBJECT、変化前の属性 ATTRB1、変化後の属性 ATTRB2 を入力する (図 4 参照)。MUSW は ATTRB1、ATTRB2 が AGENT、OBJECT の取り得る属性であるか確認し、VERB 入力の有無によって以下の 2 通りの動作を行なう。

(1) VERB が入力された場合、VERB のコア概念を検索し実体 (AGENT、OBJECT は実体の下位クラスとして扱う) の属性がコア概念の要求する属性とマッチするかを調べる。

(2) VERB が入力されなかった場合、入力された実体とその属性の変化を表現するコア概念を推測する。

— ARGUMENT INPUT TABLE —	
AGENT :	John
VERB :	take
OBJECT :	book
ATTRB 1 :	space(John)
ATTRB 2 :	space(nary)
<input type="button" value="RUN"/> <input type="button" value="DESTROY ALL"/>	

図 4: MUSW 要素入力テーブル

3.1 CORETYPE1

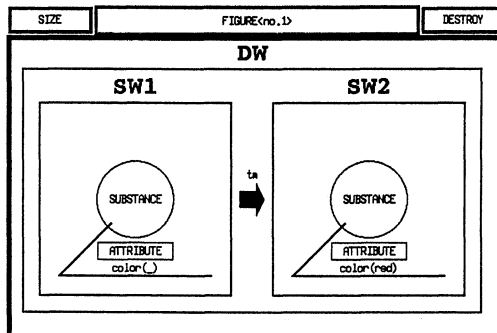


図 5: CORETYPE1

CORETYPE1 は、話者が実体の属性の時間的変化 (変化しない場合も含む) に着目し、それを表現する動詞である。be、become、remain、smile、live などがある。図 5 は、入力文 “apple become red.” (パーザとリンクしていないため文法は簡略化されている) の解析結果である。実体 apple の持つ色合いという属性 (視覚プリミティブ color) が、赤 (color(red)) に変化したということを表現している。

3.2 CORETYPE2

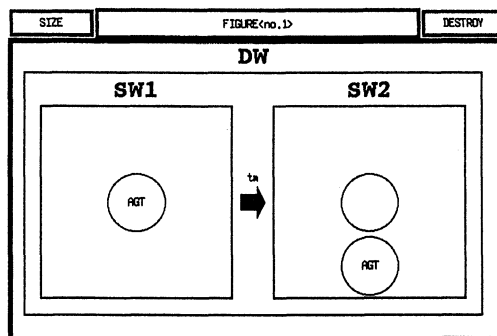


図 6: CORETYPE2

CORETYPE2 は、実体の位置属性の変化を表現する動詞である。go、come、jump、down、back、fall、sink、fly などがある。変化する属性は実体の位置であり、この点、CORETYPE2 は CORETYPE1 の特殊な場合ということが出来る。図 6 は入力文 “ball fall.” の解析結果であり、実体 ball が下方へ向かって移動したことを表現している。

3.3 CORETYPE3

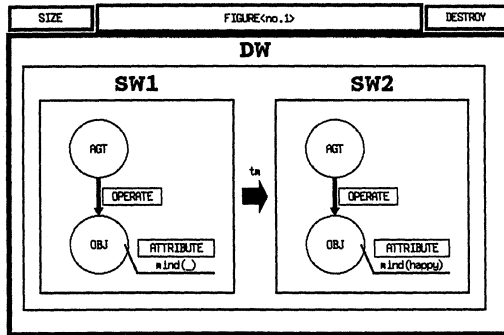


図 7: CORETYPE3

CORETYPE3 は、動作主 AGENT が目的物 OBJECT を操作することによりその属性を変化させる動詞であり、これは CORETYPE1 で扱っている実体の属性変化が、別の実体の意志的な操作によってなされる場合である。図 7 は入力文 “john make mary happy.” の解析結果である。動作主が目的物を操作し、“幸福である” という属性を与えている。これらの動詞としては keep、hold などがあ

3.4 CORETYPE4

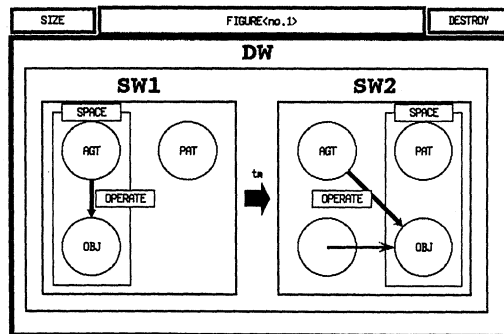


図 8: CORETYPE4

CORETYPE4 は、動作主 AGENT が目的物 OBJECT の位置属性を操作し、受益者 PATIENT へ移動させる動詞であり、これは CORETYPE2 で扱っている実体の位置属性変化が、別の実体の意志的な操作によってなされる場合である。図 8 は入力文 “john give mary book.” の解析結果である。目的物が動作主を持つ空間から受益者の持つ空間へと移動している。これらの動詞としては give、take、lend などがあ

4 おわりに

人間の言語理解のあり方に適ったより本質的な自然言語処理を実現するために、言語活動の際に人間の内部で構築されている意味認識構造のモデルを、MUSW として計算機内に再構築し動作させる方法を提案した。MUSW を用いることにより

- 言語表現の持つ本質的な意味の解析
- 似通った言語表現からの意味的差分の抽出
- 同一表現の持つ複数の意味合いの解析
- 比喩表現の解析

などが可能になる。現在、

- 言語データである実体や動詞コア概念の拡充
- 形容詞、副詞、前置詞などの品詞の追加
- 様相、時制などの表現の処理を実現する方法
- 構文解析部とのリンク
- 言語データからの英単語ネットワーク構築法

などを検討中である。

参考文献

- [1] 三浦つとむ: 認識と言語の理論, 第 1 部～第 3 部, 勁草書房 (1967/1967/1972).
- [2] 三浦つとむ: 日本語の文法, 勁草書房 (1975).
- [3] 三浦つとむ: 日本語とはどういう言語か, 講談社学術文庫 (1976).
- [4] 田中茂範: 認知意味論 英語基本動詞の多義構造, 三友社出版 (1990).
- [5] 田中茂範 編著: 基本動詞の意味論, 三友社出版 (1987).
- [6] 宮崎正弘: 言語を理解するコンピュータ 自然言語技術の展望, コンピュートロール. コロナ社, no.37, pp.75-81 (1992).