

有向グラフによる自然言語意味表現形式と付属語の意味表現

三吉秀夫¹⁾ 奥西稔幸²⁾ 阿部ひろみ²⁾ 小淵保司²⁾

¹⁾ (株) 日本電子化辞書研究所

²⁾ シャープ株式会社

自然言語文の意味をわかりやすくかつ統一的に表現するための意味表現形式とそれを用いて日本語の付属語の意味表現を行なう方式を述べる。意味表現の枠組には共有有向グラフSDGとよばれる意味ネットワークを導入している。付属語には吉田らが提唱した規格化日本語の付属語表現を採用した。付属語は日本語文において意味的に重要な役割を担っているため、意味表現においてもキーとなる。我々は規格化日本語で規定されている約350個の付属語相当表現の代表表現に対して意味表現方式を定めた。また本方式に基づいて日本語文の意味構造を抽出するシステムを開発した。

Semantic Representation for Natural Language using Directed Graphs and its Application to Japanese Fuzokugo Family

Hideo Miyoshi¹⁾ Toshiyuki Okunishi²⁾ Hiromi Abe²⁾ Yasuji Obuchi²⁾

¹⁾ Japan Electronic Dictionary Research Institute LTD.
c/o Sharp Corporation Tokyo Research Laboratories
273-1 Kashiwa, Kashiwa-shi, Chiba 277, Japan

²⁾ Sharp Corporation
492 Minosho-cho, Yamatokoriyama, Nara 639-11, Japan

This paper presents a framework to represent semantic structures of natural language sentences and a mechanism for representing the semantics of Japanese fuzokugo families within it. This framework, called SDG (Shared Directed Graph), is based on labeled directed graphs. It is well known that Japanese fuzokugos have crucial roles in sentence meaning because they describe not only the relations among objects and events but the modalities of events. We have analyzed about 350 prototype fuzokugos defined in Yoshida's "Controlled Japanese". Based on this idea, we have developed an experimental system for extracting semantic representations of Japanese sentences.

1. はじめに

自然言語の意味をいかなる形式で表現するかというテーマは、人工知能(AI)とりわけ自然言語処理において重要なテーマであり、従来からいろいろな研究が行なわれている。代表的なものとしてはAI分野の知識表現の枠組を利用したものであり、フレーム形式、意味ネットワーク、論理形式などが挙げられる。しかしこれらの形式は意味表現の枠組は提供するが、実際に個々の言語現象に対してどのような表現形式にすべきかという規定はしていない。また、言語に依存しない意味表現形式としてはSchankの提案した概念依存モデル[1]が挙げられる。概念依存モデルは言語の表す意味を高度に抽象化した概念レベルの構造を用いて表現するものであり、特定言語への非依存性、意味表現形式の一意性という点では評価できるが、構造が非常に複雑になるうえ、カバー範囲の点で問題がある。

我々は、規格化日本語[2]に基づいた自然言語処理システム[3]の開発を行ってきた。本システムはアウトプットの1つとして日本語入力文の持つ意味構造を出力する。本意味構造(意味表現形式)を構築するに当たっては、できるだけ広範囲の言語現象を扱えるとともに意味表現形式の統一性と一般性を保持できるものを念頭に置いて考えてきた。また、この意味表現は人間にとっても判り易い視認性の高さを持つことを重要視している。このような条件を満たすものとして我々は有向グラフを利用したSDG(Shared Directed Graph)と呼ばれる意味表現形式を考案し、様々な言語現象に対する意味表現を考えてきた[4]。本稿では特に付属語表現の表す意味構造をSDGで表現する方式について報告する。日本語の付属語は日本語文に於いて意味的に重要な役割を担っているため、付属語の意味を決めることにより日本語文の意味表現の大枠を決定することができる。今回は規格化日本語の付属語で規定されているすべての代表表現に対して意味構造を付与したので、表現可能な言語現象の範囲を拡張することができた。この意味構造は各種自然言語インタフェースや翻訳システムの間言等への応用が期待できる。一方、本研究内容は規格化日本語にとっては「語彙、構文の規格化」だけでなく、更に「意味構造の規格化」への拡張と位置付けることができる。

本稿ではこのSDGによる意味表現形式で中心的な役割を担う付属語の意味表現形式について報告する。2節では本研究の背景と位置付けを述べる。3節では、本研究のベースとなっている規格化日本語の概要とその付属語表現について述べる。4節では、本意味表現形式の枠組であるSDGの概要とそれを用いた各種付属語の意味表現方式について説明する。5節では、本意味表現自体を処理対象とする応用

例として縮退表現について述べる。

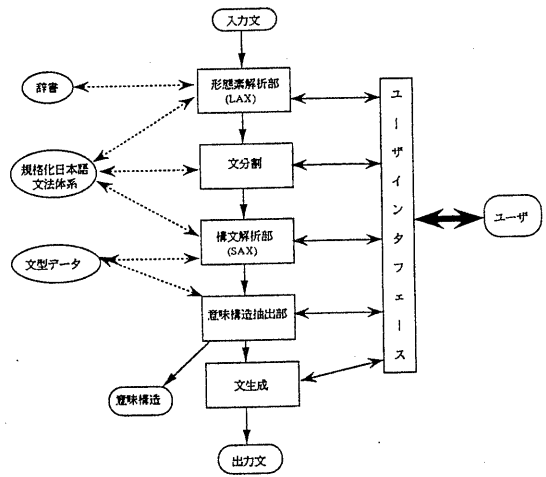


図2.1 ICOTEXTの処理概要

2. 本研究の背景と位置付け

本論に入る前に本研究を行なった背景と位置付けを簡単に述べておく。我々は規格化日本語に基づいて日本語文の解析・生成を行う自然言語処理システムICOTEXTを開発してきた。図2.1にICOTEXTの処理概要を示す。ICOTEXTは各種の自然言語インタフェースのフロントエンドとしての様々な利用形態を想定しており、現在2種類の異なるアウトプットを出力する。第1の出力は、入力文がもつ曖昧性を排除し、規格化日本語に則ってより明瞭に書き換えられた表層文である。これはICOTEXTが持つ校正支援的な機能であり、機械翻訳のように言語そのものが入力となるシステムへの利用を想定したものである。第2の出力は、入力文に対する意味処理を行なった結果抽出される意味表現形式である。これは自然言語による計算機とのインタラクションのように他のシステムとのユーザインタフェースとしての利用を想定したものである。本論文の内容は後者の意味表現形式に関するものである。我々は以下の方針に基づいて意味表現形式の設計を行なった。

- (1) 広範な言語現象に対する意味構造を表現出来ること。
- (2) 他のアプリケーションとのインタフェースに利用可能なこと
- (3) 視認性/操作性にすぐれていること
- (4) 特定言語に依存しない、

(1)は対象分野を限定せず、出来るだけ多くの種類の文を扱いたいという目論みである。逆にいえば、概念依存モデルのような深い意味分析は行なわない。(2)はユーザインタフェースとして当然の条件である。(3)は、この意味構造自体がユーザからの操作の対象にもなり得ることを想定している。(4)はまだ実現できてはいないが、最終目標としてこの意味表現形式は入力言語に依存しない普遍的なものにしたいと考えている。つまり、中間言語方式機械翻訳の中間表現的な位置付けであり、同一意味内容を持つ表現が本意味表現形式において同一の形式に変換される。

今回付属語に注目した理由は以下の2点である。

- (1) 日本語の付属語は「モノ」や「コト」の間の関係を表す表現であったり、「コト」の在り様(モダリティなど)を表す表現であるので、文の意味において重要な役割を持っている。
- (2) (1)の理由により、付属語の意味表現を決めることにより文の意味内容の大枠を定めることができる。

またその際、規格化日本語の付属語を採用することは以下のメリットがあると考ええる。

- (1) 規格化日本語は純粋な助詞/助動詞だけでなく付属語相当表現も付属語として扱い、その豊富な表現を意味別に分類しており、各グループに代表表現を与えている。

- (2) 従って、代表表現毎に意味表現を決めることにより大部分の付属語表現の意味表現を決めることが可能。つまり、意味表現構築のガイドラインとして規格化日本語は有効である。

3. 規格化日本語の概要

意味表現形式について述べる前に本研究のベースとなっている規格化日本語について簡単に紹介しておく。規格化日本語[2]は、曖昧性の少ない機械処理に適した日本語技術文書を作成することを目的として、吉田らによって提案された制限日本語文法である。その骨子は次の4つである。

- (1) 付属語的表現の規格化
- (2) 自立語の規格化
- (3) 係り受け関係の規格化
- (4) 長い文の分割

本研究に関連するのは(1)の付属語的表現の規格化であり、規格化日本語では付属語的表現を次の表3.1のように5つの大分類、計13種類に分類している。それぞれの付属語的表現には「意味、代表表現、非代表表現」が与えられており、該当の意味で使われている非代表表現が出現した場合は代表表現に置き換えられる。例えば、「目的」を表すRppとして表3.2のように定められている。

表3.1 規格化日本語の付属語的表現の分類

大分類	付属語の種類	性質	代表表現の個数	代表表現の例
関係表現	Rnp	体言から用言への係り	78	が、とは
	Rnn	体言から体言への係り	42	という、における
	Rpn	用言から体言への係り	14	ことによる、ための
	Rpp	体言から体言への係り	51	ので、た結果
副助詞的表現	副助詞的表現	副助詞相当	14	しか、ばかり
助述表現	App	用言に接続し用言の資格を持つ	96	ている、と判断する
	Anp	体言に接続し用言の資格を持つ	17	になる、にすぎない
接尾語的表現	Snn1	助数詞相当	-	年、倍
	Snn2	助数詞以外の接尾語	-	前、以上
	Snp	体言に接続して用言になる	6	できる、的である
	Spn	用言に接続して体言になる	3	さ、方、目
	Spp	用言に接続して用言になる	8	てもらう、にくい
形式名詞的表現	形式名詞的表現	形式名詞相当	16	こと、か否か

表3.2 目的を表すRppの代表表現と非代表表現

分類	意味	代表表現	非代表表現
Rpp	目的	ために	ため、に、には、目的のため 目的で、という目的で、etc.

これに則り、「煙草を買う（ため／目的で／という目的で／...）外出した」という文は「煙草を買うために外出した」と書き換えられる。

4. SDGによる付属語の意味表現

本節では意味表現形式SDGの概要と、その枠組における付属語の意味表現形式について述べる。

4.1 SDGの概要

SDGは入力文に対する意味処理により抽出した意味構造をわかり易く表示するための意味表現形式であり、基本的には有向グラフの形式を持っている。1節で述べたように本システムではさほど深い意味処理は行っていないかわりに、なるべく広範囲の言語現象の意味を統一的な形式に表現できるよう工夫を施している。また、視覚的にも解り易いものであることを心掛けている。従って、概念依存モデルのような複雑ではなく、比較的シンプルな構造をしている。体言・用言は、それぞれグラフの節点として表現される。SDGの基本的特徴は以下の通りである。

- ・用言を中心とした格関係のネットワーク（格関係は、格要素と用言節を結ぶ枝（ラベル付）で表される）
- ・グラフの枝は係り元から係り先に向かう方向を持つ（情報付加的役割しか持たない枝は方向性無し）
- ・同一格要素の共有表示（SDGの由来）
- ・等位構造を表す特別節点の導入

図4.1に「私がパンを食べる」という文に対するSDGによる意味表現の例を示す。

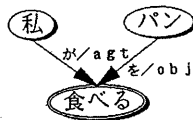


図4.1 「私がパンを食べる」の意味表現

図4.2は体言の等位構造を含む文の例であり、「健がリンゴとミカンを食べる」の意味表現である。図中の「and」節が等位構造を表している。

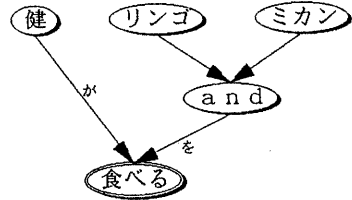


図4.2 「健がリンゴとミカンを食べる」の意味表現

4.2 付属語の意味表現形式

1節で述べたように、日本語の付属語は「モノ・コト」相互の関係や「モノ・コト」の在り様を表現するためのものであり、意味的に重要な役割を持っている。従って付属語の意味表現方式を決めれば、日本語文の意味表現の大半は定まる。規格化日本語で規定されている付属語表現があらゆる付属語的概念をカバーしているとは思えないが、我々は意味表現を構築していく上で規格化日本語は有効なガイドラインになると考える。本節では3節で述べた規格化日本語の付属語に対する意味表現形式について説明する。なお、規格化日本語には「意味」が付与されていないものや、表現が不適切と思われるものが幾つかあったので、それらには我々が意味を付加したり修正を加えた。

4.2.1 関係表現

文節間の係りを表す関係表現は次の4種類である。

- (1) Rnp（体言から用言、格助詞的表現）
- (2) Rnn（体言から体言）
- (3) Rpn（用言から体言）
- (4) Rpp（用言から用言、接続助詞的表現）

これら4種類は統語的振舞いが異なるだけで、意味的には共通のことが多い。例えば「目的」を表す「ため」から派生した代表表現はこれら4種のいずれにも含まれている。

- 例
- Rnp: のために
 - Rnn: のための
 - Rpn: ための
 - Rpp: ために

従って、これら4種は共通の形式を持つことが望ましい。

〔関係表現の意味表現の原則〕

関係表現は文節間の係り受け関係を表すので係り元の文節と係り先の文節を結ぶ枝として表す。枝には規格化日本語で規定されている「意味」をラベルとして付与する。

図4.3は「原因理由」を表すRpp「ので」含む文「彼が来ないのですぐに家を出発した」に対する意味表現である。

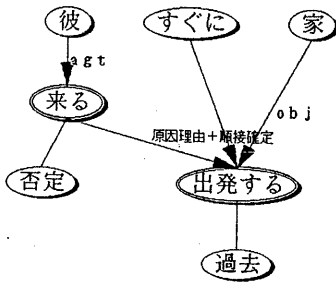


図4.3「彼が来ないのですぐに家を出発した」の意味表現

しかし、関係表現の中には上記原則ではその意味を正確に表現できないものもある。それらの意味表現方法は個別に決める。例えば「否定的限定」を表すRpp「しか(～ない)」を用いた文

「AするしかBするNpがない」 ---- (A)

の意味表現は図4.4のようにする。

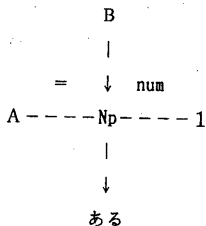


図4.4 「AするしかBするNpがない」の意味表現

図4.4の構造が表す内容は、「BするNpが唯一存在し、そ

れがAである」という言明であり、(A)をパラフレーズしたものである。これを使った例文「バスに乗るしか駅に行く方法がない」の意味表現を図4.5に示す。

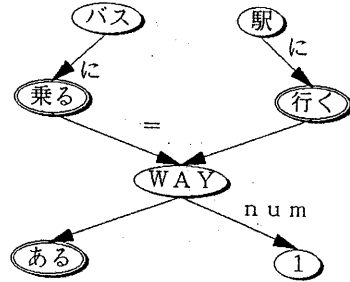


図4.5 「バスに乗るしか駅に行く方法がない」の意味表現

もちろん、「関係表現の意味表現の原則」に問題がない訳ではない。例えば、「順」を表すRpp「順に」を使った句「飲んだ順に」には少くとも、

1. 飲み終わった順に
2. 飲んだ分量の多い順に
3. 飲み始めた順に

の3通りの解釈が存在する。現在の枠組ではこれらの解釈の違いが表現できない。

4.2.2 副助詞的表現

規格化日本語には副助詞的表現として「は、も、だけ、しか、こそ、…」の14個が規定されている。これらはRnp、Rpp、体言に接続し、通常の格関係に何らかの意味を付加したり限定を行なう作用がある。そこでSDGでは副助詞的表現の意味表現を次のように定める。

〔副助詞的表現の意味表現の原則1〕

Rnp、Rppに接続する場合は、Rnp、Rppに相当する枝のラベルに副助詞的表現の意味を付加する。

本原則に相当する副助詞的表現は「は(提示)、も(添加)、だけ(限定)、こそ(強調)、でも(最低範囲)、さえ(他の強調)、までも(強意+添加)、さえ(十分条件)、ほど(比例)、ばかり(頻度付限定)」である。図4.6に「雨が

降る場合こそ延期する」の意味表現を示す。これは、Rpp
「場合」に副助詞「こそ」が接続する例である。

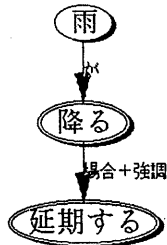


図4.6 「雨が降る場合こそ延期する」の意味表現

[副助詞的表現の意味表現の原則2]

体言に接続後、Rnpや格助詞があとに続く場合は、原則1
と同様にRnpや格助詞に相当する枝に副助詞的表現の意味
を付加する。

原則2に相当する副助詞的表現は「も（添加）、だけ（限
定）、こそ（強調）、でも（最低範囲）、までも（強意+
添加）、さえ（十分条件）、ほど（比例）、ばかり（頻度
付限定）」である。

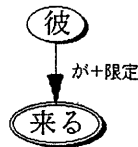


図4.7 「彼だけが来る」の意味表現

4.2.3 助述表現

助述表現は助動詞相当の表現であり、接続先が用言であ
るか体言であるかによって、AppとAnpの2種類に分類され
る。Anpは「Np1はNp2である」といういわゆる名詞文に対
する補助用言である。従ってAppもAnpも文で記述されてい
る内容に対してある種の意味を付加する機能を持つという
点で、両者の意味表現には特に大きな相違はない。

[Appの意味表現の原則]

Appを除去した文の意味表現における主文の用言文節の
節点と、Appの意味を表す節点を接続する。これら2節点
間の枝には方向を持たせない。

図4.8は「彼が本を読みはじめる」という文に対する意味
表現である。

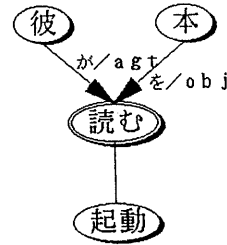


図4.8 「彼が本を読みはじめる」の意味表現

Anpの意味表現を決めるためには「Np1はNp2である」とい
う文の意味表現を決めておく必要がある。我々の枠組では
これは図4.9のように表現される。

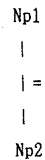


図4.9 「Np1はNp2である」の意味表現

図4.9中で「=」のラベルは必ずしも等価という意味を表
しているのではない。本来「Np1はNp2である」という文で
は、Np1とNp2の間には等価関係、上位・下位関係だけでな
く、ウナギ文にみられるような様々な関係が存在する。こ
の「=」はこれらをすべて包含したものである。これを考
慮し、Anpの意味表現を次のように定める。

[Anpの意味表現の原則]

Anpを除いた核文「Np1はNp2である」の意味表現におけ
るNp2に相当する節点にAnpの意味を表す節点を接続する。
これらの節点間の枝に方向性はない。

図4.10は「これが正解とは限らない」の意味表現である。

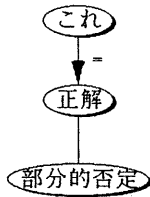


図4.10 「これが正解とは限らない」の意味表現

4.2.4 接尾語的表現

接尾語的表現には以下の5種類がある。

Snn1 (体言に接続し体言になる。助数詞相当)

例：回、月、つ、時

Snn2 (体言に接続し体言になる。助数詞以外の接尾語)

例：後、前、以上、用、度

Snp (体言に接続し用言になる)

例：する、できる、可能である、的である

Spn (用言に接続し体言になる)

例：さ、方(カタ)、目

Spp (用言に接続し用言になる)

例：くする、させる、にする、にくい、てもらう

SnpはAnpと類似しているが、Snpは前接する体言が連体修飾を受けないという点でAnpと区別される。またSppもAppと類似しているが、Sppは表層格を変化させる、という点でやはりAppと区別される。これらのうち、Snn1とSnn2については前接する体言を含んだ1個の複合名詞として扱うので特に意味表現は用意していない。Spnのうち「目」には動詞が前接するのでその動詞を含めて1語の名詞として扱う。「さ、方(かた)」にはAppやSppが伴った用言句が前接するので、その用言句が表す事象の意味表現に「さ、方(かた)」という節点を付加した形式で表す。特にこの2つは、意味表現を見易くするために5節で述べる縮退表現という簡略表示法も導入している。

図4.11に名詞句「動かしにくさ」の意味表現を示す。

Sppの意味表現は基本的にはAppと同じである。但し、Sppそのものが主体を伴っていると考えられるもの(例えば、「させる(使役)、られる(受身)、てもらう(受益)」など

は、Sppの節点に通常の格要素と同様も主体の節点を付加する。図4.12に「人は努力するほど認められる」の意味表現を示す。



図4.11 「動かしにくさ」という名詞句の意味表現

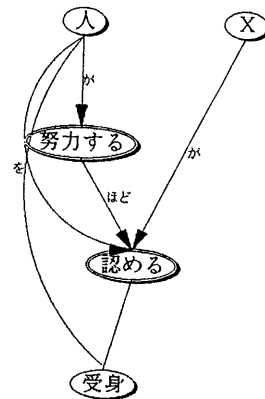


図4.12 「人は努力するほど認められる」の意味表現

4.2.5 形式名詞的表現

規格化日本語では形式名詞的表現として16個の代表表現が規定されている。SDGではこれらは基本的には普通名詞と同じく1個の節点として表現する。但し「こと、だけ、方(ほう)」は、前接する用言を体言化する機能があるため、5節で述べる縮退表現による表示も導入している。

図4.13は「煙草を吸わないこと」の意味表現である。

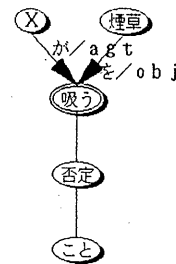


図4.13 「煙草を吸わないこと」の意味表現

なお、形式名詞的表現の意味についてはあまり深く分析していないので、今後の検討の余地が多い。

5. 意味構造に対する操作

本システムでは4節で述べたように、Spnと形式名詞的表現の一部に縮退表現という表示方式を導入している。これは意味構造が複雑になる可能性を持つ付属語に対して、見易く表示する目的で導入したものである。図5.1と図5.2はそれぞれ図4.11と図4.13を縮退した表現である。この縮退形式で表現された部分を展開して表示するためには、縮退表現の部分をクリックすればよい。



図5.1 「動かしにくい」の縮退表現

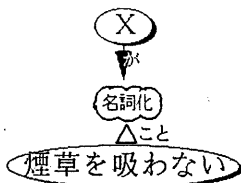


図5.2 「煙草を吸わないこと」の縮退表現

6. おわりに

本システムはICOTで開発された逐次型推論マシンPSI上を実現されている。本稿で報告した方式によればかなり広範囲の言語現象を統一的な形式で表現することができる。しかし本方式は視認性を重視するため細かい意味内容までは表現することができない。逆に細部まで正確に表現しようとするれば、構造が非常に複雑になることが予想される。このトレードオフを解決するには多重レベルの表現形式が必要となろう。縮退表現はその一つの試みである。また、名詞句の意味論や限量詞といった、現在意味論で問題になっている現象は扱うことができない。これらの解決を目指しながら、また同様の研究成果として中間言語方式機械翻訳における概念表現[5]やEDRの概念記述[6]などを参考にしながら更に精緻な意味表現形式を構築することが今後の課題である。

なお本研究は第五世代コンピュータプロジェクトの一環としてICOTからの委託として行なったものである。また、本システムのプログラムは「ICOT無償公開ソフトウェア」のひとつとして利用が可能である。

[謝辞]

本研究を行なうにあたり御指導・御助言を頂いたICOT第6研究室の田中室長に感謝致します。

[参考文献]

- [1] Schank, R. C., Conceptual Information Processing, North-Holland, 1975.
- [2] 吉田ほか, 規格化日本語による文書の作成に関する研究, 昭和61年度科研費補助金一般研究(B)研究報告書.
- [3] 小淵ほか, 規格化日本文作成支援の一方式, 情報処理学会第36回全国大会1T-8, 1988.
- [4] 三吉ほか, 規格化日本語に基づく付属語の意味表現, 情報処理学会第44回全国大会1N-8, 1992.
- [5] 牧野, 機械翻訳(電子情報通信学会編), オーム社, 1989.
- [6] EDR, 概念辞書(第3版), EDR TR-020, 1990.