経済と経営 37－1(2006.11)

〈研究報告〉

普遍言語体系，概念辞書，マルチリンガル・ワードネットの研究

工 藤 孝 史／三 上 勝 生

本研究報告の筆者は，普遍言語体系，概念辞書およびマルチリンガル・ワード・ネット等に関わる研究という共通課題をもって，平成16年度にそれぞれ慶応大学SFC研究所(工藤孝史)，スタンフォード大学CSLI研究所（三上勝生）において研修した。以下はその研修成果に基づき，「普遍言語体系」，「概念辞書」および「マルチリンガル・ワードネット」に関する現段階の研究成果を共同で報告するものである。なおこの研究は，普遍的な言語の発見とその計算処理のモデル化をめざして2000年4月に発足した「たまごプロジェクト」の研究活動と相関をなすもので，はじめにそのプロジェクトの簡単な経緯を記した。

概要

(たまごプロジェクトの活動)

2000年（平成12年）に開始した本プロジェクトのこれまでの経緯はほぼ以下の通りである。

■ 2001年（平成13年）

5カ国語（日英独仏露）間の「翻訳」（中間言語を媒介した基本語彙レベルの意味変換）の実験的システムの開発を開始。

たまごプロジェクト第一回フォーラム：「機械翻訳と次世代コミュニケーション・ツール」（2002.02.21）において上記システムの基本思想を公開する。

■ 2002年（平成14年）

5カ国語の語彙と文法を照合する作業を通して，中間言語としての普遍言語体系の構築開始。普遍言語の語彙および構文の主要な範疇のたたき台となる要素を決定する。

■ 2003年（平成15年）

語彙カテゴリー分類を基礎とする，5カ国語の意味ネットワーク（多義および同義ないしは類義関係を記したネットワーク概念辞書）の構築作業を開始する。談話類型を確定するための発話行為の分析，分類に着手する。

■ 2004年（平成16年）

旅行会話等の典型的な言語表現を対象にして，発話者の「意図」を探る指標としての助動詞ないしはモードの分類・分析に着手する。

■ 2005年（平成17年）

語彙，構文に関するネットワーク概念辞書のデータ化および談話ないし発話の構造分析を柱とし

た普遍言語体系の構築を総合的に検討するとともに，マルチリンガル・ワードネット（Multilingual Word Net: MWN）の開発に着手する。

■ 2006年（平成18年）

「普遍言語体系（UL）」と「概念辞書（DC）」と「マルチリンガル・ワードネット（MWN）」の三者の理論的関連性の検討を始める。

本稿では，以下，まず「普遍言語体系（UL）」，「概念辞書（DC）」，「マルチリンガル・ワードネット（MWN）」のそれぞれについて述べ，最後に今後の研究課題と展望について述べる。

1 普遍言語体系／UL: The System of Universal Language

普遍言語体系は，エスペラント語に代表されるような国際的な自然共通言語としての「もう一つの言語」ではなく，各自然言語間の概念および概念操作レベルでの「共通性」を手がかりにして，自然言語一般に適用可能な語彙および発話＝聴話規則を体系的に記述することをめざす人工言語体系である。いままでのところ私たちは，日本語，英語，ドイツ語，フランス語，スペイン語，ロシア語の各語彙と構文を対象に，各自然言語概念の共通項およびその組み立て操作の共通項（発話＝聴話行為規則）を照合しつつ，人類が「共有している」と仮定される基本概念（普遍範疇）とその操作（普遍操作）のありかたの発見に努めてきた。また，それと並行して，この体系を「普遍範疇」要素と「普遍操作」要素からなる具体的な言語（普遍言語＝ユニバーサル・ランゲージ）として記述し，コンピュータ上で使用できるようにするために，自然言語概念およびその操作様態の数値的体系化＝コード化に努めている。

1-1 普遍範疇／UC: The Universal Categories

例えば「上」「下」のような方向認識や「のぼる」「おりる」のような運動認識は，各自然言語による表現形式の違いにもかかわらず，メタ言語的な認識＝発話＝聴話レベルでは，人類に共通していると仮定することができる。それは，人間のあいだに身体・心理・生理等の本質的な共通性を仮定できるのと同様である。そのような仮定的な概念的共通認識内容を，言語活動における対象的な範疇として分類し体系化したものが普遍範疇である。

1-2 普遍操作／UO: The Universal Operations

日本語における助詞，印欧語の語順や格，前置詞の使用等にみられるように，思考や心理を言語として表現する仕方，すなわち自然言語の陳述形式は言語体系によって異なる。しかし，われわれは言語活動の本質的＝核心的過程を上述の普遍範疇と普遍操作の協働として類型化できると考えており，それを「普遍操作」と位置づける。

2 普遍言語体系の基本構成／The components of the system of UL

普遍言語体系は，上述のように言語活動を「普遍範疇」（概念の認識）と「普遍操作」（概念の操

作）という2個の大きな精神活動の統合と見る立場で構想される。以下に基本構成を箇条的に述べる。

2-1　普遍範疇／universal categories

普遍範疇は，人間によるメタ的＝対象的な認識内容およびその概念化の範疇を類型化するものである。以下の4種を区別する。

2-1-1　動的範疇／kinetic categories

動的範疇は，自然言語における形式的な品詞類型，例えば「動詞」「用言」といった分類に基づく呼称ではなく，メタ言語的なレベルで「動的＝時間的」と考えられる概念の範疇という意味である。したがって「遊び（名詞）」も「遊ぶ（動詞）」も同様に「動的範疇」に類型化される。日本語の場合，「～する」の形をとることのできる言葉で概念化されるものは基本的には「動的範疇」に分類されるが，たとえば「電話する」における「電話」のようなものをどう扱うかは議論の余地がある。

2-1-2　静的範疇／static categories

上記の「動的範疇」にたいして，「静的＝空間的」な概念の範疇。「山」「川」といった，それ自身のうちに時間的な運動を含まない物的な対象概念の範疇。ただし，英語における「shine（動詞／名詞）」：日本語における「輝き」のように，もともとは自然現象的な活動概念を「静的＝空間的」と見るかどうかは議論の余地がある。

2-1-3　属性的範疇／attributive categories

「大きい」「小さい」「固い」「柔らかい」「寒い」「暑い」「悲しい」「嬉しい」，また「早く」「全く」といった類いの，〝もの〟ないしは〝こと〟の属性・様態を表す概念の範疇。ただし形容詞ないしは副詞という形式的品詞分類から導かれる類型化ではなく，メタ言語的なレベルで「属性・様態的」と規定される概念の範疇。したがって「大きさ（名詞）」「深さ（名詞）」という日本語で表現される認識内容およびその概念化も普遍範疇の上では「属性的範疇」である。日本語においては「～く」「～くなる」の形をとることのできる言葉で概念化されるものは基本的には「属性的範疇」に分類されるが，例えば日本語における「暗闇」「悲哀」のような言葉についても，これを「属性的」と考えるかどうかは議論の余地がある。

2-1-4　指示的範疇／demonstrative categories

「この」「あの」「これ」「あれ」「いくつかの」「なんらかの」「ある」といった類いの，あるものを「指示」ないしは「一般的に限定」する概念の範疇。ただし指示形容詞，指示代名詞という形式的品詞分類から導かれる類型化ではなく，メタ言語的なレベルで「指示・限定的」と規定される概念の範疇。また，いわゆる人称代名詞（私，かれ……など）も基本的には「指示的範疇」に分類されるものの，これについては，発話＝聴話の主体としての本質的〝私〟と区別して考えなければならない等の問題がある。

2-2　普遍操作／universal operations

普遍操作は，人間の発話＝聴話行動において，言語主体（話者＝聴者）が自らの思考や心理を普遍範疇およびその概念化の操作（概念の主体的連結操作ないしは変形操作）をもって表現する場合の規則を類型化するものである。

2-2-1 格操作／case operators

日本語における「が（は）にを」等，印欧語における主格，与格，体格などによって表現される行為的論理認識，あるいは日本語における繋辞的な「ある」，印欧語のbe動詞相当語（copula）によって表現される存在的論理認識またはその表現規則を類型化する。言語主体（話者＝聴者）が特定する普遍範疇がどのような論理関係によって概念化（理解）されるかを，行為者格（performer-case），被行為者格（patient-case），時間／空間格（time/space-case）を基本とし，その派生的な諸格操作を分類する。なお格操作の命名法および派生的諸格操作については今後多くの議論が必要であると思われるが，格操作はあくまでも言語主体（話者＝聴者）によって理解されている「言語表現：文なりメッセージ」の論理構造を担うものであって，たとえば行為者格が話者としての言語主体を意味するといった意味ではない。「空がきれいだ」「彼女はきれいだ」「私きれい？」における「空」「彼女」「私」はいずれも言語主体が普遍範疇（空，きれい，彼女，私）を論理的に構成する際に現れる，文ないしはメッセージ内的行為者である。たとえば{「私きれい？」と彼女は私に聞くのだった}という文もしくはメッセージにおける「私」は，前者が行為者格（言語主体から見れば「彼女」），後者が被行為者格（言語主体から見れば「私」）である。

2-2-2 限定操作／determinate operators

言語主体が，普遍範疇に属する諸概念同士を組み合わせる（連携させる）ことで対象を特定する（たとえば前述「私の本（私＋本）」「my book（I＋book）」）場合の概念の操作を類型化する。この操作は，本質的には前項の「格操作」と同様に，基本的には普遍範疇の同種間または異種間の組み合わせ（連携）として言語主体が発動＝受動する論理的認識（理解）であるが，文またはメッセージを行為的論理関係もしくは存在的論理関係の面から理解する連携ではなく，特定すべき対象の限定的＝修飾的論理関係の面から理解した操作である。「美しい日曜日」「開かれた窓」「去り行く日々」「マッチ箱」「その年」「いくつかの問題」……等々，一般には「修飾」や「指示（限定）」「同列」などの呼びかたで扱われている概念と概念の連携操作である。

2-2-3 様相操作／modal operators

言語主体が文やメッセージを発動＝受動するさいに，語られている内容が展開する「場」の背景を時間的，心理的，社会的に特定しようとする操作。言語主体の発話＝聴話経験，つまり文やメッセージの発動＝受動は基本的には一回的＝現在的なものであるが，語られる内容は必ずしも一回的＝現在的なものではない。「私は宇宙人を見た」は「私」の経験を今のものとして表明する。「その頃はよく彼女とリュクサンブール公園を散歩したものだ」も同様に，繰り返された経験を今のものとして表明する。また「3650年に宇宙船フジの船長は火星人と商談した」においても，言語主体はこれを今のものとして表明している。このように，文またはメッセージは，これが発動＝受動される時点の一回性に支配されているが，一般に「時制：tens」「法：mood」「態：voice」「相：aspect」などの呼称のもとに扱われている表現操作を通して，言語主体は自らの語る内容の展開される時点を特定したり，展開の仕方（たとえば「明子は娘の手をつないで横断歩道を渡っていた」と「明子の娘は母に手をつながれて横断歩道を渡っていた」はともに「明子」とその「娘」の両方が横断歩道を渡っている様子を表明するものだが，展開の仕方は違う），さらには語る際の心理的態度や社会的力関係に対応した語り口を特定するなどといったことを行う。また「彼は来る」「彼は来るだろう」「彼は来るかもしれない」「彼は来るはずだ」「彼は来てしまった」「彼は来ている」……などの表現

は，単純に時制の問題でもなければ法の問題でもなく，言語主体が自らの表現を操作する際の様相的な価値判断と見ることができる。これらは「彼」の気持ちや考えを表現しているのではなく，言語主体の言語表現的価値判断を表すもので，「様相操作」はそうした言語主体の言語発動上の価値観を類型化するものである。

3　概念辞書／DC: Dictionary of Concepts

異なる自然言語の語句同士は意味の上で単純に１対１対応していない。言語体系が違えば，その語句が指示できる対象範囲も，その言語体系内での品詞的価値にも違いがある。私たちは，そのような差異の現実にもかかわらず，語彙の上でも発話＝聴話規則の上でも，この差異を共有カテゴリー集合の要素として扱うことができるのではないかと考える。つまり，個々の具体的自然言語語句の意味（単位）および構文規則を包括できるカテゴリー群を「普遍範疇」および「普遍操作」として理論化できるのではないかという展望である。そのためには，自然言語間で共有可能な本質的かつ一般的な言語体系が必要であるが，この体系については「普遍言語体系」として概略を上に述べた。ただしそのような範疇体系は実体的でも絶対的でもなく，あくまでも関係的かつ相対的な「価値」として定義される必要がある。この関係的かつ相対的な「価値」をメタ言語レベルで記述するのが概念辞書である。

私たちは，これまでに日本語，英語，ドイツ語，フランス語，ロシア語，スペイン語の６カ国語を対象として，語句のメタ言語的意味（概念的定義）を記述する作業を続けてきた。具体的には，概念辞書の基本となる500あまりの運動概念（前記2-1-1動的範疇に対応）と1,200あまりの静止，属性，指示概念（前記2-1-2静的範疇，2-1-3属性的範疇，2-1-4指示的範疇に対応）をとりあげ，これをメタ言語レベルで同定する作業を行った。「メタ言語」記述としては，いくつかの例外はあるものの，いまのところ各国語の代表的な国語辞書の記述をそのまま使用している。つまりそれぞれの「国語辞典」が記述する見出し語句のメタ記述（語句の意味説明）を，概念的な価値の面で照合＝評価して，共有意味区分を確立する作業である。但しそれぞれの「国語辞典」は語句のメタ記述（概念の定義）において必ずしも共通の規則をもっていないので，今後，複数の言語に共通して使えるメタ記述の方法と規則（メタ言語的記述法）の確立が必須である。また，前記「1-2普遍操作」に分類される顕在的な語句（日本語の助詞や印欧語の前置詞等）についても，メタ言語レベルで同定を試みているが，この作業は前者（普遍範疇）と同様な方法ではうまくいかない。たとえば印欧語における「語順」と日本語における「格助詞」の価値的共有性の問題は，顕在的な語句レベルだけで同定することは不可能で，方法論的にはたとえば「生成変形文法」のようにこれを本質的なレベルに還元して「共通」範疇を分類しなければならないが，その具体的＝内容的展開につては，今後の研究課題としている。

4　マルチリンガル・ワードネット／MWN: Multilingual WordNet

マルチリンガル・ワードネット（MWN）とは，既存のワードネット（WordNet：一言語内の語

彙ネットワーク）の考え方を多言語間のメタネットワークに拡張するものである。私たちは，実験的な試みとして上述の「概念辞書」を土台に，複数の言語体系にまたがる語句の意味対応関係のデータベースを構築中である。ここではマルチリンガル・ワードネットの基本的な構成要素（データベース・レコードのおもなフィールド）について，ごく簡単に解説する。

4-1 見出し語／Index

自然言語の語句が登録される。登録される語は一語でなくてもよい（例えば，英語の「go out」，日本語の「通り過ぎる」のような，本来「通る」＋「過ぎる」の合成と考えられるような単語，ドイツ語やフランス語に見られる「再帰動詞／代名動詞」など）。英語の「good morning」やフランス語の「s'il vous plaît」のような慣用表現も登録可能である。その他，ひとまとまりの意味をもつと考えられる語群は単語として扱われる。最終的には，この「ひとまとまり」という意味単位を，何を基準にどう確定するかという最重要課題を解決しなければならいが，現在はいくつかの実験的試みをくり返しながら，この意味単位を模索している段階である。方法論としては，あらかじめ抽象的に「単語」つまり意味解析の最小単位となる要素を定義するのではなく，各自然言語がもつ多様な意味的現実を精査する作業を通じて，帰納的に意味単位を確定する方法をとっている。現在のところ，日本語（JP），英語（EG），ドイツ語（GM），フランス語（FR），ロシア語（RS），スペイン語（SP）について，それぞれの国語辞典が見出し語としてリストアップしている「単語」が，基本的にはそのまま見出し語として登録されているが，一部例外もある。なお現段階の見出し語数は各言語について基本単語3,000～5,000語程度である。

4-2 多義（ポリセミ）／Polisemi

それぞれの「国語辞典」を見て明らかなように，自然言語の単語は通常複数のメタ言語的意味（抽象的意味）をもつ（とくに動作概念を表す単語の場合その数が多くなる）。同一の見出し語（Index）は，発話内容に応じてそれぞれ別の現実的意味を獲得するが，その各々を多義として登録する。たとえば「行く」という単語（見出し語）は，「学校へ行く」という場合と「うまく行く」という場合とでは，現実的意味（辞書が「見出し」としている抽象的意味に対立する）の上での相違がある。多義はこの相違を表したもので「見出し語＋数字」（「行く1」「行く2」）の形で登録される。見出し語に多義が認められない（メタ言語的定義がひとつしかない）場合も，その意味は便宜上「多義1」として登録される。したがって，MWNデータベースは多義項目1件につき1レコードを形成することになり，これがMWNデータベースの実質的な見出し語＝語彙単位となる。国語辞典における「見出し語」は，それが使用されるケース（語結合）によって現実的意味を変化させる以上，これは，あくまでも複数の現実的意味を束ねている「指標」（抽象的意味）に過ぎず，実質的な語彙単位（現実的意味）としては扱われないということである。

4-3 説明文／Explanation

多義には説明文が登録される。これは語彙単位のメタ記述であるが，現在のところその記述は，前述のように一般的辞書にある説明すなわち定義を参考にして，各見出し語と同一の自然言語で記述される。上述の通り（3概念辞書）今後このメタ記述の記述法の確立は必須である。

4-4　例文／Illustrative Sentence

多義は，他の多義との現実的結合ないしは発話によって現実的意味を獲得する。一般的辞書ではそれを用例で示している。現在のMWNデータベースではそれらが引用され多義の例文として登録される。例えば，「買う」には「車を」につながる用例と「顰蹙（ヒンシュク）を」につながる用例とがある。

4-5　メタ記述／Meta-Description

現在のところは，各国語の代表的辞書の説明文（抽象的意味の定義）とその例文によって記述される。メタ記述法が確立されれば，複数の言語体系に対して同一の記述が対応することになるであろう。

4-6　類義語／Synonym

多義のひとつひとつについて，それと同じような意味に使われる（同じ言語体系内の）別の単語すなわち「類義語」が登録される。多義の類義語として登録される単語は以下の２種にわけられる。

4-6-1　相関類義語（類義語１種）／correlative synonym (synonym 1st group)

ある見出し語の多義（polisemi）が他の見出し語（index）である場合には，これを相関類義語として登録。例えば，「行く２」が，見出し語「来る」を類義語として持っている（「君も行く？」が「君も来る？」と同じ意味を表す）場合，「来る」（index）は「行く１」（polisemi）の相関的類義語である。

4-6-2　付帯類義語（類義語２種）／incidental synonym (synonym 2nd group)

ある見出し語の多義（polisemi）が類義語（synomym）をもっているけれども，それが見出し語（index）として登録されていない場合には，付帯類義語として登録される。例えば，「行く１」が，見出し語としては登録されていない「赴く」を類義語としてもつ場合，「赴く」は「行く１」の付帯類義語である。相関類義語は，ある見出し語に束ねられている現実的意味（多義）のあるものが，別の見出し語に束ねられている現実的意味（多義）のあるものと「類義」であることを示すものであるのに対して，付帯類義語は，ある見出し語に束ねられている現実的意味（多義）が，それと同じ意味で使われる別単語をもつということのみを示すもので，常に「一義」，すなわち関連する多義と同一の意味として扱われる。

5　今後の課題および展望

[三上勝生]

やや見通しにくいかもしれない普遍言語体系，概念辞書そしてマルチリンガル・ワードネットの三者の理論的関係について簡単に説明しておきたい。

先ず，概念辞書とは，普遍言語体系の中の普遍範疇（カテゴリー）に基づいて六カ国語の基本的語彙を記述（コード化）したものである。次に，マルチリンガル・ワードネットとは，概念辞書に基づいて六カ国語間に類義性と多義性のリンク（関係）を張ったもの，つまり多言語間意味ネット

ワークである。

そもそも普遍言語体系は，各言語に関する様々な知見と経験を総動員することを通じてあくまで経験的に構築されるものである。したがって，その「普遍性」は相対的なものにすぎない。しかし，各自然言語の実態に照らし合わせながら概念辞書を，さらに多言語間の意味ネットワークを制作する過程において得られる知見は，普遍言語体系の構成にフィードバックされ，体系は再構成される。このような終わりない循環的な作業を通じて，はじめて普遍言語体系はそう呼ばれるに相応しいものへと少しずつ進化していく。われわれは決して何らかの実体ないしは本質を「普遍」として先取りしているのではない。不可避の循環的作業を通じて「普遍」を浮き彫りにしようとしているのである。

[工藤孝史]

三上勝生氏はスタンフォード大学において，おもに人間の言語活動における「言表」や「論理」という問題を，哲学や論理学あるいは広く文学の立場からも研究されたようであるが，今回その三上氏との意見交換の上に，現時点での普遍言語に関する研究成果の概略を共同でまとめたのがこの研究報告である。当然にも，上記各項目の基本的考え方には，2000年4月に私も含め札幌大学内外の研究者らによって「たまごプロジェクト」が発足して以来私自身が「普遍言語」のありかたについて考察してきた内容が反映されている。これらの各項目については，それを詳細に論じる必要があるが，それについては稿を改めてそれぞれの立場で逐次発表して行きたいと考えている。

私は，三上氏がスタンフォード大学において研究されているのと同時期に，慶応大学藤沢キャンパス（SFC研究所）において，多言語の意味ネットワークを「概念辞書」の土台の上にどう築きあげられるかという問題について，それを自然言語処理の立場から研究していたが，その過程でつねに問題となったのは「そもそも人間の言語活動を支えている，世界の〝概念的な〟理解またはその言動化にはヒトとしての共通性を確認できるのか」という問題であった。それは「疑念」でもありまた一方，多分そのようなものがあるはずだという「予測」でもあった。現時点においても，この「共通性」についていえば，その実在を確定することはできていないし「疑念」と「予測」は一種のアンビバレンツとして私の中にある。しかし私は少なくとも，形式を異にする自然言語体系の使用者同士が接触した時，それぞれの言語使用者は，相手の言葉を自分の言葉に「置き換えて」理解しようとする傾向のあることを経験的に知らされてきた。この「置き換え」がどうして可能なのかに注目した時に，そこにはヒトという存在に共通する概念形成や概念操作があるのではないかという予測を立てたのである。本報告はそうした概念把握や概念操作のありかたを「普遍言語」というキイワードに即して展開したものである。

言語的現実は，自然科学の対象となりうる「物質」とは違って，そもそもひとつの「精神」的現実（もっとも音声や文字はそれ自体をとって見れば「物質」にすぎないが）であって，これを扱うには，何よりも言語活動の現実を「身体的」であると同時に「心的」である（いいかえれば「物質」と「精神」の絶え間ない交渉によって条件づけられた）人間存在そのものありかたに遡って考える

ことが不可欠である。その意味で私にとって「普遍言語」の研究は，「心身問題」研究の一部門であるのだが，今回ここにまとめた各議論につては，今後テーマを改めて展開したい。

【本稿は一部平成16（2004）年度国外留学研修の成果（三上勝生）および国内留学研修の成果（工藤孝史）に基づく】