

意味類型構築のための文接続表現の体系化

衛藤純司*1 池原悟*2 佐良木昌*3 宮崎正弘*4
池田尚志*5 新田義彦*6 白井諭*7 柴田勝征*8

*1 (有)ランゲージウェア etoh@titan.ocn.ne.jp *2 鳥取大学工学部情報工学科 ikehara@ike.tottori-u.ac.jp
*3 長崎純心大学 saraki@st.rim.or.jp *4 新潟大学工学部情報工学科 miyazaki@ie.niigata-u.ac.jp
*5 岐阜大学工学部応用情報学科 ikeda@info.gifu-u.ac.jp *6 日本大学経済学部・理工学部 nitta@eco.nihon-u.ac.jp
*7 NTT-アドバンステクノロジー株式会社 shirai@nlp.ntt-at.co.jp *8 福岡大学理学部応用数学科 kshibata@vin.sm.fukuoka-u.ac.jp

「意味類型論(セマンティック・タイポロジー)」と「類推思考」の2つの原理に基づいた機械翻訳システムを構築している。これは(1)原言語と目的言語の表現から単純な重ね合わせでは対応づけられない非線形な表現構造を文型パターンとして抽出し、それを意味的な同等性にしたがって類型化すること、(2)意味的に類型化された文型パターンを「類推思考の原理」によって対応づけること、の2つの仕組みから構成される。本稿では、このような意味類型の考え方にしたがって重文(接続表現で結ばれた前後二つの節からなる)の意味を把握することを提案する。そして、二つの節の間の論理関係を表す接続表現の体系化を試みたので説明する。

キーワード: 機械翻訳、意味類型、類推思考、重文、論理関係、接続表現

Semantic Classification of Conjunctive Expressions based on Semantic Typology

Junji Etoh*1 Satoru Ikehara*2 Masashi Saraki*3 Masahiro Miyazaki*4
Takashi Ikeda*5 Yoshihiko Nitta*6 Satoshi Shirai*7 Katsumasa Shibata*8

*1 Languageware Inc. *2 Faculty of Engineering, Tottori University
*3 Nagasaki Junshin Catholic University *4 Faculty of Engineering, Niigata University
*5 Faculty of Engineering, Gifu University
*6 College of Economics, College of Science and Technology, Nihon University
*7 Advanced Technology Co.Ltd. *8 Faculty of Science, Fukuoka University

This paper describes the application of the Theory of Semantic Typology(TST) to the analysis of clause relations in complex and compound sentences. TST is one of the two principles of the new machine translation system under construction, through which semantically non-linear expressions can be extracted from the huge bilingual corpus and patternized by meaning. Another principle is the Analogical Mapping Theory(AMT), through which a set of meaning patterns of a source language is mapped to that of target language via common concept to both. Two principles set forth have been proposed by the special interest group of "Language, Cognition and Expression".

keyword: Machine Translation, Semantic Typology, Analogical mapping, Complex Sentences, Logical Relation

1 はじめに

従来の機械翻訳システムは、ほとんどがトランスファー方式で行われてきた。この方式は、構文解析結果を用いて原言語の表現構造を目的言語の構造に変換したのち、原言語で使われた語や句を目的言語のそれに置き換えることを基本としている。文の構造は意味と独立に扱われるが、対応する語や句を重ね合わせていけば、その結果として同じ意味を表す文を合成することができるかと仮定している。

しかし、実際の言語表現では、このような語句の重ね合わせで意味を保存できるとは限らない。とりわけ、日本語と英語のような語族の異なる言語間では構造と意味とのずれは大きく、日本語らしい日本語、英語らしい英語に翻訳しようとするれば、むしろ、そのようなずれをこそ積極的に活用しようとするだろう。

そこで、語や句の単なる重ね合わせではない、文の構造と意味を一体のものとして扱う仕組みが求められる。そのような仕組みの実現を目指したものとして多段翻訳方式[池原 87]の研究がある。この方式は、原言語から意味のまとまりを表現する構造を取り出し、それを「構造的な意味の単位」として目的言語に変換することを目指している。「構造的な意味の単位」としては、用言、および、それと共起する名詞の関係に着目して、これを結合価パターン¹の形式で記述している[池原 97]。

本稿で提案する「意味類型を用いた類推による機械翻訳」方式は、この研究の延長線上にあるもので、(1)原言語と目的言語の表現から単純な重ね合わせでは対応づけられない非線形な表現構造を文型パターンとして抽出し、それを意味的な同等性にしがたって類型化すること、(2)意味的に類型化された文型パターンを「類推思考の原理」[市川 63]によって対応づけること、の2つの仕組みから構成される。

この方式を実現するには、(1)原言語と目的言語に対して網羅性の高い意味類型パターン集(文型パターンを意味類型化したもの)を構築

すること、および、(2)各パターン間を意味的等価性に基づいてマッピングするための論理的意味範疇を体系化することが必要である。

2 文型パターン

文型パターンとは、原言語と目的言語の対訳文から、単純な重ね合わせでは対応づけられない非線形な表現構造を取り出したものである。よい翻訳を得るには原言語と目的言語の実状に即した個別的な文型パターンが望ましいが、より広範な文を翻訳できるようにするにはなるべく汎用的な文型パターンが望ましい。そこで、個性性と汎用性を同時に実現するために、対訳コーパスから単語レベル、句レベル、節レベルの3つのレベルの文型パターンを作成する。

3種の文型パターンでは、句レベルの文型パターンを標準とする。これと、個別的な単語レベルの文型パターンと汎用的な節レベルの文型パターンを組み合わせることによって、個性性と汎用性の両立を図る。

(1) 単語レベルの汎化例

和文：彼らは日本に慣れない外国人と見ると
こないかさま物売りつける。

英文：When they have to do with foreigners
new to Japan, they foist such fakes
upon them.

和文 P：N1 は[N2 に慣れない]N3 と見ると[こ
んな]N4 を売りつける。

英文 P：When N1 have to do with N3 [new to
N2], N1 foist [such] N4 upon
N3_pro.

(2) 句レベルの汎化例

和文：昨日出した手紙が宛名不明で返って
きた。

英文：The letter I mailed yesterday came
back marked addressee unknown.

和文 P：[N1 が]VP2.ta NP3 が Adj4 で
V5.tekita

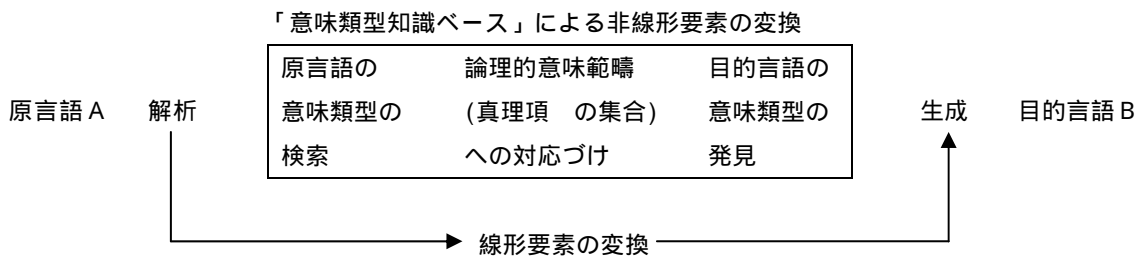


図 1 言語の等価変換方式の構成

英文 P : NP3 [that] N1 VP2.ed V5.ed
VP.en(Adj4)

(3) 節レベルの汎化例

和文：いつまで親の厄介になっているかと思
うと肩身が狭い。

英文：I feel small when I think I am still
dependent on my father.

和文 P : いつまで CL1 かと思うと AP2

英文 P : I feel AP2 when I think CI1

註：N, V, NP, VP, Adj, AP, CL は、それぞ
れ名詞(句)、動詞(句)、形容詞(句)
節を表す。

_pro は代名詞化、.tekita .ed .ed など
は活用を表す。

数字は日本語と英語とで対応する語句を
示す。

3 意味類型

言語は、話者が対象を認識し、そこから概念
を取り出し、それを語のある配列で表現したも
のである。対象を認識する仕方、概念を取り出
す仕方、そして言語で表現する仕方、それぞ
れに、文化・社会による違いもあれば、違いを超
えた共通性もある。

「意味類型」は、認識と概念と表現のこのよ
うな差異性と共通性を背景にして考えられたも
ので、言語表現のパターンを意味の同等性に注
目して類型化したものである。すなわち、原言
語と目的言語の対訳文から上述のような手順で
取り出した文型パターンを、例えば、因果関係

を表すもの、比較を表すもの、対比の関係を表
すもの、等々に分類する。もし望むならば、ど
のようなタイプの因果関係か、比較は優劣の比
較か同等の比較か、対比される二つの事象はそ
れぞれどのようなタイプの事象か、というよう
により詳細な観点から分類することもできよう。
このようにして得られた、意味と表現の一体化
した知識構造を「意味類型」という[有田 87]。

意味と表現が一体化したものであるから、そ
れぞれの言語に個別的である。日本語には日本
語の「意味類型」があり、英語には英語の「意
味類型」がある。したがって、「意味類型」は、
すべての言語に共通な意味構造があると考え、
その意味構造を記述しようとするいわゆる「中
間言語」とは違う。「中間言語」はあくまで言語
の「ユートピア」であって実在するかどうか疑
わしいものであるが、「意味類型」は実際の言語
表現から上述のように明確に規定された手順に
したがって取り出されたものである。

和文：先週風邪を引いたばかりなのに、今週
また風邪を引きました。

英文：I just had a cold last week, but I've
got one again this week.

意味類型：前件(過去のネガティブな身体変
化の完了)/論理関係(因果・逆接・
意外)/後件(現在のネガティブな
身体変化の反復)

4 論理的意思範疇と真理項

一般に、個々の文の意味類型は、複数の概念

が組み合わせられた複合的な構造を持つ。上記の例でいえば、「過去」「現在」「ネガティブ」「身体変化」「完了」「反復」「因果」「逆接」「意外」等々といった概念が組み合わさって、文全体の意味を表している。このような、意味類型を構成する概念の体系を「論理的意味範疇」と呼び、その要素である個々の単一な概念を「真理項」と呼ぶことにする。

本方式を実現するには、論理的意味範疇の体系をどのようにデザインするか、また、真理項の集合をどのように用意しておくかが、きわめて重要である。そこに求められるのは次の3点である。

- (1) 網羅性：ある限定されたジャンルだけでなく、多種多様な言語表現に対して記述可能であること。
- (2) 体系性：一定の秩序を持ち、新しい表現・意味をその秩序の中に容易に包括しうること。
- (3) 記述容易性：具体的な言語表現と結びつき、誰でも一定の意味類型を記述できること。

特に、体系性について言えば、人間の認知・判断作用の背後にある構造と照応するようなものであることが必要である。

5 重文の意味類型

日英機械翻訳では、「日本語語彙大系」を用いることによって単文レベルの翻訳の品質は大幅に向上してきたのに対して、重文や複文の翻訳はとても満足のいくレベルに達してはいないというのが現状である。そこで、本方式ではこれら重文や複文に対して高い品質の翻訳を行うことを主要な目標にしている。

本稿ではそのうち、重文の意味類型について説明する。重文は、接続表現で結ばれた前後二つの節からなる。そこで、前節と後節それぞれの意味と、それらを結ぶ接続表現の意味との総

合として重文の意味を記述する。

例：先週風邪を引いたばかりなのに、今週末風邪を引いた。

意味類型：過去のネガティブな身体変化の完了/因果・逆接・意外/現在のネガティブな身体変化の反復

接続表現：因果・逆接・意外（のに）

前節：身体変化/ネガティブ（風邪を引く）

過去（先週）

完了（た）

後節：身体変化/ネガティブ（風邪を引く）

現在（今週）

反復（また）

5.1 接続表現の意味

ここで接続表現というのは、いわゆる接続助詞だけでなく、ある種の形式的な語句を伴う接続助詞相当語句とでもいうべき表現を含んでいる。

接続表現の意味を網羅的につかむために、次のような手順を踏んだ。すなわち、電子化された日英・英日辞典の例文対訳コーパス約100万文から150,753文の重文を抽出した。その文節構造を解析して[岸井03]、246の接続表現を得た。以下にその一部を挙げる。

あげくに、いじょう、おりに、が、かいなか、かぎり、かたわら、かどうか、かのように、から、からには、かわりに、くせに、けれども、こそ、ことなく、し、ずに、そばから、だけに、だったり、つつ、て、で、ていらい、てから、と、ところが、どころか、とたんに

多義性を考慮して、これらの接続表現を含む353の文を選び、そこから以下に述べるような論理的意味範疇の体系を抽出した。

5.1.1 時間

時間関係は、ある時点を基準にして、それ「以

前」と「同時」と「以後」の3つに分類できる。また、これとは別の系列として、特定の時点を基準とすることなしに事態の「継起」を表現する範疇がある。

以前

期限：汽車が出るまでだいが間がある。

同時

関連：少年が思い入ったような態度を見るにつけ、私にはすべてが恐ろしかった。

機会：私は貴殿が9月にローレンスを訪問された折りにお目にかかれて嬉しく思いました。

推移：彼は年をとるとともに体つきが父親に似てきた。

随伴：我々は風景を眺めつつ山を下る。

対立：彼は人を疑いながらも重要な事を託している。

非随伴：眉ひとつ動かさずに1万ドルぼんと出した。

即時：一番鶏が鳴いたとたんにその亡霊は消え失せた。

交互：彼は喜んだり悲しんだりした。

以後

以来：妻を失って以来、彼は氣力が衰えた。

継続：卒業してからも研究を続けるつもりです。

対偶：双方の申し立てを聞いてからでなくては判断は下せない。

対照：彼は刑務所を出てからはまともに暮らしている。

5.1.2 因果

人間の認知・判断作用にとって、因果関係というのはきわめて重要である。また、因果関係を表す表現形式は多岐にわたっており、「時間」や後述する「関係」に比べても格段に複雑である。これは、次のような事情によるものと思わ

れる。客観的な世界にある（と想定される）因果関係に対して、人間はさまざまな態度・視点を取ることができる。原因->結果の通常の関係に沿って認識することもできるし、逆に結果->原因というように逆方向的に認識することもできる。また、原因に働きかけることによって結果を積極的に生起させようとすることもできるし、逆に結果をめざして行動することもできる。さらに、原因と結果の関係の如何によってさまざまな感情を抱くこともできる。つまり、因果関係をめぐって人間はさまざまな意志を働かせ、さまざまな意味を汲み取って、それを多様な表現形式で表しているのである。

順接：ある原因から予期されるとおりの結果が現れること

原因=>結果：私どもの家は南向きだから冬でも暖かい。

悪因=>悪果：わたしが注意してなかつたばかりに、子どもたちにけがをさせてしまった。

結果=>原因：川の水は岸とすれすれになるまで増水した。

条件：原因を条件とみなして、その条件のもとに結果が生起するというを表す。

仮定：前例がまったくないなら新たに作らなくてはならない。

確定：明日になれば帰ってくるだろう。

既定：ああ決心している以上止めたってむだだ。

限定：生命のある限り希望もある。

否定：他の事をしたいと思わないようなら本当の仕事ではない。

反実：病気でなかつたら行くのだった。

十分：おとなしくしてさえいれば、ここにいてもよい。

理由：暗いからお宅まで送りましょう。

根拠：彼に金をやるとは君もどうかしている。

手段：医師はヴァリアムを与えて彼女を鎮静

させた。

註)原因に働きかけることによって結果を人為的に生起させようとするとき、手段という関係をとる。

目的：膜が破れないように気をつけないければならない。

註)手段と反対に、結果をめざしてあることをしようとするとき、目的という関係が現れる。

逆接：ある原因から予期されるとおりの結果が現れないこと

原因 >結果：彼は年をとっているけれども、まだ活動的である。

結果 >原因：君ともっと話していたいけれど、時間がないんだ。

留保：いろいろと欠点はあつても、すばらしい人だ。

註)ある事態を一部承認するという意味。「留保」というのは、「判断」を留保するというこであるから、順接の「(判断)根拠」と逆の関係にある。

条件：逆接の因果関係でも条件があり、仮定以下、順接と同様の区分がある。

仮定：だれに話しかけられても顔を赤らめる。

既定：カウンセリングを受けても効き目がなかった。

否定：道を聞かないでも彼女の家はすぐに分かった。

反実：あの時にはすぐ出かけたとしても、やっぱり遅れたでしょう。

感情：論理関係にかかわる限りでの「感情」とは、因果関係に対する感情的な反応として現れるものである。例えば、ある原因から予期されたとおりの結果が生じたときに「満足」したり、予期されたとおりの結果が生じなかったときに「意外」の感にうたれるというように。

順接

慨嘆：こんな失敗をするとは彼の手腕を疑

いたくなる。

非難：彼に金をやるとは君もどうかしている。

驚き：そんな装備で冬山に登るなんて無謀だ。

疑問：彼が交通事故を起こすなんてちょっと考えられない。

憤慨：それを私のせいにするなんて、あなたも恥知らずだ。

願望：彼女が快方に向かってくれたらとせめてもの希望をつなぐ。

Etc.

逆接

不服：こっちが一日じゅう汗水垂らして働いているというのに、あいつらは家でのおんぴり過ごしている。

失望：寒かったけれど部屋は暖房をつけなかった。

感嘆：彼は年をとっているけれども、まだ活動的だ。

非難：あの人はぴいぴいしているくせに人におごるのが好きだ。

Etc.

5.1.3 関係

時間関係でも因果関係でもない、一般的な関係である。大きく「同一」「差異」「存在」の3つに分けられる。

これまで述べてきた「時間」も「因果」も関係には違いないのだから、ここでことさらに「関係」というカテゴリーを持ち出すのはおかしいと思われるかもしれない。

しかし、見方を変えて人間の認識・判断作用という観点から見ると、「同一」「差異」「存在」といった概念は時間や因果よりもっと根源的であるとも言える。このような最も基礎的な認識・判断の上に立って、時間とか因果といったより高度で複雑な認識・判断がなされる。

同一

形容：航空隊はロンドンを襲撃するばかりに準備を整えて待っていた。

類似：何かを言おうとするかのように唇を開いた。

差異

対立：冬の寒さはなんともないが夏の暑さはこたえる。

比較：怒ったというよりは、悲しそうな顔でした。

註) 差異を前提として、その差異がどの程度のものかを判断するのが比較。差異が極小の場合は同一性になるので、比較は「同一」と「差異」の両方にまたがっているとも考えられる。

存在

限定：あなたには受け入れるしか選択の余地はない。

並存：彼は小説家であるのみならず詩人でもある。

排存：彼は来るかもしれないし来ないかもしれない。

換言：彼女は攻撃的すぎるというか、自己主張が強すぎるんだ。

5.2 節の意味

一つの文には多層な意味が積み込まれている。命題の意味、モダリティの意味、発話行為の意味、価値評価、等々。また、命題の中にも、中核をなす格関係と、それらを修飾するさまざまな意味がある。これら多層な意味のいずれかが、何らかの視点に応じて浮かび上がってくる。例えば、「逆接」関係を表す「あの人はぴいぴいしているくせにおごるのが好きだ」という重文では、2つの節のそれぞれの多層な意味のうち、逆接関係に置かれた「ぴいぴいしている」と「おごる」という表現が格別にクローズアップされることになる。

節の意味を「意味属性」「モダリティ」「価値評価」という3つの観点から記述する。

5.2.1 意味属性

「日本語語彙大系」の用言意味属性と名詞意味属性を与える。ただし、「語彙大系」の意味属性は、結合価ないし格関係をコントロールするために設けられたものなので、重文の意味を表現するにはいささか違和感のあるものである。例えば「言う」という動詞の意味属性は「精神的移動」であるが、重文の意味を構成するものとしては「言語行為」といったような意味属性を使いたいところである。そこで、「語彙大系」の意味属性を、より一般に理解しやすいような属性に写像する。

5.2.2 モダリティ

発話者の意図や事態の様相といった本来の意味でのモダリティよりもやや広い意味で使っている。命題の中核的な意味を何らかの意味で修飾する要素を含めている。節と節との論理関係に関わるものとしては、名詞や動詞の意味属性よりもあるいは大きな意味あいを持つかもしれない。

否定：たとえ親の命令であろうとも、正しくないことはしない。

部分否定：それは事実をすべて述べていないまでも事実に近い。

全否定：誰も手を触れないのに、自然にドアが閉まった。

完了：あの人は日本へ来たばかりなのに、もう日本語が話せます。

継続：こっちが一日じゅう汗水垂らして働いているというのに、あいつらは家でんびり過ごしている。

反復：何度読んでも、ここの所の前後の関係が分かりません。

状態：その時酔っばらっていたからということで彼の行為を許すわけにはいかない。

可能：ハリウッドのスターに会えるというのでわくわくしていた。

不可能：彼の心の裏を見抜けないようではき

みも甘い。

不可避：どうせやらなければならぬのなら
楽しくやりなさい。

蓋然：彼は或いは来るかもしれないし来ない
かもしれない。

受動：人から頼まれるままに承知した。

使役：軍事行動はテロをやめさせられないば
かりかいつそう誘発させることになる。

義務：何を買うにしてもお金はもっと有効に
使うべきだ。

願望：君ともっと話していたいけれど、時間
がないんだ。

意志：政府が当初の公約を果たすつもりなら
その力を結集する必要がある。

強度：その原因は非常に曖昧だったので、誰
もはっきりとそれを指摘できなかった。

過度：彼女は攻撃的すぎるというか、自己主
張が強すぎるんだ。

5.2.3 価値評価

ある事態が「好ましいことか、好ましくないことか」「快をもたらすものか、不快をもたらすものか」あるいは、一般的に「善いことであるか、悪いことであるか」など、事態に対する価値評価を表すものである。

例：歌は好きなんだけどひどい音痴なんだ。

この文では「好き」がポジティブ、「音痴」がネガティブとなる。「だけど」という接続表現は逆接関係を表すが、この関係を担っているのが「ポジティブ」「ネガティブ」という価値評価である。

価値評価は、因果関係の論理計算に特に重要な役割を果たす。因果の項で述べたように、人間は因果関係に対してさまざまな態度・視点をとることができるし、また、因果関係の如何に応じてさまざまな感情を抱く。特に感情は価値評価と切り離すことができないものであり、ほとんど表裏の関係にあると言ってもよい。

価値評価の仕方は多様でありうる。節と節との論理関係が複雑なものになるにしたがって、それに関与する価値評価もまたそれ相応に複雑なものになると予想される。

6 おわりに

「意味類型を用いた類推による機械翻訳」方式を提案し、「重文」の意味を表現するための論理的意味範疇の体系について説明した。

論理的意味範疇は、日本語と英語のそれぞれの意味を表現するものであると同時に、日本語から英語への翻訳を媒介する仕組みでもある。したがって、機械翻訳の中心的な課題である多義性の問題を解決するための一つの有力な方法となることが期待される。

謝辞 この研究は、科学技術振興事業団（JST）の戦略的基礎研究事業（CREST）の支援と科学研究費補助金 基礎研究（B）（課題番号 13480091）を受けています。

参考文献

- (1) 有田潤：ドイツ語講座，南江堂（1987）
- (2) 市川亀久彌：創造的研究の方法論（増補版），三和書房（1963）
- (3) 池原悟，宮崎正弘，白井諭，林良彦：言語における話者の認識と多段翻訳方式，情報処理学会論文誌，Vol. 28, No. 12, pp1269-1279（1987）
- (4) 池原悟，宮崎正弘，白井諭，横尾昭男，中岩浩巳，小倉健太郎，大山芳史，林良彦：日本語語彙大系，岩波書店（1997）
- (5) 岸井謙一，伊佐治和哉，高木優紀江，池田尚士：文節解析システム ibukiB と大規模コーパス中の文節パターンの分布について，言語処理学会第9回年次大会，2003
- (6) 大野晋，浜西正人：類語国語辞典，角川書店（1985）
- (7) 国立国語研究所：分類語彙表，秀英出版（昭和40年）