

# カテゴリ理論による言語処理概念の再構築

新 田 義 彦

## 1. はじめに

従来、自然言語処理における諸概念は、処理アルゴリズムと処理対象データベースを基盤に据えて構築されてきた。具体性が強いために実装システムの構築や現場作業員の理解が容易になるという利点を持つ。一方、ソフトウェア工学の技術的側面を重視するあまりに、言語がもつ概念構造、理論構造、特に代数的構造、といったような普遍性や一般性を持つ考え方から離れすぎため、時に近視眼的評価や重複技術の開発などに陥りやすい欠陥も不可避であった。

本論文では、これらの欠点を克服する目的で、データや情報構造の変換を、1つの圏から他の圏への射 (morphism) として捉え、射の間の変換をさらに考察できる圏論 (Category Theory) を援用して、自然言語処理の諸概念の再構築を試みる。

取り上げる自然言語処理概念には、情報検索、情報抽出、文章生成、文章要約、質問応答、対話処理、暗号処理、機械翻訳、などが含まれる。

圏論における射の概念、関手の概念、極限の概念、錐の概念、自然変換の概念、さらに随伴の概念が、自然言語処理システムの開発工程に、見通しのよいガイドラインを与えうることを示すことが本論文の眼目である。

以下の章では、カテゴリ理論の基礎概念の概観をした後、下記のような自然言語処理の諸概念をカテゴリ論により統合的に掌握する工夫について述べる。自然言語処理は、Natural Language Processing の頭文字を取って、NLP と略記することにする。

取り上げる NLP の諸概念のうち代表的なものを列挙すると下記ようになる。

- (1) : 自然言語処理とは → NLP 全般の定義と処理の概要を定式化する
- (2) : 語の処理 → 語の辞書情報、語が文章全体を代表する機能 など
- (3) : 句の処理 → 複数の語の緊密なつらなり (結束) がもつ機能そして形態について
- (4) : 文の統語構造 → 文の構成をその形式だけから定義し処理する
- (5) : 文の統語構造 → 文の機能や効果をその使用環境・存在環境との関係で定義し処理する
- (6) : 情報検索
- (7) : 情報抽出・知識発掘
- (8) : 文章生成
- (9) : 文章要約
- (10) : 質問応答・対話処理
- (11) : 暗号処理
- (12) : 機械翻訳

## 2. カテゴリ理論の基礎概念

カテゴリ理論の基礎概念や技法が、自然言語処理 (NLP) の再構築に有効な理由の概略を、Eilenberg, S., Mac Lane, S. (1945), Esterbrook, Steve (1999), Freyd, P. (1964), Goguen, Joseph (1989), Mac Lane, Saunders (1997) などに準拠して述べる。

- 1) アルゴリズムやデータ構造に軸足を置いて、論じられてきた NLP の多様な技法やアプリケーションを、一般性や普遍性の高い形式 (た

たとえば数学における定理や定義)で記述することができる。この記述を通してある種の様式美や論理的整合性をも入手できる。

- 2) 記述された NLP の様々な技法や手法は、数学的な証明の対象とすることすら可能となる。
- 3) NLP の様々な分野を縦断的に観察すること、共通するより高次の概念や技法を発見することも可能となる。
- 4) NLP のある処理を関手 (functor)  $F$  として定式化できたならば、その随伴  $G$  を仮定して探索あるいは構築することにより、新しい処理や別の観点の処理を構築できる可能性がある。
- 5) 同型、自然性、普遍性という概念を、より厳密に NLP に持ち込むことが可能となる。
- 6) 対象間の半順序関係、全順序による列挙などから一般性の高い対象に辿りつけることがある。
- 7) 直和や双対 (カテゴリ  $C$  の射の向きを逆にして、双対カテゴリ  $C^{\text{op}}$  が得られる) から一般性の高い概念に辿りつけることがある。
- 8) リミット (射影的極限, projective limit) やコリミット (帰納的極限, inductive limit) から一般性の高い概念に辿りつけることがある。
- 9) 図式 (diagram) やその上の経路 (path) を追跡することにより一般性の高い新概念に辿りつけることがある。

### 3. 自然言語処理の主要概念のカテゴリ化

辞書検索、形態素解析 (Saraki, M. and Nitta, Y. (2005, 2008, 2011), 久光徹, 新田義彦 (1991, 2004), Tohru Hisamitsu, T., & Nitta Y. et al. (1996)) 構文解析 (Maurice, Gross (1972, 1989),) 意味解析, などの基礎処理, そして, 情報検索, 知識発掘, 質問応答, 機械翻訳, 自動要約, などの応用処理は, すべて異なる対象を持つカテゴリ (圏) として, 再定義できる。ひとたびカテゴリとして定式化できたならば, カテゴリ間の射 (つまり関手) を考えることにより, 処理の共通化や汎化ができ, システム開発の効率向上が期待できる。

たとえば辞書検索や形態素解析は, 単語  $w$  を対象とするカテゴリの射 (morphism) として定式化できる。構文解析や意味解析は, 句  $p$  を対象とするカテゴリの射 (morphism) として定式化できる。

情報検索や知識発掘は, キーワード  $k$  やキーワード  $kp$  を対象 (object) とするカテゴリと文献やドキュメント  $D$  を対象とするカテゴリの間の関手 (functor) として定式化できる。

質問応答は, 質問文集合を対象とするカテゴリと回答文集合を対象とするカテゴリ間の関手として定式化できる。

機械翻訳には, 様々な関手をその定式化として採用できるが, 原型となる関手は, 言語  $A$  で記述された文の集合を対象とするカテゴリと, 言語  $B$  で記述された文の集合を対象とするカテゴリ, の間の関手である。「言語  $A$  で記述された文集合」の部分のバリエーションを考えることにより, 翻訳の種々の位相を定式化して取り扱うことが可能となる。

自動要約は, 大きな (長い) 文の集合を対象とするカテゴリと, 小さい (短い) 文の集合を対象とするカテゴリの間の関手として定式化できる。その際に何らかの意味記述を対象とするカテゴリ (圏) を, 随伴させておく必要がある。意味の保存を担保する必要があるからである。

随伴関手による意味の保存という注意は, 情報検索, 機械翻訳などにも適用される。

### 4. 言語表現の担う意味のパターン化

本章では, 自然言語処理の中核をなす種々の言語要素のパターン変換 (cf. 新田義彦 (2012), Ikehara, S. et al. (2002)) を, 具体的対象としてカテゴリ論的再構築を試みる。以下, 4~8章のパターンの等価変換と計算の議論は, 文献 [新田義彦 (2012)] の第4章を縮約しつつ引用して, カテゴリ論的考察を加えてまとめた用のものである。

まず, 意味類型パターンを表層単語列から抽出

する例を取り上げる。

日本語文や英語文のような自然言語文を、変換したり生成したりする計算機プログラム（応用システム）を実現するためには、統語知識だけでは不十分であり、意味知識の参照が必要となる。たとえば、下記の例のような翻訳や言い換え（1）と（2）を考えてみよう。

- (1) この虫は紙を食う。  
 ⇒ This bug eats paper.  
 ⇒ この虫は紙を栄養源とする。
- (2) この趣味は金を食う。  
 ⇒ This hobby wastes money.  
 ⇒ この趣味にはお金がかかる。

上記のような翻訳や言い換えを正しく実行するには、「食う」という日本語動詞の意味を、その支配する名詞の意味との関係で精密に把握する必要があるため、(3)、(4)、(5)、(6)のような統語知識だけでは不十分である。

- (3) この虫／名詞句 (Subject) は／助詞 (Case Marker) 紙／名詞 (Object) を／助詞 (Case Marker) 食う／動詞 (Predicate)
- (4) This bug/Noun Phrase (Subject) eats/Verb (Predicate) paper/Noun (Object)
- (5) この趣味／名詞句 (Subject) は／助詞 (Case Marker) 金／名詞 (Object) を／助詞 (Case Marker) 食う／動詞 (Predicate)
- (6) This hobby/Noun Phrase (Subject) wastes/Verb (Predicate) money/Noun (Object)

翻訳や変換・言い換えを適正に行なうためには、下記の(7)や(8)のような語彙レベルの意味知識の参照が最低限必要となる。

- (7) 虫 (+ animal) 食う (+ animal action to edible thing) 紙 (+ edible)
- (8) 趣味 (+ abstract, - animal) 食う (+ abstract

action to resource) 金 (+ resource, - edible)

意味知識ベースとして既に確立している体系は「辞書 (dictionary, lexicon)」であるが、“語と語の意味関係”や“語と語の可能な接続関係”などが、計算機処理の要求を満たすほど十分に提示されているわけではない。しかし近年徐々にその増強と完備化が進められている。特にインターネット上の膨大な文書集合をコーパスとして整備されつつある語彙体系はその効力が期待できる。

単語間の同義関係、類義関係、含意関係、などの語彙知識を与える知識ベースとしての「シソーラス」のうち現行の代表例をあげると下記のようになる。

- i) 分類語彙表, 国立国語研究所 発行: 1つの見出し語には1つの語義しか与えない。そして単語間の関係(文脈)情報に関する情報は与えない。単語とその意味を規範的に分類するという観点から体系化している。
- ii) 角川新類語辞書, 角川書店 発行: 同上。2つ単語間の意味類似性の記述は詳しい。
- iii) EDR 電子辞書, 日本電子化辞書研究所 (= EDR) 発行: 種々の計算機応用プログラムでの利用を想定して構築されているため、文脈情報が付与されているが、2つの単語の共起関係に限られている。また語義の汎化は扱っていない。
- iv) IPAL 計算機用日本語基本辞書 (IPAL (1987)), 情報処理振興事業団 (= IPAL) 発行: 基本語を精選して精緻な語釈と文脈情報を付与することを目的としているため、語彙数が小さい。動詞、名詞、それぞれの基本語を約1,000語ずつ扱っている。

通常の“文”は、3つの単語以上から構成される場合が多いが、上記で見たような既存のシソーラスは、そのような“文の意味”を扱うためには

情報不足である。またシソーラスへの登録語彙数も十分とは言い難い。このような欠点が比較的少ないシソーラスとして、最近開発された下記のシソーラスが注目される。

v) 日本語語彙大系, 岩波書店 発行: 用言とそれが支配する名詞句や修飾語句の意味的依存関係を表記している点が特徴である。単語の意味属性を意味分類コードにより弁別的に表現している。意味分類コードは12段の階層構造の上に配置されており、全体として約3,000個のコードから構成されている (Ikehara, S. et al. (1997))。

この「日本語語彙大系」を1つの出発点データとして文の意味パターンを観察することにする。複数の単語群が一行に並んだ文が記述提示している情報、つまり“文の意味”を表現する形式概念として、「意味類型パターン」を仮定し、その具備要件や実現・実装の方法を議論することがNLPの一つの重要テーマとなりうる。

上述の議論は、単語の集合を対象とするカテゴリと、意味素性の集合を対象とするカテゴリ、の間の関手として考察可能である。また「意味パターン」は、単語列パターンを対象とするカテゴリ、と意味素性列パターンを対象とするカテゴリ、の間の関手として考察可能である。

意味類型パターンの観察をもう少し続ける。

下記(9)のような表層文に対応する意味類型パターンの例は、下記の(10)と(11)のようになる。ただし、N1, N2, N3などは名詞または名詞句の存在を示す汎化記号である。それらが(社長, 工場, 自動溶接機)のような名詞〔句〕集合の場合、意味類型パターンが(10)であり、(給仕, 私, コーヒー)のような名詞〔句〕集合の場合、意味類型パターンが(11)である。

(9) N1 が N2 に N3 を入れる。

(10) N1 〈#3 主体〉が N2 〈#389 施設〉に N3 〈#962 機械, #2007 設備〉を入れる。

(11) N1 〈#4 人〉が N2 〈#4 人〉に N3 〈#857 飲み物〉を入れる。

上記の日本語文パターンを英語文パターンへ書き換えた結果は、たとえば下記の(12)と(13)のようになる。

(12) N1 installs N3 in N2.

(13) N1 makes N3 for N2.

上記の(10)と(11)に出現する“#3, #4, #389, #962, #2007”のようなコード番号は、日本語語彙大系において定義されている意味類型コード番号である。

「日本語語彙大系(岩波書店刊行の全5巻の辞書)」自体に記述されている、意味類型パターンの総数は、約14,000パターンである。各々のパターンは、用言1個とその前後に配置された主語、目的語、補語、修飾語、などからなる日本語の単文にほぼ対応している。各単文の構成単語は、日本語動詞約6,000語から選んだ1語、および日本語名詞30万語から選んだ数語である。

約14,000個の単文パターンが、多様な日本語単文全体をどの程度被覆しているのか評価することは難しい。しかし、これらの単文パターンは、先ほど見た意味類型パターンの例題(12)や(13)に示したように、意味類型コードによる“表層単語の意味の汎化”を行なっているため、存外に被覆率は高いと予想される。

## 5. 等価変換における意味の保存

自然言語処理の基本目的は、下記の2つのカテゴリに典型化・対照化して把握すると分かりやすい。

- i) 自然言語文の担う意味（つまり文や文章が伝えようとしている語用論的な効果）やメッセージ内容）を、計算処理により積極的に変形・発展・転換すること。
- ii) オリジナルの自然言語文の意味を保存（不変化）したまま変形・発展・転換すること。

オリジナルの文の意味は、意味類型パターンの中に抽象化・汎化して取り込むことができると仮定する。換言すれば、意味類型パターンに取り込めないような“精緻な意味内容（ニュアンスなど）”は忘却される。「忘却」という考え方もカテゴリ理論に由来する。

意味類型パターンの中に取り込まれた「文や文章の意味」を不変のまま変形・変換する方法、つまり“意味保存型の変換法”は、「等価変換」という名称をもつ古典哲学概念に基礎を置いている。この概念は理論物理学における基礎概念でもあり、有名な相対性理論をアインシュタインが考察する際の思想的基礎になっていたことも知られている。

工学や技術分野にける新発見や発想の基盤を哲学的に研究した市川亀久彌（1963）の定式化に依拠して、パターン変換を見ていくことにする。市川亀久彌は、“独創性”の背景にある“等価的類推思考”を下記のような記号式により定式化した。

$$(14) C (A/a = / \varepsilon B/\beta)$$

ただし、式（14）は“条件Cの下で、=の左辺と右辺の間で意味的等価関係が成立している”というように解釈する。ここで“意味的等価関係”とは、条件Cの立て方により、“類推”であったり“翻訳”であったり“言い換え”であったりする。“=”の右辺と左辺の間には、“等価性”という関係があるだけであり、“方向性”や“非対称性”はない。

もう少し詳細に記号の意味を検討する。A/aは言語体系 $\alpha$ の上で記述された表現であり、B/ $\beta$ は言語体系 $\beta$ の上で記述された表現である。換言すれば、言語体系 $\alpha$ や $\beta$ の上で思考した際の、“対象となった事象・事物”や“結果（あるいは過程）”としてのイメージや想念”を表す記号がA/aやB/ $\beta$ である。ギリシャ文字の $\alpha$ 、 $\beta$ 、 $\varepsilon$ は、思考や言語表現の基底にある系（システム）を表す。特に“ $\varepsilon$ ”は“=”の左辺と右辺に共通する思考系／言語表現系を表す。たとえば、 $\alpha$ が感情や感覚系であり、 $\beta$ が音楽系であり、 $\varepsilon$ が音楽家一般の思考系を表現するものとすれば、式（14）は音楽家が自身の感情や感覚を等価変換して楽譜に変換するという行為、つまり“作曲”という知的行為を定式化したこととなる。同様に、 $\beta$ を図形・色彩・構図などの画像系とすれば、画家の“描画”という芸術的創造行為を定式化したことになる。

このように原典（市川亀久彌（1963））が提唱する“等価変換”は、創造、工夫、発明、発見、など、非常に広範囲の知的所為の基底にある哲学的構造を扱うものであるが、本論文では、定式化（14）の作用域を言語現象のみに限定して考察することとした。具体的には、下記のような枠組みa)、 b)、 c)の下で考察を進める。

- a) 条件Cは、文の意味のうち意味類型パターンにより抽出（写像）できるものを扱う。
- b) 言語表現AおよびBは、自然言語の文とする。文の構造は単文・複文・重文などであるが、特に本研究では複文に関心がある。単文のパターン化にはある程度の見通しが、日本語語彙大系の開発により得られている（Ikehara, S. et al. (2002)）からである。
- c) 表現の基底にある表現系 $\alpha$ や $\beta$ は、日本語や英語、ドイツ語、フランス語、といった自然言語の区別を表すこととする。一般に $\alpha \neq \beta$ の場

合には、等価変換は言語  $a$  から  $\beta$  への翻訳を表すこととなる。また  $a$  と  $\beta$  とが同一の言語系に属していて、単に語彙集合や統語解析規則の相違などがある場合には、言い換え・要約を表すことになる。

d) 等価変換においては、双方向性や対称性を仮定せず、左辺から右辺への一方向への変換を考える。双方向性を仮定すると、翻訳や言い換えの場合に、双方向翻訳や相互（可逆）変換を実現しなければならなくなるが、これは現状の自然言語処理技術の水準に照らして困難である。ゆえに、式 (14) における等号 “=” の代わりに、等価変換 “ $\Rightarrow$ ” を使う。等価変換または等価変換は、カテゴリ理論における自然変換として、再評価することが可能である。

e) 等価変換 “ $\Rightarrow$ ” における基底表現系は、述語論理 (predicate logic) とする。つまり等価変換 “ $\Rightarrow$ ” において、不変に保存される意味情報は、述語論理で表現可能なもの限定することにする。

結局、本論文で扱う等価変換の定式化は (14) 式の代わりに下記 (15) 式として記号表現できる。

$$(15) C(A/a) \Rightarrow / \varepsilon C(B/\beta)$$

上式における記号の読み方を念のために再論する。“ $C(A/a)$ ” は言語  $a$  で記述されている文（一般的には言語表現） $A$  を、意味類型パターン記述文法  $C$  によって記述（表現）したものである。記号 “ $C(B/\beta)$ ” の解釈も同様である。記号 “ $\Rightarrow / \varepsilon$ ” の解釈は、述語論理 ( $\varepsilon$ ) で表現（抽出）される意味を保存しつつ、左辺の言語表現を右辺の言語表現に変換（変形）することである。

言語  $a$ （たとえば日本語）から言語  $\beta$ （たとえば英語）への機械翻訳の場合を具体例として、等

価変換の手順を以下に示す。

第1段) 言語  $a$  で記述されている文  $A/a$  を、“言語  $a$  の意味類型知識ベース  $KB a$ ” と照合させつつ、意味類型パターン  $C(A/a)$  に変換する。

第2段) 意味類型パターン  $C(A/a)$  を、述語論理系  $\varepsilon$  の形式  $A/a/\varepsilon$  に変換する。この変換処理においては述語論理規則系と共に、表層文  $A/a$  および意味辞書  $SD a$  などを参照する。

第3段) 述語論理形式  $A/a/\varepsilon$  を参照しつつ、“言語パターン変換規則  $PT a \beta$ ” にしたがって、意味類型パターン  $C(A/a)$  を意味類型パターン  $C(X/\beta)$  に変換する。ここで、 $X/\beta$  は、言語  $\beta$  による記述文を変数化（汎化）した表現を表しており、これに言語  $a$  と言語  $\beta$  に共通する意味類型コード（つまり厳密に言うとは回りくどくなるが、意味類型パターン記述文法  $C$  が定めるコード）を付与した表現が、 $C(X/\beta)$  である。 $C(X/\beta)$  は、厳密に言うとは回りくどくなるが、意味類型パターン記述文法  $C$  が定めるコードである。

第4段) パターン  $C(X/\beta)$  を、“言語  $\beta$  の意味類型知識ベース  $KB \beta$ ” と照合させつつ、意味類型パターン  $C(B/\beta)$  に変換する。多義性の処理は、前述のように補助的な意味処理や辞書参照により適宜行なう。

第5段) 意味類型パターン  $C(B/\beta)$  から、言語  $\beta$  による表層文  $B/\beta$  を生成する。生成においては、“言語  $a$  と言語  $\beta$  の語彙の対照辞書  $TD a \beta$ ” などを援用する。

上述の変換（変形）過程を “ $\Rightarrow$ ” により記号化すると下記 (16) 式のようになる：

- (16)  $A/a$   
 $\Rightarrow /KB\ a, SD\ a$   
 $C(A/a)$   
 $\Rightarrow / \varepsilon, SD\ a$   
 $A/a/\varepsilon$   
 $\Rightarrow /PT\ a\ \beta, C(A/a)$   
 $C(X/\beta)$   
 $\Rightarrow /KB\ \beta, SD\ \beta$   
 $C(B/\beta)$   
 $\Rightarrow /SD\ \beta, TD\ a\ \beta$   
 $B/\beta$

ただし、“ $\Rightarrow /x$ ”は左辺から右辺への変換において、“知識  $x$  の制御を受けること”，もしくは“知識  $x$  を参照・照合しながら変換すること”を意味する記号である。また  $KB\ a$ 、 $KB\ \beta$  はそれぞれ、“言語  $a$  の意味類型知識ベース”および“言語  $\beta$  の意味類型知識ベース”である。 $SD\ a$ 、 $SD\ \beta$  はそれぞれ“言語  $a$  の語彙的意味辞書”および“言語  $\beta$  の語彙的意味辞書”である。また  $TD\ a\ \beta$  は、“言語  $a$  と言語  $\beta$  の語彙対照辞書”である。 $\varepsilon$  はある種の“述語論理系”であるが、その詳細は本報では述べない。 $PT\ a\ \beta$  は、“言語  $a$  から言語  $\beta$  へのパターン変換（トランスファー）知識（＝パターン変換規則系）”である。その他の記号、 $A/a$ 、 $B/\beta$ 、 $C(A/a)$ 、 $C(B/\beta)$  などの意味は従前の通りである。

このように文の意味を代数記号で表現してみると、カテゴリ理論における射（morphism）の考え方、そして種々の関手や極限が、設定でき、処理の透明化や汎化、共有への道が拓けることがわかる。

## 6. 述語論理による文の意味記述

第4章で“ $\varepsilon$ ”で表記した論理系の概観を、論理系による“文の意味記述”の例により示す。下記の例題から、論理系は、文の意味をマクロなレベルで記述するのに適していることが分かる。つ

まり、複数個の単文の複合体（＝複文）における“単文と単文の意味的關係”が論理記号により端的に表現可能であることが分かる。このような論理系記述を、パターンマッチングの標識として利用すると、複文レベルの“意味類型パターン”の検索照合やパターン変換が効率的に、同時にまたマクロな意味関係を精密に保持しながら実行可能となることが予測できる。

### (23) 相反事象の論理構造：

常識推論 ( $P1 \rightarrow P2$ )  $\wedge$  /contradict 現実 ( $P1 \cap \text{not } P2$ )

文例：彼は全面的に協力すると言いながら、こちらが頼んだことはすべて理屈を付けて断ってくる。

$P1$  = 彼が全面協力すると言明する

$P2$  = 彼が当方の依頼を引き受ける

$x \wedge$  /contradict  $y$  = “ $x$  と  $y$  が相反する”

### (24) 相反事象の論理構造：

事実 ( $P1$ )  $\wedge$  /contradict 実感 ( $P2$ )

文例：あの小説は今年最高の文学的収穫と評価されているが、読んでみるとすぐに眠くなるような陳腐凡庸な内容であった。

$P1$  = あの小説は今年最高の文学的収穫という評価を得ている

$P2$  = あの小説は読み始めるとすぐに眠くなるような駄作である

### (25) 模範に対する常識補完構造：

教科書の規範 ( $P \rightarrow G$ )  $\wedge$  常識（過剰・固執 ( $P \rightarrow \text{not } G$ )

文例：いくら健康食品だからといって、毎日鰯団子ばかり食べていては貧血を起こすよ。

$P$  = 鰯団子を食べる

$G$  = 健康によい

### (26) 模範に対する常識補完構造：

教科書の規範 ( $P \rightarrow G$ )  $\wedge$  常識（過剰 ( $P$ )

→ not G)

文例：鞆は大きいほど沢山ものが入って便利だが、大き過ぎると持ちにくい。

P = 鞆が大きい

G = 便利である

(27) 規範的含意に対する常識補完構造：

規範的含意 (P1 → not P3) ∧ 常識推論 (P1 ∩ P2 → P3)

文例：語学の才能が無いからと言っても、10年も米国に住んでいれば買い物ぐらひは英語でできるようになるだろう。

P1 = 語学の才能を持たぬ

P2 = 10年間米国に住む

P3 = 英語で買い物ができる

(28) 論理的真理とは異なる感覚的・直観的真理の陳述構造

論理的真理 (P1 ≠ P2) ∧ 信念的・感覚的真理 (P1 = P2)

文例：炭鉱で働くことはかつては奴隷労働と変わらなかった。

Working in coal mines was formerly tantamount to slavery.

P1 = 炭鉱で働く

P2 = 奴隷労働をする

上記の論理記述においては、主語述語や提題などの変数記述、変数(変項)に対する量子子記述は省略した。複文の意味構造を、単文レベルの意味を基本要素(つまり1つの命題)として論理記述している。

### 7. パターン変換の計算

機械翻訳を典型例とするパターン変換の計算内容や特徴を以下にスケッチする。

第1段) 言語  $a$  で記述されている文  $A/a$  を、意味類型パターン  $C(A/a)$  に変換する計算：

$A/a \Rightarrow /KB a, SD a \quad C(A/a)$

まず言語  $a$  で記述されている表層文  $A/a$  に、パーズング処理を施してして品詞タグ付けをする。品詞タグ付けにおいては、機能語の表示と意味類型コードの付与が大切である。パーズングは、文脈自由度言語としての統語解析ではなく、正規言語 (cf. Hopcroft, J. E. & Ullman, J. D. (1970)) としてのパターンマッチング処理により行なう点が本質的である。“自然言語文の意味はパターンに写像できる”という前提を置いているので、パターンへの写像はあくまで“パターンマッチング処理”つまり“決定性有限状態状態オートマトンによる「正規言語の計算」”でなければならない。

パーズングにおけるあいまい性解消処理には、“言語  $a$  の語彙の意味辞書  $SD a$ ” が参照されることもある (cf. Maurice, Gross (1989))。知識ベース(つまり  $KB/a$  や  $KB/\beta$ ) はその見出し項目(エントリ)が、文パターン(つまり何らかの汎化コードのシーケンス)であるのに対して、意味辞書  $SD a$  の見出し項目(エントリ)は、言語  $a$  の表層単語もしくは語句である点を注意する。

パーズングにより品詞タグ付けがなされ、ある程度の汎化処理が施された表層文(入力文  $A/a$ ) は、“言語  $a$  の意味類型知識ベース  $KB a$ ” と照合されつつ、最適マッチングした意味類型パターン  $C(A/a)$  に変換(置換)される。アノテーション処理(=品詞タグ付け)によりある程度汎化された表層文  $A/a$  を、しばらく、 $G(A/a)$  と表記することにする。 $G(A/a)$  は、変数項を含む汎化シーケンスである。 $G(A/a)$  を知識ベース  $KB a$  に引き当てて照合検索し、最適な  $C(A/a)$  を検出する計算の本質は“部分マッチング”であろう。ここでいう「部分マッチング」は、 $G(A/a)$  に含まれる“変数部分”と  $KB a$  のパターンに含まれる“定数部分(註：意味類型コードや具体的な単語など)”との間の“ローカルなマッチング”

という意味である。この部分マッチングにおいては、意味辞書  $SD\alpha$  が齎す情報を援用することもある。一般にパターンマッチング計算では、一致の程度を表す評価関数が使われるが、これについては続報で述べることにする。

第2段) 意味類型パターン  $C(A/\alpha)$  を、述語論理系  $\varepsilon$  に従う形式  $A/\alpha/\varepsilon$  に変換する計算：

$$C(A/\alpha) \Rightarrow / \varepsilon, SD\alpha \quad A/\alpha/\varepsilon$$

この変換処理においては述語論理規則系  $\varepsilon$  と共に、表層文  $A/\alpha$ 、および意味辞書  $SD\alpha$ などを参照する。述語論理表現  $A/\alpha/\varepsilon$  の最大の効用は、文  $A/\alpha$  全体のマクロな意味的性質を、簡潔にカテゴリ表示できる点にある。たとえば“相反する2つの事象の陳述”や“規範に矛盾する直観・感覚・常識”、“常識推論による不言及事項の補完”といった、マクロなパターン・カテゴリが表現できる。文(特に複文)パターンのマクロな意味カテゴリも論理系  $\varepsilon$  により表記できる。

意味類型コードのシーケンスとしての  $C(A/\alpha)$  を、そのマクロな意味カテゴリとしての論理式  $A/\alpha/\varepsilon$  へ写像する計算も、パターンマッチングにより実行する。このパターンマッチング計算で参照される(註：換言すれば、“パターンマッチング計算を制御する”)知識データベースは、したがって、“意味類型コードパターン”対“述語論理式”の多対多の対応表であると言える。

第3段) 意味類型パターン  $C(A/\alpha)$  を意味類型パターン  $C(X/\beta)$  に変換する計算：

$$C(A/\alpha) + A/\alpha/\varepsilon \Rightarrow / PT\alpha\beta \quad C(X/\beta)$$

ここで、 $X/\beta$  は、言語  $\beta$  による記述文  $B/\beta$  に到達する以前のいわば“虫食い文”である。“虫食い”の部分は“変数  $x$  や  $y$ ,  $z$  などにより汎化されている”部分である。このような、虫食い状態の“汎化部分の残る文パターン”が得られるの

は、2つの言語  $\alpha$  と  $\beta$  の間では、一対一の語句対応が成立しないことによっている。完全対応が取れぬ部分は、変数により汎化(“虫食い”状態化)して対処しているのである。

このような虫食い型の汎化表現  $X/\beta$  に、言語  $\alpha$  と言語  $\beta$  に共通する意味類型コード  $C$  を付与した表現が、 $C(X/\beta)$  である。

また“ $\Rightarrow$ ”の左辺に“述語論理表現  $A/\alpha/\varepsilon$ ”が“+”記号により付加されているが、その意味は、 $A/\alpha/\varepsilon$ (つまり入力文  $A/\alpha$  のマクロな論理カテゴリ情報)を参照しつつ、言語変換規則に則して、 $C(A/\alpha)$  を  $C(X/\beta)$  に写像しているということである。この写像計算も、本質的にはパターンマッチング処理(Koskenniemi, Kimmo (1990))により行なう。変数記号の導入による“汎化”は、パターンマッチングの計算処理にいささかの複雑性を加味することは、既に述べた通りである。

第4段) パターン  $C(X/\beta)$  を、意味類型パターン  $C(B/\beta)$  に変換する計算：

$$C(X/\beta) \Rightarrow / KB\beta, SD\beta \quad C(B/\beta)$$

この変換写像の計算は、“言語  $\beta$  の意味類型知識ベース  $KB\beta$ ”との照合(パターンマッチング)処理が本質であるが、補助的な意味処理のために意味辞書  $SD\beta$  の参照も適宜行なう(註：本論文では論述は略した)。要するに、変数によって“汎化(=虫食い)”されている部分を、知識ベース  $KB\beta$  や意味辞書  $SD\beta$  を参照・照応しつつ、補完処理して、言語  $\beta$  の上の意味類型パターン  $C(B/\beta)$  を構成することが、この第4段の使命である。

第5段) 意味類型パターン  $C(B/\beta)$  から、言語  $\beta$  による表層文  $B/\beta$  を生成する計算：

$$C(B/\beta) \Rightarrow / SD\beta, TD\alpha\beta \quad B/\beta$$

生成においては、言語  $\alpha$  と言語  $\beta$  の語彙対照辞書  $TD\alpha\beta$ 、言語  $\beta$  の意味辞書  $SD\beta$  などを参照

する。(註:対象言語生成技術の詳細は次報に回した。)機械翻訳が、文法に適合する滑らかな訳文を、原文(A/a)の意味を適正に保存しつつ生成できるか否かは生成写像の計算に強く依存する。

変換写像計算の全般に関わる付帯的注意を述べる。元の文A/aをパターンC(A/a)に変換する計算、あるいは、パターンC(A/a)をパターンC(X/β)に変換する計算、さらにまた、パターンC(X/β)をパターンC(B/β)に変換する計算、パターンC(B/β)から出力文B/βを生成する計算、等においては、種々の多義性やあいまい性が発生するが、その解消処理方法については次の論文で扱う。あいまい性解消のためには、入力文の内部に存在する情報(“文内情報”)以外に、当該文の前後に存在する他の文との関係情報(“文間情報”,いわゆる文脈情報)、常識や状況・背景・前提知識・世界知識などの“文外情報”も参照されることを注意しておく。

この意味で、文を言語aから言語βに変換(翻訳)する等価変換計算は、パターンマッチングを何重にも行なう総合的知識処理計算である。

次にパターンマッチング計算の根底にある正規言語(註:Regular Expression, Chomsky階層の計算方法について簡単に触れる。正規言語の文法(生成規則)は、下記(29)式のような形式をしている。

$$(29) A \rightarrow a \text{ or } A \rightarrow aB$$

ただしA, Bは非終端記号(non-terminal symbol)を表し、aは終端記号(terminal symbol)を表す記号とする。

前章までで見た“変換もしくは変形⇒”の左辺と右辺に出現するパターン、文字列、単語列、または語句、などの集合を適宜、適当な正規言語として定義すると、パターンマッチング計算が正規言語上の形式的計算として見通しよく実行でき

る。正規言語の上に写像された記号列(つまりパターン)は、たとえばPerlやSed, awk, Lexといったスクリプト処理系言語プロセッサで扱うことが可能である。Perlにおける“エスケープ表記”としては下記のようなものがあり、パターンマッチング処理計算手順を記述する便宜が図られている。本研究ではPerl言語によるパターン変換計算の実装を考えている。

このようにカテゴリ理論、特にカテゴリ間の種々の変換(関手)は正規表現の変換と強く関係しているので、この関係に立脚した研究開発を進める意義が大きいように思う。

## 8. 意味類型パターンの汎化と意味シソーラスの階層関係

表層文の例とその汎化の様々なレベルをまず示し、次に汎化に関与する意味シソーラスや文法(統語や意味)コード系について述べる。“汎化”は“パターン変換”や“パターン間写像”の本質であることを示した。そして汎化は、カテゴリ間の変換(関手)によりさらに汎化できることを強調したい。

汎化の種々相の一部を以下に列挙する。

i) 文を構成する語や句を、統語コード(たとえば品詞タグ)や概念コード(意味類型)に変換する汎化。

この“語句レベルの汎化”は、その“汎化の程度(=語句の意味を抽象化する程度、意味の不要部を捨象する程度)”の制御を“意味シソーラスの階層構造”に従って行なうという点が重要である。統語コード系のみによる汎化は、通常の状態素解析(=品詞タグ付け)や統語解析(=パーシング)であり、この汎化(解析)処理では、“汎化の程度の調整”はできない。品詞コード系を定める辞書や統語カテゴリコード系を定める文法が、通常固定されているからである。

ii) 単文レベルのパターン(=単一述語パターン)を概念コード(=意味類型)に変換する汎化。

iii) 複文・重文レベルのパターン (= 複数パターン間の関係) を述語論理形式に汎化しつつ, 適切な概念コード (= 意味類型) を付与する汎化.

これまでの例で示したように, 文や文章の“意味の領域”に踏み込むと, タグ付けにおける“汎化”には様々なレベルが存在することとなるが, カテゴリ理論の方法論を使うと, 検討課題の整理統合そして透明化が一気に進むように思われる. “意味と汎化の関係”を, カテゴリ間の変換として統合的にかつ厳密・透明に扱い, ラックボックス処理的なヒューリスティックスに頼らずに扱う意義を大きいと言えよう.

## 9. おわりに

“意味類型パターン”という形式の変換に重点を置いて, その変換をカテゴリ論的に捉える考え方の素描を与えた. カテゴリ理論の形式の上に, 自然言語文の意味を最大限写像する方法とその妥当性について検討することが直近の課題である.

意味類型パターンの主要な記述要素は“意味分類コード”であるが, これについては, 既存のシソーラス大系を調査検討して, 岩波書店刊行の「日本語語彙大系」の意味分類コードが, 拠点コード系として利用可能であることを述べた.

自然言語文の“意味”を“計算可能なパターン”に写像する理由は, 下記3点に要約できる.

- i) 翻訳, 書き換え, 言い換え, 要約, などの自然言語処理の基本的問題のすべてが, “意味類型パターン間の変換・写像”として統一的に定式化できる, という期待.
- ii) 自然言語処理全般における計算方法が, 理論的にも実用的にも透明化・簡潔化できる, という期待.

そして

iii) “パターン間の等価変換または汎化変形”は, カテゴリ論の変換処理と馴染みやすく, 自然言語処理を統一的かつ形式的に計算可能化する道

が拓ける.

“言語表現を典型とする人間の知的所為”を“表現形式間の等価変換”として捕らえる思想は, 「等価変換理論」と呼ばれる古典的な哲学概念に依拠している. 等価変換理論では, 表現形式間の変換の双方向性, つまり変換が可逆であり変換された2つの表現形式間に対称性や互換性が成立することを, 根源的に要請している. 本論文では, 自然言語の計算機処理が, 本質的に“汎化”による“写像”もしくは“忘却”であるという理由で, 変換を“一方向変換”に制限して, “不可逆変換”もしくは“非対称変形”として考察した. しかしながら, 自然言語処理において, 双方向変換や可逆変換を考察する意義は大きい. この問題は今後の課題としたいが, やはり圏論における自然変換の概念が有力な指標を与える.

カテゴリ論における種々の表現そして変換は, 正規表現 (regular expression) の変換と相性がよいこと, 今後の有望な研究対象として期待できること, などにも言及した.

## 参考文献

- 市川亀久彌 (1963) 創造的研究の方法論 (増補版), 三和書房
- 梶博行, 新田義彦 (1979), 文献検索システムのファジイモデルとその実現, 電子通信学会論文誌, Vol. J62-D, No.4. (1979)
- 言語特集 (1988), 機械翻訳の現状と未来, 言語, 1988年1月号, 大修館
- 清水義夫 (2013) 記号論理学講義, 東京大学出版会
- 新田義彦 (1988), 自然言語理解の基礎: 意味論と語用論, 情報処理, Vol.30, No.10, 情報処理学会 pp.1182- (1988, Oct.)
- (2012) 機械翻訳の原理と活用法, 明石書店
- , 田中穂 (1988), 機械翻訳のむずかしさと課題, 言語, Vol.17, No.1, 大修館 pp.50- (1988, Jan.)
- 丹羽芳樹, 新田義彦 (1993), 日常辞書の語義文参照

- ネットワーク上での単語間距離計算に基づく単語の数値列表現方法, 信学技法, Tec. Repo. IEICE, NLC 93-40, 電子情報通信学会 pp.73-79 (1993, Jul.)
- (1997), 単語ベクトルを用いた多義語の意味推定: 共起ベクトルと定義距離ベクトルの比較, 情報処理学会, NLP-102-7, pp.49-56 (1997, Jul.)
- 久光徹, 新田義彦 (1991), 日本語形態素解析における非サ変動詞の分割単位について, 情報処理学会論文誌, Vol.32, No.3, 情報処理学会 (1991, Mar.)
- (1994) 尤度付き形態素解析用の汎用アルゴリズムとそれを用いた尤度基準の比較, 電子情報通信学会論文誌, Vol. J 77-D-II, No.5, pp.959-969 (1994, May)
- (1998), 文書走査を用いた複合名詞解析, 言語処理学会論文誌, Vol.5, No.4, 言語処理学会, pp.35-60 (1998, Apr.)
- 松岡潤, 新田義彦 (1978), 語彙出現率に関する推定法, 情報処理シンポジウム論文集, 情報処理学会 (1978)
- 吉村賢治 (2000) 自然言語処理の基礎, サイエンス社 183P
- Ikehara, S. et al. (2002) 池原悟, 佐良木昌, 宮崎正弘, 池田尚志, 新田良彦, 白井諭, 柴田勝正, 類推思考の原理に基づく言語の意味的等価変換方式, 人工知能学会論文誌, (2002)
- IPAL (1987), 計算機用日本語基本動詞辞書 IPAL, 情報処理振興事業協会 (1987)
- Nitta, Yoshihiko (1988), 新田義彦, 自然言語理解の基礎: 意味論と語用論, 情報処理, Vol.30, No.10, 情報処理学会 pp.1182- (1988, Oct.)
- Saraki, M. and Nitta, Y. (2011), 英語複合前置詞の諸形態と日本語複合辞との対照—明晰・精密な表現への志向, 電子情報通信学会 (IEICE) 思考と言語研究会 (TL研) 研究会研究資料 (2011-2-5) 「文型と意味」特集
- Aho, Alfred V., Hopcroft, John E. & Ullman, Jeffrey D. (1976) The Design and Analysis of Computer Algorithm, Addison-Wesley
- Bentivogli, L. and Pianta, E. (2005) Exploiting Parallel Texts in the Creation of Multilingual Semantically Annotated Resources: the MultiSemiCor Corpus. *Natural Language Engineering* 11(3): 247-261
- Brown, P. F. et al. (1993) The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational Linguistics* 19(2): 263-311
- Chomsky, Noam (1964), Syntactic Structure, Mouton and Co.
- Church, Kenneth W. (1983) A Finite-state Parser for Use in Speech Recognition, Proc. of ACL-83, 21st Meeting of the Association for Computational Linguistics
- Eilenberg, S., Mac Lane, S. (1945) General Theory of Natural Equivalences, Trans. Am. Math. Soc. No.58 (1945) pp.231-294
- Esterbrook, Steve (1999) An Introduction to Category Theory for Software Engineers, Dept. of Computer Science, Univ. of Toronto, <http://www.cs.toronto.edu/~sme/presentations/cat101.pdf> S
- Freyd, P. (1964) Abelian Categories: An Introduction to the Theory of Functors, New York, S Harper and Row
- Goguen, Joseph A. (1989) A Categorical Manifesto, Technical Monograph PRG-72, Oxford University Computing Laboratory, Programming Research Group (March 1989)
- Grice, H. P. (1975) Logic and Conversation, in: P. Cole and J. L. Morgan (eds.) *Syntax and Semantics*, vol.3, New York, Academic Press, pp.41-58
- Hobbs, J. R., Stickel, M., Appelt, D. & Martin, P. (1993), Interpretation as Abduction, Artificial Intelligence, Vol.63, No.1-2, pp.69-142 (1993) Also published as SRI International Artificial center Technical Note 499, December 1990
- Hopcroft, J. E. & Ullman, J. D. (1970) Formal Languages and their Relation to Automata, Addison Wesley (1970)
- Hornby, A. S. (1978) *Guide to Patterns and Usage in English*, Oxford University Press (Japanese Translation: Kenzo Ito.1978. *An English Model, Usage*. Oxford University Publication Office)
- Ikehara, S. et al. (1997) *Japanese Lexical Compendium*, Iwanami Shoten, Tokyo, Japan
- Karttunen, Lauri (1994), Constructing Lexical Transducers,

- Proceedings of COLING-94, Papers presented to the 15th International Computational Linguistics, Kyoto (1994)
- Kinyon, A. (2001) A Language-Independent Shallow-Parser Compiler, *Proc. 39th ACL Ann. Meeting (European Chapter)*: 322-329
- Koehn, P (2010) *Statistical Machine Translation*, Cambridge University Press, 433p
- Koskenniemi, Kimmo (1990), Finite-state Parsing and Disambiguation, Proceedings of COLING-90, Papers presented to the 13th International Computational Linguistics, Vol.2, pp.229-232, Helsinki (1990)
- Lehmann, W. P. (1980) *The METAL System*, Linguistic Research Center, University of Texas, Texas, USA
- Lehrberger, J. J. (1978) *Automatic Translation and the Concept of Sublanguage*, Groupe de Recherché en Traduction Automatique (TAUM), Université de Montreal, Canada
- Mac Lane, Saunders (1997) Categories for Working Mathematician, Springer-Verlag, New York LLC. (March 27, 1997)  
(訳本: S. マクレーン 著, 三好博之, 高木理 訳, 圏論の基礎, 丸善出版 (2012-3))
- Maimon, O and Rokach L (eds.) (2010) *Data Mining and Knowledge Discovery Handbook, Second Edition*, Springer Verlag 1285p
- Main, Michael G. and Benson, Dravid B. (1983), Denotational Semantics for “Natural” Language Question-Answering Programs, *American Journal of Computational Linguistics*, vol.9 no.1 (1983) pp.11-21
- Marcus, M. P. (1980) *A Theory of Syntactic Recognition for Natural Language*. The MIT Press
- Maurice, Gross (1972) *Mathematical Methods in Linguistics*, Prentice Hall Inc.
- (1989) The Use of Finite Automata in the Lexical Representation of Natural Language, *Lecture Notes in Computer Science*, No.377, Springer Verlag
- Mihalcea, R. and Simard, M. (2005) Parallel Texts. *Natural Language Engineering* 11 (3): 239-246
- Moon, R. (1987) The Analysis of Meaning, in: (Sinclair (ed.), 1987) Chapter 4:86-103
- Munday, J. (2008) *Introducing Translation Studies*, Taylor & Francis Group : 訳本 ジェレミー・マンデイ 著, 鳥飼玖美子 (監訳) 翻訳学入門, みすず書房 (2009)
- Nagata, M. (2003) Natural Language Processing by Statistic Model, in: Amari, S. et al (eds.), *Statistics for Language and Psychology*, Chapter 2: 59-128 (甘利俊一 他 編 統計科学のフロンティア 第10巻 言語と心理の統計 第2章 永田昌明, 確率言語モデルによる自然言語処理, 岩波書店 (2003))
- Nakamura, Y. (1983) *How far can we go in translation? (翻訳はどこまで来たか?)* Japan Times, Tokyo, Japan
- Nida, E. and C. Taber (1969) *The Theory and Practice of Translation*, Leiden: E. J. Brill (1969)
- Nierenburg S. et al. (eds.) (2003) *Readings in Machine Translation*, The MIT Press
- Nitta, Y. (1993) Referential Structure: A Mechanism for Giving Word-Definitions in Ordinary Lexicons. in: *Language, Information and Computation*, LSK (Linguistic Society of Korea)
- (2002a) A Study of Semantic Typology Patterns and their Transformations, *Economic Review of Nihon University*, 71 (4) Nihon University, Tokyo:131-155
- (2002b) Problems of Machine Translation: From a Viewpoint of Logical Semantics, *Economic Review of Nihon University*. 72(2) Nihon University, Tokyo: 23-42
- (2002c) A Study of Descriptive Language for Sentence Patterns, *Economic Review of Nihon University*. 72(3) Nihon University, Tokyo: 35-59
- Nitta, Y. et al. (1982) A Heuristic Approach to English-into-Japanese Machine Translation. in: J. Horecky (ed.). *Proc. COLING 82 (at Prague) (= Proceedings of the 9th International Conference on Computational Linguistics)*, North Holland Publishing Company: 283-288
- (1984) A Proper Treatment of Syntax and Semantics in Machine Translation, *Proc. of COLING 84 (at Stanford)* (= *Proceedings of the 10th International*

- Conference on Computational Linguistics*), Association for Computational Linguistics: 159-166
- Nitta, Yoshihiko (1986a), Idiosyncratic Gap: A Tough Problem to Machine Translation, *Proc. Comp. Linguistics, COLING'86 ACL (Assoc. Comp. Ling.)* (1986)
- (1986b), Problem of Machine Translation Systems: Effect of Cultural Differences on Sentence Structure, *Future Generation Computer System, Vol.2, No.2, North-Holland* (1986)
- (1986c), Machine Translation: A Problem of Understanding, *Japan Computer Quarterly, No.64, JIPDEC* (1986)
- (1987), Natural Language Understanding Viewed as Human Interface, *Proc. Human Interface, SICE (Society for Information and Control Engineering)* (1987)
- Nitta, Yoshihiko, et al. (1982) A Heuristic Approach to English-into-Japanese Machine Translation, in: Horecky, J. (ed.): *Proc. COLING 82 (at Prague)* (=Proceedings of the 9th International Conference on Computational Linguistics), North Holland Publishing Company (1982) pp.283-288
- (1984): A Proper Treatment of Syntax and Semantics in Machine Translation, *Proc. of COLING 84 (at Stanford)* (=Proceedings of the 10th International Conference on Computational Linguistics), Association for Computational Linguistics (1984) pp.159-166
- Perila, Fernando C. N. & Wright, Rebecca N. (1991) Finite-State Approximation of Phrase-Structure Grammars, In 29th Annual Meeting of the Association for Computational Linguistics, pp.246-255, Association for Computational Linguistics (1991)
- Pim, A. (2010) *Exploring Translation Theories*, Routledge, Taylor & Francis Group : 訳本 アンソニー・ピム著, 武田珂代子 (訳) 翻訳理論の探求, みすず書房 (2010)
- Roche, Emmanuel & Schabeseds, Yves (1997) *Finite-State Language Processing*, The MIT Press
- Saraki, M. and Nitta, Y. (2005), The Semantic Classification of Verb Conjunction in the “Shite” Form (日本語文のシテ型用言接続が英語で連体修飾に転化する現象), *Proceedings of Spring IECEI Conference, IECEI Japan*
- (2008), *Regular Expression and Text Minig* (正規表現とテキスト・マイニング), *Second Printing*, 明石書店, 312p
- Slocum, J. (1985) Machine Translation — Its History, Current Status and Future Prospects, *Computational Linguistics*, 11(1)
- Tohru Hisamitsu, T., & Nitta Y. et al. (1996), Analysis of Japanese Compound Nouns Using Direct Text Scanning, *Proc. COLING'96 (the 16th International Conference on Computational Linguistics)* Vol.1, pp.550-555 (1996, Aug.)
- Woods, W. A. (1970) Transition Network Grammars for Natural Language, *Comm. of the ACM, Vol.13, No.10*, pp.591-606
- (1980) Cascaded ATN Grammars, *Computational Linguistics, Vol.6, No.1*
- Yoneda, Nobuo (1954) On the Homology Theory of Modules, *J. Fac. Sci. Tokyo, Sec. I .7* pp.193-227
- (1960) On Ext and Exact Sequences, *J. Fac. Sci. Tokyo, Sec. I .8* pp.507-526
- Yoshiki Niwa, Yoshihiko Nitta et al. (1994), Co-Occurrence Vectors from Corpus vs. Distance Vectors from Dictionaries, *Proc. COLING'94 (15th International Symposium on Computational Linguistics at Kyoto)* pp.304-309
- (1995), Statistical Word Sense Disambiguation Using Dictionary Definitions, *Proc. Natural Language Pacific Rim Symposium, KISS (Korean Information Science Society) & IPSJ (Information Processing Society of Japan)* pp.665-670 (1995, Dec.)
- (1997), Topic Graph Generation for Query Navigation: Use of Frequency Class for Topic Extraction, *Proc. NLP'97 (Natural Language Pacific Rim Symposium '97 at Phuket)* pp.95-100 (1997, Dec.)

以上