

日英機械翻訳のための意味属性体系

宮崎 正弘*1 池原 悟*2 横尾 昭男*3 白井 諭*3

- *1 新潟大学工学部情報工学科
〒950-21 新潟市五十嵐2の町8050
miyazaki@info.eng.niigata-u.ac.jp
- *2 鳥取大学工学部知能情報工学科
〒680 鳥取市湖山町南4-101
ikehara@ike.tottori-u.ac.jp
- *3 NTTコミュニケーション科学研究所
〒239 神奈川県横須賀市光の丘1-1
{ayokoo,shirai}@cslab.kecl.ntt.co.jp

あらまし： 日英機械翻訳における高品質な意味解析を実現するため、語彙に関する知識を収録した単語意味辞書と用言を核とした表現構造を収録した構文意味辞書を結び付けるものとして、意味属性体系を提案し、その具体的構成を示した。意味属性体系は、対象の見方、捉え方、すなわち、対象を概念化する際の視点（単語意味属性）を約3,000に分類し、体系化したもので、新聞記事など現代日本語の記述文で使用される、単語意味辞書に収録された固有名詞20万語を含む合計40万語を対象に意味属性が付与されている。
キーワード： 機械翻訳、意味解析、シソーラス、単語意味辞書

Semantic Attribute system in Machine Translations

Masahiro MIYAZAKI*1, Satoru IKEHARA*2,
Akio YOKOO*3, Satoshi SHIRAI*3

- *1 Faculty of Engineering, Niigata University
8050 Ikarashi 2-nocho, Niigata 950-21, Japan
miyazaki@info.eng.niigata-u.ac.jp
- *2 Faculty of Engineering, Tottori University
4-101 Koyamacho-minami, Tottori 680, Japan
ikehara@ike.tottori-u.ac.jp
- *3 NTT Communication Science Laboratories
1-1 Hikarinooka, Yokosuka 239, Japan
{ayokoo,shirai}@cslab.kecl.ntt.co.jp

Abstract: This paper presents semantic attribute system with 3,000 attributes classified by the viewpoint of conceptualization of the object, for semantic analysis in machine translation. A semantic word dictionary for knowledge related to semantic use of word and a semantic structure dictionary for knowledge related to meanings of expression structures are connected by the semantic attributes. We gave semantic attributes for 400,000 words of modern Japanese stored in the semantic word dictionary for machine translation.

Key words: machine translation, semantic analysis, thesaurus, semantic words dictionary

1. はじめに

機械翻訳システムにおいて、高品質な訳文を得るには、言語表現中の単語がどのような約束（語義）で使用されているか正確に決定できる質の良い意味解析が要求される。ここで、概念化の過程と概念を単語に対応させる方法について考えてみると、これらは、対象とする実体の見方、捉え方に大きく依存し、同一の対象でも見方、捉え方によって使用される単語に違いが生じることがわかる。また、逆に、一つの単語を一つの語義で使ったとしても、その表す概念は様々である。このような現実の言語表現上での用法に関する知識を計算機に処理可能な形に整理する必要がある。

本稿では、上記のような考えに基づき、対象の見方、捉え方、すなわち、対象を概念化する際の視点（単語意味属性）を分類・整理することにより名詞の意味体系化を行うものとして名詞の意味属性体系を提案し、その具体的な構成法、構築法、および機械翻訳システムにおける検証結果について述べる。

2. 意味属性体系化の方針

2.1 意味属性体系化の狙い

話者の対象認識の結果として得られた実体の概念は表現上、単語（名詞）に対応づけられる。単語と概念の対応関係は必ずしも一対一とは限らず、複数の概念が一つの単語と対応づけられることも多い。通常、単語の意味と言われているものは、このような単語と概念の対応関係に関する約束のことであり、厳密には語義である。従来、単語の意味（語義）を扱う方法として、意味素、意味索性、意味標識などを用いる方法が考えられている。その違いは必ずしも明確でないが、大きくみて、単語の意味をさらに細かな要素に分ける立場と、単語の意味を対象の持つ特徴の総体で捉える立場に分けられる。ここでは、部分的表現の意味は全体の中で決まることを基本的な立場とし、その点で、単語の意味について還元論的立場をとらない。すなわち、名詞の語義で表される概念は、認識の単位（第1回概念化の結果）として一定のまとまりを持った総体であり、意味の要素に分けられないものとする。そのような意味で、ここでは、対象の見方、捉え方に着目して名詞を分類する基準として単語意味属性を用いる。単語意味属性とは、対象を概念化する際の視点を整理したものであり、話者からみれば、単語の用法に相当する。

従来、同義語や類義語を語義により体系的に分類・整理し、階層的な木構造にしたものとして、分類語彙表¹⁾や角川類語新辞典²⁾など人間用のソーラスがある。近年これらは、電子化され、計算機処理に使えるようになってきているが、以下の点で意味処理向きではない。

(1) 分類観点が必ずしも明確でなく、上位-下位関係や全体-部分関係が識別子なしに混在していたりする。また、角川類語新辞典では、種々の連想関係をもつものが混在している。さらに、この辞典では、図書の十進分類法に基づき、大、中、小、最小分類の4階層に分類しているが、単語の意味分類は、本来このような固定的な枠組みになじまない。

(2) 意味処理では語義だけでなくその語義で表される対象概念の種々の見方、捉え方が必要である。例えば、「学校」は「組織」、「建物」、「場所」など種々の見方、捉え方があがるが、1つの分類項目（ソーラス上のノード）にしか掲げられていない。また、分類語彙表では原則として単語の多義は考慮せず、多義語を最も基本となる分類項目にのみ掲げている。

一方、意味処理を前提としたものに、分類観点の明確なソーラスの構築を目指した東工大ソーラス³⁾・⁴⁾、概念を分類・体系化したEDR概念辞書⁵⁾・⁷⁾などがある。

これらは、上記の(1)～(2)の問題点を解決しようとするものである。前者は分類項目数約1,000である程度の大きさがあるが、収録語数が少ない試作段階のものである。後者は収録語数が20万語と大規模で、概念分類項目も約6,000と多いが、分類項目の大半は終端ノードに偏っており、日英機械翻訳に適用した場合、文型パターン記述の柔軟さに欠ける問題がある。また、新聞記事に頻出する固有名詞に対して特別な意味体系を備えていない。語彙統計によれば、新聞記事に出現する単語数（短単位）は3,000語で80%カバー率となること⁸⁾、英会話でよく使われる単語数は2,000～3,000語とみられること、通常使いこなされる漢字数は2,000～3,000であることなどを参考に、分類項目数は3,000を当面の目標にした。

以上の考えに基づき、対象の持つ特殊性を捨象する立場から概念化の視点（単語意味属性）を約3,000に分類し、名詞の意味属性体系を構築する。主たる分類対象語は、新聞記事などのような現代日本語の記述文で使用される、日英翻訳用の単語辞書に収録された一般語12万語、固有名詞20万語、専門用語（電気電子、情報関連）5万語、その他（時事用語など）3万語の合計40万語である。なお、意味属性体系の構成や構築にあたっては、単に新聞記事だけでなく、現代日本語の記述文全般に適用できるよう汎用性を重視する。

2.2 意味属性体系化のポイント

名詞の意味属性体系の基本構成を検討するにあたり、以下の点に考慮して体系化を行った。

(1) 分類観点について

上位-下位関係（is-a関係）の他に、全体-部分関係（has-a関係）にも着目して、対象を概念化する際の視点を階層的な木構造形式にまとめる。なお、図1に

示すように、意味属性間の関連がis-a関係かhas-a関係かを示す識別子を木構造の枝に付与することにより、分類観点を明確にする。

階層的な木構造を基本構成とすることにより、以下の利点がある。

- 1) 用言の文型パターンにおける名詞の意味制約条件を記述する際、必要に応じて下位の意味属性を用いて細かく記述したり、上位の意味属性を用いて荒く記述できるようになる。
- 2) 任意の意味属性のすぐ上位にある意味属性が高々1個しか存在しないことにより、意味属性体系の効率的な探索が可能となる。
- 3) 上位の意味属性(ノード)の性質(属性)を下位の意味属性(ノード)に伝搬・継承できるようになり、下位の意味属性を定義するための記述量を削減できる。
- 4) 上位の基本的な分類体系を崩すことなく、必要に応じて最下位の分類を細分化することにより、階層的な意味属性を拡張できる。

(2) 固有名詞の扱い

種類も多様で語数も多い固有名詞を含む複合名詞の解析などでは、固有名詞について、一般名詞意味属性より細かい精度の意味属性の分解能が必要になるため、部分的に細分化した別の意味属性体系とする。

なお、一般名詞意味属性により名詞の意味制約条件を記述した用言の文型パターン対を用いて、格要素に固有名詞を含む文の意味解析を可能とするため、固有名詞にも固有名詞意味属性に対応する一般名詞意味属性を付与する。

(3) 多義の扱い

多義語には、複数の意味属性を付与する。

例1) 「木」には、『樹木』<『植物』と『木材』<『人工物』の2つの意味属性を付与する。

例2) 「犬」には、『獣』<『動物』と『スパイ』<『人』の2つの意味属性を付与する。

なお、多義語などに複数の意味属性を付与した場合、

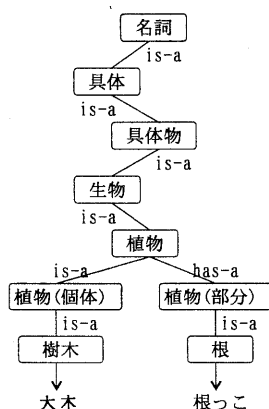


図1. 分類観点の明確化

基本的な意味属性と考えられるものほど前方に配列し、派生的な意味属性と考えられるものほど後方に配列する。上記の例1では、『樹木』、『木材』の順に意味属性を配列し、例2では、『獣』、『スパイ』の順に意味属性を配列する。

(4) 種々の対象の見方、捉え方の扱い

その語義で表される対象概念に種々の見方、捉え方があるものには、複数の意味属性を付与する。

例1) 「本」には、「物体」と「内容(情報)」の2つの捉え方があるので、『本(物体)』<『人工物』と『本(内容)』<『抽象物』の2つの意味属性を付与する。

例2) 「計画」には『思考活動』と「思考活動の結果生み出される情報」の2つの捉え方があるので、『計画』<『人間活動』と『案』<『抽象物』の2つの意味属性を付与する。

例3) 「水上飛行機」には、「飛行機」と「船」の両方の機能があるので、「飛行機」と「船」に付与される意味属性を付与する。

(5) 用言や用言性名詞の扱い

用言は、それが名詞化した場合の意味属性に写像される。動詞および動詞から名詞に転生したサ変動詞型名詞や連用形名詞については、動作を表すものは『事』の下位の意味属性を付与し、状態や関連を表すものは『抽象的關係』などの下位の意味属性を付与する。

例1) {「記述」、「書き」、「書く」…}

<『書き(一般)』…<『人間活動』<『事』

例2) {「類似」、「似」、「似る」…}

<『類似』…<『関連』<『抽象的關係』

形容詞および形容詞から名詞に転生した「形容詞語幹+さ、み…」型の名詞やいわゆる形容動詞語幹については、このような語の表わしている属性値に対応する属性(『性質』『状態』など)を示す意味属性を付与する。

例3) {「賢い」、「賢さ」、「有能」…}

<『能力』…<『性質』

例4) {「広い」、「広さ」、「広大」…}

<『度量衡』…<『数量』

(6) 並列語・縮退語の扱い

異なった意味属性をもつ単語が並列または縮退して構成された並列語・縮退語には、原則として並列または縮退された複数の単語の共通の直近上位の意味属性を並列語、縮退語の意味属性とする。

例1) 並列語「鳥獣」は『鳥』と『獣』の共通の直近上位の意味属性『動物(个体)』を意味属性として付与する。

例2) 縮退語「出入国」は「出国」と「入国」の共通の直近上位の意味属性『出・入り』を意味属性として付与する。

(7) 分類項目である意味属性の記述

階層的な木構造で構成された名詞意味属性体系のノードにあたる意味属性の名は、概念化の視点を表すのに最も適切と思われる単語(名詞)を組合せて表現する。なお、通常の単語は多義性があるのに対して、意味属性を表す単語は、一語一義で使用する。

3. 意味属性体系の構成法

本節では、2節の方針に基づき構築した意味属性体系の基本構成と一般名詞と固有名詞の意味属性体系の具体的構成について述べる。

3.1 意味属性体系の基本構成

意味属性体系は、意味属性をノードとする木構造で構成され、各ノードには概念化の視点を表すのに適切な単語(名詞)を組合せて表現した意味属性名(ノード名)とそれに対応するノード番号を付与した。

ノード名の記述においては、以下の記号を導入することにより簡略化を図った。

- 1) ・ : 並列 (or) を表示 例) おじ・おば
- 2) / : 列挙 (and) を表示 例) 自称 (単数/男)
- 3) () : 補足, 制約を付加 例) 兄弟 (年上)
- 4) < > : 分類観点を表示 例) 人間<職業>
- 5) [] : 読みを補う 例) 風 [ふう]

ノード番号は、図2の例に示すように、縦型探索法によってルートノード(ノード番号:1)から木構造探索する際に辿るノードの順に通番を付与した。このようなノード番号を付与することにより、任意のノード(ノード番号:n)に対して、その親ノードのノード番号 n_p と子孫ノードのノード番号の最大値 n_{cmax} を保持していれば、子孫ノードのノード番号が $(n+1, n+2, \dots, n_{cmax})$ であることが直ぐに分かるため、構文意味辞書の用言の文型パターンにおける名詞の意味制約条件などのチェックを簡単に行える。

3.2 一般名詞意味属性体系の具体的構成

一般名詞(固有名詞以外の名詞)を、外部世界に実在する実体(『具体』)と人間の頭の中に観念として存在する実体(『抽象』)に大きく二分した。具体は、また人間活動の主体となるもの(『主体』:人間の他に、人間の集合体としての組織、疑似人間<準人間>

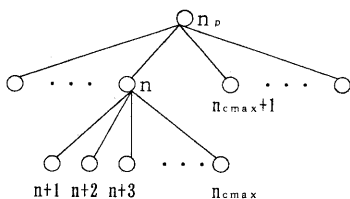


図2. 単語意味属性体系のノード番号

としての神仏などを含む)、人間活動の具体的な場となるもの(『場所』)、人間活動の対象となるもの(『具体物』)に三分した。抽象は、また人間活動の対象となるもの(『抽象物』)、動的属性を固定的に実体化して捉えたもの(『事』)、実体や属性間の種々の関係や静的属性を固定的に実体化して捉えたもの(『抽象的關係』)に三分した。

さらに、上記の『主体』、『場所』、『具体物』、『抽象物』、『事』、『抽象的關係』は、以下のように細分化した。

『主体』:『人』(神仏など準人間を含む)とその集合体である『組織』に細分化した。さらに、『人間』は『人間<人称>』、『人間<性別>』、『人間<老若>』、『人間<身体的状況>』、『人間<親族関係>』、『人間<対人関係>』、『人間<社会階層>』、『民族・人種』、『国籍』、『人間<能力>』、『人間<性向>』、『人<職業>』、『人<地位>』、『人<役割>』などの様々な観点から細分化した。

『場所』:人間がある利用目的のため自然界の場や人工的な領域に構築した建造物で人間活動の場としての役割を果たすもの(『施設』)、『領土』、『都市』、『村落』、『地域(範囲)』など人工的な領域(『地域』)、『陸地』、『山』、『川』、『海』、『宇宙』などの自然界の場(『自然』)に細分化した。

『具体物』:生命があるか否かにより『生物』と『無生物』に細分化した。さらに、『生物』は『動物』と『植物』に細分化し、『無生物』は、人間活動の結果作り出されるもの(『人工物』)と人間活動とは直接に関係なく自然界に存在するもの(『自然物』)に細分化した。

『抽象物』:人間活動(精神)の結果として生じる抽象物(『抽象物(精神)』)と人間活動(行為)の結果として生じる抽象物(『抽象物(行為)』)に細分化した。

『事』:人間の『精神』や『行為』などのような種々の活動(『人間活動』)、『物象』、『気象』、『天象』、『生命現象』などのような自然界の種々の現象

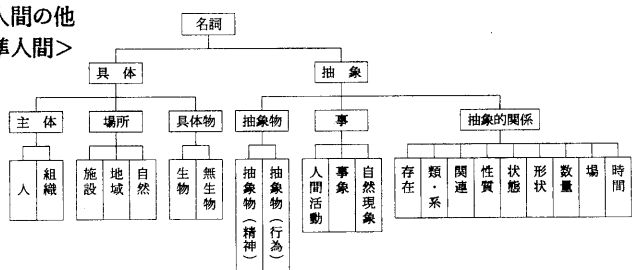


図3. 一般名詞の意味属性体系(上位4段まで)

(『自然現象』), および社会現象や抽象的関連の変化など(『事象』:『存在』,『関連』,『性質』,『状態』,『形状』,『数量』,『場』,『時間』の変化(『変動』)など)に細分化した。

『抽象的關係』:種々の『抽象的な關係』(『存在』,『類・系』,『関連』,『形状』,『数量』,『場』,『時間』)や『静的な關係』(『存在』,『関連』,『性質』,『状態』)に細分化した。

以上のようにして、図3に示すような上位分類をもち、全体で2,700余りの意味属性をもつ木構造(最大12段)で、一般名詞意味属性体系を構成した。なお、意味属性体系の具体的構築にあたっては、核となる語義について、既存の人間用のシソーラス(分類語彙表、角川類語新辞典)の分類観点を参考にした。

3.3 固有名詞意味属性体系の具体的構成

固有名詞を以下の通り分類した。

(1) 『地名』(準地名を含む)

- a) 『国際地域名』:「アジア」,「中東」などのような複数の国にまたがる領域の名
- b) 『国名』:『独立国』『植民地』(他国の属領・植民地<例:ジブラルタル>)に細分化
- c) 日本の行政区画名:都道府県レベル{『都』『道(日本)』『府』『県(日本)』}, 支庁郡レベル{『支庁』(例:北海道「渡島支庁」/東京都「小笠原支庁」)『郡(日本)』}, 市区町村レベル{『市』『特別区』(東京23区)『その他の区』(東京23区以外の政令指定都市の区)『町』(『郡(日本)』)『村』}, 字名レベル{『大字(町)』(『市』『区(日本)』で「~町」の形式のもの)『大字(その他)』(『市区町村』で「~町」以外の形式のもの)『小字』}に細分化
- d) 外国の行政区画名:『国(連邦内)』(旧ソ連邦内の15共和国,旧ユーゴスラビア内の6共和国,旧チェコスロバキア内の2共和国,ロシア連邦共和国内の自治共和国などの名)『州』(米国,ドイツなどの州<自治州,準州なども含む>の名)『省』(中国,ベトナムなどの省の名)『道(外国)』(韓国,北朝鮮などの道の名)『県(外国)』(フランス,エジプトなどの県の名)『区(外国)』(区<特別区,自治区,辺境区などを含む>の名)『郡(外国)』(郡の名)『州・省等(その他)』(上記以外の行政区画<都市や村落を除く>の名)『都市』(都市の名)『村落』(村落の名)に細分化
- e) 行政区画以外の領域の名:『地方名』(複数の行政区画にまたがる広い領域の名<例:関東/近畿/湘南>,日本の旧国名<例:武蔵/相模>など広い領域の名)『地区名』(景勝地,繁

華街,地区など狭い領域の名)に細分化

- f) 『地形名』:『陸上地形名』(大陸,島,半島,岬,地峡,山,山脈,峠,平野,盆地,峡谷,高原,丘陵,砂漠などの名)『河川湖沼名』(川,湖,沼,滝などの名)『海洋名』(大洋,海,湾,海峡などの名)『海底地形名』(海溝,海淵,海盆,海山,海底山脈,大陸棚などの名)に細分化
 - g) 『天体名』:天体,星座,星雲などの名
 - h) 『建造物名』:『建物名』(ビル,タワー,団地,アパート,工場,基地,ダム,神社,寺院,城郭,宮殿などの名)『公園名等』(公園,広場,競技場,遊園地,動物園,植物園,遺跡,古墳,墓地などの名)『農場名』(農場,牧場,養殖場などの名)『鉱山・泉等の名』(鉱山,炭田,油田,塩田,温泉,泉などの名)に細分化
 - i) 『交通路名』:『路線名』(鉄道やバスの路線,航路,航空路,道路,通り,運河,水路などの名)『交通施設名』(橋,トンネル,交差点,関などの名)『駅名等』(駅,バス停,港,空港などの名)に細分化
- #### (2) 『人名』(準人名を含む)
- a) 『姓』
 - b) 『名』:『名(男)』『名(女)』に細分化
 - c) 『有名人名』(歴史上または現代の有名人の名<架空の人物も含む>):『姓+名』(姓+名の形式の有名人名)『姓+名以外』(姓+名の形式以外有名人名<例:「老子」「卑弥呼」「明治天皇」「安芸の海」>)に細分化
 - d) 『称号』(人に関する称号<例:本因坊>):『称号(男)』『称号(女)』に細分化
 - e) 『神仏名』:『神仏名(男)』『神仏名(女)』に細分化
- #### (3) 『組織名』
- a) 『機関名』:『公共機関名』(役所,郵便局,病院,図書館,公民館,博物館,試験観測所などの名)『企業名』(会社,公社,公団,特殊法人などの名)『商店名』(商店,飲食店などの名)に細分化
 - b) 『団体・党派名』:『組合名』(労働組合,協同組合などのような組合の名)『同盟名』(同盟,連盟などのような同盟の名)『団の名』(職能別団体,教団,楽団,劇団,球団などのような団体の名)『会の名』(協会,学会,大会,委員会,審議会,連合会,クラブ,サークルなどのような会の名)『党派名』(政党,派閥などの名)に細分化
 - c) 『学校名』:『大学・高専』(大学,短大,高専などの名)『中・高校』(高校や中学校の名)

『小学校』（小学校の名）『幼稚園等』（幼稚園、保育所などの名）『学校名（その他）』（専門学校、塾、予備校などの各種学校の名）に細分化

d) 『国際組織名』：『国際機構名』（国際機関、国際機構の名）『条約名』（条約・同盟の名）に細分化

(4) その他の固有名詞（上記以外の時、事、物の名）

a) 歴史名：『年号』（年号の名）『時代名』（時代の名）『事件名』（事件、変事、事故などの名）に細分化

b) 文化名：『行事名』（祭、節句などのような行事の名）『言語名』（言語の名）『宗教名』（宗教、宗派の名）『流派名』（流派の名）『作品・出版物名』（芸術作品、新聞、雑誌、書籍などの名）『理論・方式名』（理論、方式、モデルなどの名）『法律名』（憲法、法律、政令、条例、規則などの名）『制度名』（制度の名<賞や勲章の名を含む>）に細分化

c) 『民族・人種名』：民族、人種の名

d) 愛称等：『動物名』（競走馬、ペットなどのような動物の愛称）『植物名』（固有名のついた植物の愛称）『乗り物名』（列車、船、飛行機、宇宙船などの愛称）『商品名』（商品、システム、ソフトウェアなどの名）『プロジェクト名』（プロジェクト、計画などの名）『宝物名』（固有名のついた宝物の愛称）『現象名』（風や海流などのような自然現象の名）『愛称等（その他）』（上記以外の愛称等）に細分化

e) 上記以外の固有名詞

以上のようにして、全体で130の意味属性をもつ木構造（最大9段）で、固有名詞意味属性体系を構成した。

4. 意味属性体系の構築とその検証

4.1 見出し語への意味属性の付与

意味解析用の単語辞書（単語意味辞書）には、一般語・固有名詞・専門用語など膨大で多様な語が収録され⁹⁾、それらの語に2節で述べた意味属性が付与される。ここでは、単語意味辞書の見出し語に意味属性をどのように付与したか、また付与される意味属性の一貫性をどのようにして確保について述べる。

(1) 固有名詞への意味属性の付与

固有名詞は種類も多様で語数も多いが、固有名詞の見出し語は、その種類毎に収集することとし、収集の段階で、それぞれ固有名詞意味属性を付与した。

固有名詞に一般名詞意味属性を半自動的に付与するため、固有名詞意味属性と一般名詞意味属性の対応表を作成し、それに基づき固有名詞に一般名詞意味属性

のデフォルト値を機械的に付与した。なお、機械的に付与された意味属性が正しいか否か人手でチェックし、必要ならば意味属性の修正、削除を行った。

上記により、「東京」には、{『都』『駅名等』『大学高専』『企業名』}の4つの固有名詞意味属性と{『行政区画名』『駅名等』『港』『学校名』}、4つの一般名詞意味属性が付与された。これは、固有名詞辞書項目として、実在する駅や大学などの知識を持たせたことに等しい。従って、「東京/駅」などの複合固有名詞の示す対象が実在するかどうかの判断にも使用できる。

(2) 固有名詞以外の名詞への意味属性の付与

固有名詞以外の名詞のうち、基本的な単語については、既に分類語彙表や角川類語新辞典など人間用のシソーラスで意味分類が行われた例がある。そこで、日本語意味辞書で定めた一般語に対する意味属性体系とこれらの既存の意味分類との関係を調べると、分類の考え方の違いにより、日本語意味辞書の意味属性と既存のシソーラスの分類番号は必ずしも1対1に対応しないが、1対 n ($n \geq 1$)に対応する部分が存在することが分かった。この関係を持つ単語の場合は、意味属性のデフォルト値を半自動的に付与することが期待できる。例えば、意味属性体系のある意味属性 α が、既存のシソーラスの n 個の意味分類番号 $\beta_1, \beta_2, \dots, \beta_n$ に対応するものとする、既存の辞書に収録された単語のうち分類番号 β_i ($i = 1, \dots, n$)に属する単語については、初期値として意味属性 α を付与できる。

そこで、このような関係にある単語に対しては、まず、機械的に意味属性 α を付与した後、意味属性 α が付与された単語の中で、 α 以外の意味属性を付与すべき単語を手で抽出し、これらの単語については意味属性 α を削除し、必要ならば他の意味属性を付与した。上記の関係にない基本的な単語については、はじめから人手による属性付与を行った。これらの結果を見て一つの単語に複数の意味属性が付与された場合、より基本的なものからより派生的なもの順に意味属性が付与されているか否かチェックし、必要ならば意味属性の配列を修正した。

既存のシソーラスに収録されていない名詞相当の複合語、派生語などの一般名詞意味属性については、主名詞がこれらの語の最後部の単語であるが多いことに着目して、すでに上記の方法で付与された基本語（短単位の一般語）の意味属性を参照する方法で、半自動的に付与した。以下にその手順を示す¹⁰⁾。

- 1) 意味属性を付与すべき語の後方からの最長一致法で、既に意味属性が付与された一般名詞（基本語）を収録した一般名詞（基本語）辞書を検索する。
- 2) カタカナ語を除き、意味属性を付与すべき語の2文字以上の文字列が一般名詞（基本語）辞書の見

出し語と部分一致（カタカナ語の場合単語単位で一致）すれば、当該見出し語の意味属性を、意味属性を付与すべき語の意味属性とする。

例)「自動制御」の意味属性は、後方最長一致法で検索された基本語「制御」の意味属性を付与する。

- 3) 機械的に付与された意味属性が正しいか否か人手でチェックし、必要ならば意味属性を修正・削除（後方最長一致法で検索された一般名詞<基本語>辞書の見出し語に複数の意味属性が付与されていた場合、不要な意味属性を削除）する。

上記の方法でも一般名詞意味属性が付与されない固有名詞以外の名詞については、既に一般名詞意味属性を付与された固有名詞以外の名詞を参考にして、以下のようにして人手で一般名詞意味属性を付与した。

意味属性を付与したい語Wと類義な語W'を連想する。もし、既にW'に意味属性が付与されているならば、WにW'の意味属性を付与する（W'に複数の意味属性が付与されている場合、Wの意味属性として不適当な意味属性は削除する）。もし、W'に意味属性が付与されていないならば、別な類義語を連想し、上記の処理を繰り返す。

複合名詞内で固有名詞と結合して複合固有名詞を形成する固有名詞以外の名詞や接辞（固有名詞承接語；「大学」、「銀行」、「空港」、「駅」、「港」…）には、それらに複合固有名詞を形成した場合の「固有名詞意味属性」を付与した。これにより、複合固有名詞が固有名詞承接語と結合してさらに新しい複合名詞を形成し、これを繰り返すことによりより長い複合名詞を形成する場合に対処できる¹¹⁾。

例)羽田+空港→羽田空港（固有名詞意味属性=『駅名等』）。

羽田空港+ビル→羽田空港ビル（固有名詞意味属性=『建物名』）

(3) 用言への意味属性の付与

用言（基本語）についても固有名詞以外の名詞（基本語）と同様に、既に分類語彙表や角川類語新辞典など人間用のシソーラスで体系的な意味分類がなされていることに着目し、それらをも参照しながら用言への一般名詞意味属性の付与を行った。

上記の方法でも一般名詞意味属性が付与されない用言については、一般名詞意味属性が付与されない固有名詞以外の名詞と同様に、既に一般名詞意味属性を付与された用言を参考にして、人手で一般名詞意味属性を付与した。

(4) 意味属性付与の一貫性確保

見出し語へ意味属性の付与に関しては、どのようにして付与される意味属性のばらつきをなくして一貫性を保つかが問題となった。そこで、意味属性体系の基本構成の検討を研究者1人で担当すると共に、意味

属性の付与を少人数の分業体制で行い、最終的には意味属性体系の設計者が基本語については一語一語チェックすることにより解決を図った。

4.2 意味属性体系の検証

本稿で述べた意味属性体系とこれに基づき作成した日本語意味辞書（単語意味辞書、構文意味辞書）を日英翻訳システムALT-J/E¹²⁾に組み込み、各種の翻訳実験を通じて意味属性体系の構成の適切性、その内容の良否をチェックしてきた。

種々の実験を通して、機械翻訳における構文パターンの選択に用いた場合の本意味属性体系の有効性について確認することができた^{12)、13)}、以下に示すような問題点も明らかになった。

(1) 意味属性名の定義

階層的な木構造で構成された意味属性体系のノードにあたる意味属性の名は、概念化の視点を表すのに最も適切と思われる単語（名詞）を組合せて記述されているが、本来曖昧性のある名詞では意味属性名を十分に表現しきれない。今後、計算機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する方法を検討し、意味属性体系における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

意味属性体系における分類観点は上位-下位関係、全体-部分関係であるが、これ以外に静的属性、機能、構成要素など種々の分類観点が存在する。

また、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』のように分類する観点は、現状の意味属性体系では上位-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と考えられる。さらに、語と語の関係は、その観点によって、遠くなったり、近くなったり変化する。例えば、「鳥」と「飛行機」は意味属性体系上、『動物』と『乗り物』のように比較的遠い関係にあるが、「飛ぶ（空中移動）」という動的属性に着目すると、両者は非常に近い関係にある。

今後、意味属性体系における分類の多観点化について検討し、語を種々の分類観点から分類し、観点による語と語の関係の変化を扱えるようにし、分類観点をより明確にした多次元シソーラス¹⁴⁾を構築する必要がある。

(3) 意味属性の細分化と類語間の弁別特性の付与

現状の意味属性体系では、一つの意味属性内に種々の観点から見た語が混在している。

例えば、意味属性『鳥』には、{「燕」、「雀」}などのような種としての「鳥」だけでなく、{「雄鳥」、「小鳥」/「海鳥」、「渡り鳥」}のように静的属性や動的属性からみた「鳥」などが混在している。今後、名詞句の意味解析や名詞の訳語選択などに意味属性体

系を適用し、名詞と名詞の間の意味的関連を扱えるようにするには、種々の観点から見た語が混在する意味属性を細分化して分類観点を明確化すると共に、一つの意味属性に含まれる類語間の語義や用法の差異を弁別できるような弁別特性を付与した類語弁別ネットワーク¹⁰⁾を構築する必要がある。

(4) 形容詞・形容動詞や副詞への意味属性付与に伴う意味属性体系の拡張

現状の意味属性体系では、実体(名詞)の静的属性の値(属性値)を表現する形容詞・形容動詞については、このような属性値に対応する属性を示す意味属性を付与することとしているが、一つの意味属性内に種々の属性に対応する語が混在している場合がある。例えば、意味属性「度量衡」には、{「長さ」、「面積」、「体積」、「重さ」…}のような種々の属性が対応しており、属性の観点から見たとき意味属性の粒度が大きすぎ、意味属性の細分化が必要である。また、現状の意味属性体系に組込まれていない副詞について、その意味的用法の分析を進め、どのように意味属性体系に組込むかについて検討する必要がある。

(5) 意味属性体系の拡張に伴うノード番号の拡張性

階層的木構造で構成された意味属性体系のノードにあたる意味属性の名は、人間が意味属性体系を検索・更新する際に利用する人間用の識別子であり、計算機で意味属性体系を利用する場合、意味属性にふられたノード番号を識別子として用いる。しかし、現状のノード番号は3節で述べたようにルートノードを'1'とする連番がふられており、あるノードを細分化したり、新しいノードを木構造内に新設する場合、ノード番号をどのように付与するかが問題となる。今後、意味属性体系を拡張する場合、ノード番号をどのように拡張すべきかについて検討する必要がある。

5. おわりに

対象の持つ特殊性を捨象する立場から概念化の視点(単語意味属性)を約3,000に分類した名詞の意味属性体系を提案し、日英翻訳用の単語辞書に収録された新聞記事などのような現代日本語の記述文で使用される40万語を対象に意味属性を付与し、意味属性体系と単語意味辞書を構築した。新聞記事文は現代日本語の記述文の代表的なものの一つであり、本意味属性体系の構築にあたって、現代日本語の記述文全般に適用できる汎用性を重視した。その意味で、本意味属性体系は、人間用のシソーラスである分類語彙表や角川類語新辞典のように現代日本語の記述文に適用できる汎用性をそなえている。また、従来、例をみない種々の固有名詞を対象とした本格的な固有名詞の意味属性体系と固有名詞意味辞書を構築した。

今後、今回構築した意味属性体系について、その内

容の細部にわたる検証、文型パターン対との整合性の検証、不要なノードや細分化すべきノードの分析などを進め、意味属性体系の拡張を行っていく予定である。

[参考文献]

- 1) 国立国語研究所：分類語彙表，秀英出版(1964)
- 2) 大野，浜西：角川類語新辞典，角川書店(1981)
- 3) 田中，仁科：上位/下位関係シソーラス I S A M A P 1 の作成 [I]，情報処理学会研究報告，87-NL64-4，pp. 25-34(1987)
- 4) 田中，仁科：上位/下位関係シソーラス I S A M A P の作成 [II]，情報処理学会研究報告，87-NL64-5，pp. 35-44(1987)
- 5) 日本電子化辞書研究所：概念辞書(第2版)，EDR Technical Report, TR-012, (1989)
- 6) 日本電子化辞書研究所：概念辞書(第3版)，EDR Technical Report, TR-020, (1990)
- 7) 荻野，仲尾，長沢，小笠原：EDR電子化辞書における概念体系，言語処理学会第1回年次大会，pp. 197-200(1995)
- 8) 国立国語研究所：電子計算機による新聞の語彙調査，国立国語研究所報告，No. 37，秀英出版(1970)
- 9) 宮崎正弘：辞書の記述と利用—機械辞書の観点から—，日本語学，Vol. 14, NO. 4, pp. 52-61(1995)
- 10) 池原，白井，横尾，F. Bond, 小見：日英機械翻訳における利用者辞書の意味属性の自動推定，自然言語処理，Vol. 2, No. 1, pp. 3-17(1995)
- 11) 宮崎正弘：係り受け解析を用いた複合語の自動分割法，情報処理学会論文誌，Vol. 25, No. 6, pp. 970-979(1984)
- 12) 池原，宮崎，白井，横尾：An Evaluation Method for MT Systems and its Application to ALT-J/E, 人工知能学会誌論文，Vol. 7, No. 6, pp. 1077-1086(1992)
- 13) 池原，宮崎，横尾：日英機械翻訳のための意味解析用の知識とその分解能，情報処理学会論文誌，Vol. 34, No. 8, pp. 1692-1704(1993)
- 14) 川村，片桐，宮崎：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研究報告，NLC94-48, pp. 33-40(1995)
- 15) 中田，宮崎：類語弁別ネットワークによる語義記述法，自然言語処理の基本問題シンポジウム論文集，pp. 1-8(1992)