

# 非線形な言語表現と文型パターンによる意味の記述

池原 悟

鳥取大学工学部知能情報工学科

〒 980-8552 鳥取市湖山町南 4-101 ikehara@ike.tottori-u.ac.jp

**あらまし** 対訳文から非線形な表現構造を取り出すための判断基準と文型パターン化の基本的な方法について検討した。文型パターン翻訳方式を実現するには、大規模な文型パターン辞書の開発が必要であり、対訳コーパスから半自動的に文型パターンを生成することが期待される。本稿では、非線形な言語表現の意味と構造を一体的に文型パターンで記述することを念頭に、まず、表現の全体と部分に対する線形性と非線形性の定義を行い、表現構造と意味の関係の線形性と非線形性を判定するときば意味の基準として、「言語規範としての意味」を用いること、および、その意味を記述する方法として工学的立場から目的言語の表現を使用することを提案した。次に、このような方法で作成される文型パターンの性質と意義を明らかにした。最後に、表現構造の線形性、非線形性が表現要素の選び方に依存していることに着目して、日本語表現の非線形な構造を段階的に文型パターン化する方法を示した。

キーワード : 言語表現、線形表現、非線形表現、本来の意味、規範としての意味、文型パターン、機械翻訳

## Non-linearity of Linguistic Expressions and their Meaning Descriptions by Sentence Patterns

SATORU IKEHARA

Faculty of Engineering, Tottori University

**Abstract:** Conventional approaches in Natural Language Processing have employed the principle of Compositional Semantics assuming the linearity of the relation between linguistic expressions and their meanings. However, the most actual expressions are not linear and it causes the limitation of Natural Language Processing. In order to address this problem, this paper clarifies the standard to decide whether a linguistic expression is linear or nonlinear and proposed the method to describe nonlinear sentences structure by using sentence patterns. First, the difference of an actual meanings and the meanings as a linguistic norm was defined for linguistic expressions. And linearity was defined for the relation between expressions and the meanings as linguistic norms. Second, assuming that the meaning of Japanese expressions are defined by using English expressions, the standard of judging linearity and nonlinearity of Japanese expressions are proposed. Finally, based on these definitions, the method of describing the structure of nonlinear linguistic expressions by sentence patterns

Key Words: Linguistic Expression, Linear Expression, Sentence Pattern, Machine Translation

### 1. はじめに

従来、機械翻訳などの自然言語処理の研究では、要素合成法を基本とした研究開発が行われてきた。これは、言語表現の意味が線形であると仮定したものと見える。しかし、多くの言語表現は非線形であり、このことが、従来の言語処理の限界を生む原因となっていたと思われる<sup>1)</sup>。

例えば、トランスファー方式を基本とする従来の機械翻訳システム<sup>2), 3)</sup>では、原文の構文構造を目的言語の構造に変換する過程と原文の各要素を翻訳する過程を持ち、両者から得られた結果を合成することによって目的言語の表現が生成される。これは、構文構造と表現の意味が別々に処理されることを意味している。非線形な言語表現においては、言語表現全体の意味はそれが構成要素に分解される過程で失われ、目的言語を合成する過程で復元されなくなることが問題である。

この問題を解決する方法として、古くから意味のまとまる表現の単位をパターン化した「パターン翻訳方式」が試みられてきた。この方式は「テンプレート方式」とも言われる方法で、トランスファー方式と組み合わせて使用される例が多い。最近では、「翻訳メモリ」とも併用される傾向にある。

しかし、いずれのシステムも組み込まれている文型パターンは 200 ~ 300 件程度にとどまっており、その用途も、特定の専門分野など限定的である。これは、大規模な文

型パターン辞書を開発するには、多大なコストが見込まれるだけでなく、パターン相互の意味的相互干渉が問題となるためと考えられる。コストの問題を解決するには、大規模な対訳コーパス等からの半自動的に文型パターン化が行えるよう文型パターン作成の基準を明らかにしていくことが大切であり、また、文型パターンの意味的な相互独立性を実現するには、文型パターンの適用条件が的確に表現できる方法を実現する必要がある。

ところで、線形な表現は、従来の要素合成法によって翻訳可能と見られるから。文型パターン方式において、パターン化すべき表現は、非線形な表現構造を持つものだと考えられる。言語表現の非線形性の問題を解決する方法としては、既に、構造と意味の関係を考慮した多段階翻訳方式が提案されている<sup>4)</sup>。この方式は、話者の対象認識を表現する客体的表現と話者の感情や意志を直接的に表現する主体的表現を分離、融合する仕組みと客体的表現の構造を意味を失わないように目的言語に対応づける仕組みから構成されており、後者には、結合価文法が使用されている。

その結果、単文のレベルでは精度の高い意味解析が実現されるが、この方式では、複文(埋め込み文を持つ文)、重文(接続のある文)の持つ非線形性が扱えないこと、また、原文に対して単一の目的言語表現が対応づけられる仕組みであるため、文脈に応じた表現選択ができ

ないことが問題となっている<sup>1</sup>。

これらの問題を解決するため、最近、さらに言語表現の意味的類型化を基本とする意味的等価変換方式が提案された<sup>5)</sup>。この方式では、言語表現を意味的に類型化する方法として、非線形な言語表現の構造を文型パターンで表現する方法が提案されている。この方式を実現するにも、与えられた言語表現が線形であるか、非線形であるか、もしくは、どの要素が線形要素であり、どの要素が線形でないかを判断する基準が問題となる。

そこで、本稿では、この問題を解決するため、言語表現の線形性と非線形性を判断する方法を明らかにし、非線形な言語表現の構造を文型パターンによって記述する方法を提案する。言語表現の線形性と非線形性は、表現とその意味の関係に関するものであるため、言語表現の意味をどのように定義するかが問題となる。そこで本稿では、まず、言語表現が持つ「本来の意味」と「言語規範としての意味」を定義した後、論理式での説明を参考に、言語表現と「言語規範としての意味」がいかなる関係にあるとき、その表現は線形といえるかを明らかにする。次に、機械翻訳への適用を念頭に、工学的立場から、原言語表現の意味を目的言語の表現で表すことを考え、原言語表現の線形性と非線形性を判定する基準を提案する。最後に、この基準の相対性に着目して、非線形な言語表現の意味を文型パターンによって表現する方法を提案する。

## 2. 線形性と非線形性の基本検討

### 2.1 非線形な言語表現の必然性

従来の自然言語処理では、言語が線形であることを前提とした要素合成法を基本にさまざまな研究開発が行われてきたが、以下に述べる2つの理由で言語表現は本質的に非線形であると考えられる。

第1の理由は、表現多次元性と表現媒体の1次元性に起因する。話者の認識する対象世界は、三次元的な空間と時間軸を加えた四次元の世界である。話者はこのような4次元の対象の認識に話者自身の感情や意志の加わった認識を表現するため、表現内容は多次元的である。これに対して、このような多次元的な認識を対応づけようとする表現媒体は、空気の粗密波か文字列であり、いずれも一次的である。表現の枠組みとしては、あちらを立てれば、こちらが立たず関係が生じ、例外の多い、すなわち非線形性を持った複雑な規則とならざるを得なくなると考えられる。

もう一つの理由は、言語が社会的に開かれたシステムであることに起因する。言語は社会的に自然発生した表現の枠組みであり、オープンなシステムである。既存の枠組みで表現できないような思想や考えが生まれるとそれを表現するための工夫が行われ、既存の枠組みが変更されたり拡張されたりするなど、常に変化している。慣用表現もそのような過程で生まれたものと見ることができる。

このような既存の枠組みは、例えそれが初期段階では線形で整合した体系であったとしても表現能力を拡張するため、数多くの非線形な表現構造を持ち込まざるを得なかったものと考えられる。

以上のように、言語の非線形性は必然的なものであり、従来の原語処理の方式的限界を克服するためには、言語表現の意味と構造の関係を明確にし、非線形な表現構造を扱うことができるような方法論を確立することが必要と考えられる。

### 2.2 フレーゲの原理と表現の線形性

自然言語表現の線形性と非線形性を識別する問題では、論理学における論理式の線形性に関する議論が参考となる。論理学では論理式全体の意味と部分の意味の関係について、「フレーゲの原理」が成り立つとされている。この原理は、論理式の線形性を示すもので、各要素が論理結合子で結合された論理式において、各要素(単文)の真理値の任意の組が与えられると、結合子の選択によって生成された論理式(複合文)全体の真理値は完全に決定される<sup>6)</sup>ことを意味している<sup>6)</sup>。「論理結合子は、全体の外延を諸部分の外延の関数として決定する機能」を持つことから、「結合表現の意味はその諸部分の意味の関数である」と言い換えることができる。

この考えを言語表現に当てはめ、線形な言語表現は、「全体の意味が部分の意味の和で表される表現」、すなわち、「全体の意味はその構成要素に還元できる表現」であり、「重ね合わせの原理」が適用できる表現だと考える。逆に、非線形な言語表現は、「全体の意味が部分の意味の和で表されないような表現」、すなわち、「単語や句の意味が合算できない表現」だとする。

さて、論理学では表現全体の意味と部分の意味はいずれも真理値で表されるものとしているが、自然言語の表現においては、表現全体の意味と部分の意味をどのように定義するかが問題となる。また、併せて言語表現では、「意味を合算する」とはどういうことを明確にすることが必要である。そこで次節では言語表現の持つ意味を定義する方法を述べる。

### 2.3 言語表現の意味の二重構造

#### (1)言語表現の本来の意味

言語表現の意味については、古くから、多数の説があり、それらは、「対象意味論」、「認識意味論」、「形式意味論」、「解釈意味論」のほか、それらを折衷した意味論や「関係意味論」に分類される<sup>5)</sup>。本検討では、三浦つとむの言語過程説の「関係意味論」<sup>7)</sup>の立場から、言語表現の意味を考える。

三浦文法では、「言語表現の意味」は「言語表現とそれに対応づけられた話者の認識との関係」であり、「話者の発話した(実際に使用された)表現」に対してのみ定義されるとされる。以下では、この意味のことを「本来の意味」と称す。

ところで、この定義によれば、言語表現の意味は、それ

<sup>1</sup> 結合価パターン<sup>4)</sup>の方法は、体言と用言の間の意味的に非線形な関係を定義するには適しているが、助詞、助動詞を含む非線形な文型構造の記述には不適である。多段翻訳方式では、助詞、助動詞などは主体的表現として分離しており、このことも、「多段翻訳方式」で残された問題点と言える。

ぞれの表現に固有で客観的なものであり、解釈によって変わるものではないため、自然言語処理にとって都合がよい。しかし、与えられた言語表現の一部でも他の要素に置き換えたりすると、話者の認識との関係(すなわち意味)は失われるため現実の言語表現はすべて非線形形だと言うことになる。

## (2)言語表現の規範としての意味

ところで、三浦の指摘する「概念の二重構造」に従えば、言語表現は、一般に上記のような「本来の意味」のほか、「規範としての意味」を持つと考えられる。「規範としての意味」は、言語上の約束から見たときの表現と表現内容との関係である。話者の認識(すなわち「本来の意味」)とは相対的に独立しており、現実には使用されていない言語表現に対しても定義される。

三浦は言語において、「話者の認識の中に概念が形成される過程」と形成された概念を表現に結びつける過程」を分け、「話者の認識として形成された概念」と「言語規範としての概念」は必ずしも一致するものではないとし、話者は自分の認識として形成された概念を相手に伝えるため、社会的に共有された言語規範としての概念の中から自分の概念と最も近いものを選択して表現に結びつけることを指摘している。

この考えを背景に池原<sup>8)</sup>は、言語表現には単語や句、節などさまざまな表現の単位が存在することに着目して、これらの表現も「概念」の表現であるとし、単語で表現される概念を「単一概念」、句や節などの表現で表される概念を「複合概念」に分類している。また、この区別は言語表現過程における制約から生じるものであり、相対的な分類であることを指摘している。

表現過程において、話者は、自分が伝えたい概念に対して、適切な「言語規範としての概念」を探し、それを表すための単語を使用しようとするが、自分の使用する言語に適切な概念が存在しないとき、より低位の概念に分解して表現する。このようなとき使用されるのが句や節の表現である。従って、言語表現では、単語だけでなく表現そのものが話者の認識で形成された概念と対応関係を持つことになる。

以上の議論によれば、言語表現によって表される意味とは、概念化された話者の認識、すなわち、概念であり、その概念は、規範としての概念を使用することによって近似的に表現される<sup>1)</sup>。そこで、以下では、「言語表現の意味」と「言語表現の表す概念」を同じ意味で使用する。

そこで、いずれの概念(意味)を機械翻訳の対象とするかであるが、言語理解は聞き手が「規範としての概念(意味)」を介して「話者の概念(本来の意味)」を類推することだと言える。この点に着目すると、計算機による意味処理は以下の2つの過程に分けることができる。一つは与えられた表現の表す「規範としての概念(意味)」を特定する過程(意味解析)と言う)であり、第2のステップは「規範としての意味」から「本来の意味」を推定する過程

(意味理解)と言う)である。

このうち自然言語処理において当面重要な課題は第1のステップを実現すること、すなわち、聞き手が話者と共有する「言語規範としての概念」の処理を実現することである。機械翻訳においてこの処理が実現できれば、訳文品質は大幅に向上することが期待できる。すなわち、機械翻訳において、目的言語の表現で表された「規範としての意味」を目的言語の「規範としての意味」に対応づけることができれば、「話者認識としての意味」は目的言語側の人の理解に任ずることができるからである。また、「規範としての意味」は、文脈や聞き手の立場などによる解釈の問題とは相対的に独立しているため、処理しやすい対象と言えることも利点である。

以上から、以下では言語表現の線形性と非線形性を判定するための基準として「言語規範としての意味」を使用する。

## 2.3 言語表現の意味の記述方法

さて、個々の言語表現に対して「規範としての意味」を具体的に定義する方法であるが、世界の言語をカバーするような共通の中間言語が設計できれば、それを使用すればよい。しかし、そのような言語は実現不能である。人間用の辞書では、語や表現の「規範としての意味」が、別の言語表現を使用して記述されている。しかし、計算機では、これで意味が理解されることにならない。現在の計算機から見れば、どのような書き方も単なる記号にすぎないから、逆に、どのような書き方でも、それが排他的な体系となっていればよいと言える。そこで、本稿では、原言語の単語や表現の意味を目的言語の表現で記述することにする。これは、機械翻訳システムにとって大変好都合である。

例えば、「私は彼に手紙を送る」の意味は、「I send him a letter」だと定義し、「猿も木から落ちる」の意味を「Any body can make a mistake」だと定義する。

このように言語表現の意味を他の自然言語で記述する場合、通常は使用した言語側での意味的な多義が問題となるが、機械翻訳の場合は、翻訳結果の意味を理解するのは目的言語側の人間であるので、あまり問題にはならないと期待できる。

## 3.言語の線形性と非線形性の定義

前章では、言語表現の意味として「規範としての意味」を取り上げること、また、それを目的言語の表現を用いて記述することを述べた。本章では、言語表現の「部分の意味」と「全体の意味」の関係によって言語表現の線形性、非線形性を定義し、両者を判定する方法を示す。

### 3.1 線形性と非線形性の定義案

2.1節では、言語表現において、「意味を合算できる要素」が線形な要素であることを述べた。ここでは、「意味を合算する」を「意味を引き算する」に置き換えて考える。

<sup>1)</sup> 言語間の翻訳において、厳密な意味で意味的に正解の翻訳は存在せず、翻訳はあくまで近似であると言われているが、その意味では、原言語表現も話者の認識に対する近似である。このように言語が近似的な表現であることは、概念の二重性の議論から説明されるもので言語の本質的な特徴である。「本来の意味」と「規範としての意味」を明確に区別することは言語の意味処理方式を検討する上でも、大変重要と考えられる。

例えば「私は彼に手紙を送る = I send him a letter」では、日本語側から「彼に」を削除したときの意味は、英語側から him を削除したときの意味と同じになる。従って、「彼に」は、引き算ができる要素で、すなわち線形要素だと言うことになる。

しかし、「私は」を削除すると対応する英語表現が成り立たないし、「送る」を削除すると日本語表現も成り立たないから、これらの要素は非線形要素である。また、「手紙を」を削除したときは、「I send him」とはならず「I send to him」になるから、「手紙を」も非線形要素となる。すなわち線形要素は「彼に」だけである。

この定義では、線形要素はいかにも限定的である。第2章の議論から、フレーゲの原理は、「複合表現の意味はその諸部分の意味の関数である」ことを意味し、「式や記述の一部をそれと等価なものに置き換えても全体の評価値や意味は変わらない」ことが導かれる。そこで、本検討では、まず、 $\rightarrow$ に従って、表現中の要素(部分的表現でも良い)を同義語に置き換えても全体の意味が変わらないとき、その要素は線形要素だと考え、次のように定義する。

#### 定義 1: <表現要素の線形性と非線形性>

言語表現を構成要素(単語、句、節、等)のうち、それと同等の意味を持つ他の要素に置き換えても元の言語表現全体の意味が変化しないような要素を「その言語表現に対する線形要素」と言う。逆に、言語表現全体の意味が変化するとき、その要素を「その言語表現に対する非線形要素」と言う。

次に、この定義を用いて表現全体の線形性を以下の通り定義する。

#### 定義 2: <言語表現の線形性と非線形性>

線形要素のみから構成される言語表現を「線形な言語表現」と言い、1つ以上の非線形要素を有する言語表現を「非線形な言語表現」という。

### 3.2 線形性と非線形性の定義の見直し

ここで、無限とも言える言語表現を線形要素に着目して縮退させることを考える。しかし、前節の定義では、線形要素は、意味的に置き換え可能な同義語を持つ要素に限定されることから、それを変数化によって縮退させたとしても、やはり無限とも言える表現構造となりそうである。非線形な表現(もしくは表現構造)は、要素に分解すると全体の意味が失われるから、人間がそれを使用するためには、表現(もしくは表現構造)全体を記憶している必要があるが、有限の記憶能力の人間には、それは不可能だと言うことになる。

ところで、人間の概念形成過程と表現過程<sup>3)</sup>から見て、概念は単一の単語だけでなく、複数単語からなる表現でも表されること(前者は「単一概念」、後者は「複合概念」と称される)については前に述べた。この議論では、概念は、全体として階層関係をもつことが指摘されている。

これに着目すると、通常、文のような表現で表された概念にも上位概念が存在し、線形要素の変数化などによる

言語表現の汎化は、上位概念に対する表現への縮退を意味することになる。

例えば、単語で表現される「比較」、「因果関係」などの概念は、「単一概念」であるが、その配下には、さまざまな表現形式(これには「複合概念」が対応する)が存在する。このような配下の表現の線形要素を適切な方法で汎化することができれば、上位概念「比較」、「因果関係」等の表現構造が得られる可能性がある。

実際の日本語を自立語と助詞、助動詞などの辞に分け、前者を変数、後者を演算子に対応させて考えると、表現構造は論理式相当する。この場合、変数が線形要素であり、その値域は外延である。しかし、言語表現の場合は、演算子と変数の選び方は任意であるから、本検討では、線形要素を変数化したものを表現の構造と考えることにする。

以上の考えに基づいて、定義 1 を下記の通り拡大する。表現構造全体の線形性と非線形性に関しては、定義 2 と同じである。

#### 定義 1改: <表現構造における線形要素>

特定の概念(複合概念)を表わす表現(又は表現構造)の要素のうち、他の要素(意味の異なる要素でも良い)に置き換えても表現(又は表現構造)全体の意味(複合概念)が変わらないとき、その要素をその表現(表現構造)の線形要素」と言う。

これは、元の表現の意味が多少失われるとしても、上位概念の意味を失わない範囲での変数化を可能にしようと言うものである。厳密な意味において機械翻訳は、近似であることから考えて、この近似は、訳文品質を低下させるものではないと考えられる。

上記の定義によって、線形要素では、目的言語表現に意味的に対応する要素が存在し、両者を意味の異なる別の要素に置き換えても、表現全体の意味的な対応関係は保たれることになる。

これを原言語と目的言語の対訳文型で考えると、要素置き換えによって対応する目的言語の表現構造が変化しない限り、その要素を線形要素とすることである。

これは、言語表現の構造の持つ意味を考えることに相当する。文献<sup>3)</sup>によれば、「比較」、「因果関係」などの複合概念に対して、各言語はそれを表現するための構造的な仕組みを持つ。通常の言語表現では、構成要素が異なる意味の要素に置き換えられても表現の構造が変化しない限り、表現構造の持つ複合概念としての意味は変わらない場合が多い。

従って、線形要素を持つ言語表現は多いが、全体が線形であるような文は殆ど存在しないと予想される。逆にすべての要素が非線形な文は、成句、ことわざのような特殊な表現に限られることから、大半の文は、線形要素と非線形要素を含む文であることが予想される。

### 3.3 線形要素の重要な性質

定義 1改と定義 2 によれば、非線形な表現構造の処理方式を考える上で重要な点は、以下の3点である。

#### (1)線形要素の制約条件

まず第1の点であるが、定義1改では、置き換え可能な要素を線形要素としているが、これは実際にどんな要素に置き換えても良いことを意味しない。

例えば、「私は彼に手紙を送る」の例では、「私は」を「彼女は」に変えたり「手紙を」を「贈り物を」に変えたり「送る」を「あげる」に変えたりしても、その意味は、「I send him a letter」の対応する要素の「I」「She」、「letter」「gift」、「send」「give」のような置き換えで表現されるから、いずれも線形要素だということになる。しかし、格要素「私は」を動詞「食べる」に置き換えるようなことはできない。全体の意味は、英語側の対応する要素の置き換えでは表現できなくなるどころか、日本語側でも意味をなさなくなる(そもそも言語規範としての表現の意味が定義できない)。あくまで、置き換え先となる要素が存在するということであり、置き換え先となる要素に制約があっても問題ない。

なお、「猿も木から落ちる = Any body can make a mistake」などともなると、どの要素も置き換え不能となるから、いずれの要素も非線形だと判断される。

### (2)要素の選び方と全体の線形性

第2は、表現要素の線形性と表現全体の線形性の関係である。定義2によれば、すべての要素が線形の場合に限り表現は線形だとしている。これによれば、与えられた表現をどのような要素に分けるかによって、全体の表現が線形になったり非線形になったりする場合が考えられる。

例えば、「私は彼に手紙を送る。 = I send him a letter」の表現において、「私は = I」、「彼に = him」、「手紙を = a letter」、「送る = send」の4つの要素に分解すると、どの要素も置き換えても、その意味は格要素に対応する英語表現の置き換えによって表現され、英語表現の構造は変わらないから、全体が線形だと言える。しかし、「私」、「は」、「彼」、「に」、「手紙」、「を」、「送る」のように全単語にばらすと、「は」、「に」、「を」は置き換え不能(但し、「は」は「が」に、「に」は「へ」に置き換えができそうだが)となり、表現全体は非線形だということになる。

このことから、線形、非線形の定義は相対的である。しかし、指定された要素の線形、非線形の区別は、その要素が意味的代替要素を持つか否かによって一意に決定できる。代替要素を持てば、それを値域とする変数に置き換えることが可能となるため、実用上問題は発生しない<sup>1)</sup>。

### (3)表現全体の線形性と要素自身の線形性

第3は、線形、非線形の区別は表現の部分と全体の関係を言うものであり、線形要素だと言ってもその要素自身が線形であることを意味しないことである。線形要素の内部構造は非線形であっても良い。

このように、線形、非線形の分類が再帰的な構造を持つことは、非線形な言語表現に対する処理方式を考える上で大変重要な点である。

## 3.4 非線形な表現構造の有限性

### (1)人間の記憶容量と言語表現の多様性

記憶容量が有限の人間が、どのようにしてほぼ無限とも言える言語表現を使いこなせるのかという問題は、古くから知られているプラトンの問題<sup>9)、10)</sup>と見ることができる<sup>2)</sup>。この問題にどのように答えるかは、新しい自然言語処理方式を考える上で大変重要な問題である。

この問題に対して、チョムスキーは、意味記述方法として深層構造を導入し、表層の表現は、深層構造からの変形によって生成されると説明した<sup>3) 11)</sup>。深層構造と変形規則は有限であるが、その組み合わせから無限の表現が生成されると考えれば、有限の記憶容量の人間が無限の言語表現を理解し使いこなせることが説明できる。

しかしこの説明は、意味を変えない変形規則の認めた点で行き詰まり、変形規則を認めない方向に進んだことからこの問題の解決には至らなかった。

そこで、本節では、言語表現の線形性、非線形性を考える観点からこの問題について考える。

### (2)非線形な表現構造の有限性

一般に非線形な表現は、要素合成の方法では生成できないため、表現全体を覚えておかなければ使用することはできない。従って、ほぼ無限とも言える言語表現のすべてが非線形であるとするなら、記憶容量の有限性から見て、人間はこれを使いこなすことはできないことになる。従って、人間がこれを使いこなせると言うことは、言語表現は無限にあるとは言え、非線形な表現構造は有限であることを意味する。

3.1節と3.2節で述べた言語表現の線形性、非線形性の定義によれば、言語表現は線形要素と非線形要素から構成され、そのうち非線形要素を1つ以上持つ表現が非線形な表現構造である。このことから、多くの言語表現は非線形であるとは言え、その中に線形要素を持つことが多いと推定される。

従って、線形要素には値域があり、代替できる要素を持つこと、また、非線形構造自身の線形結合によってより大きな表現が生成されると考えれば、記憶容量の有限性と言語表現の無限性の問題は矛盾なく説明できる。

この説明は、非線形な表現構造の数の有限性を示唆するものであり、非線形な言語表現の構造を文型パターン化する方法について実証的な研究を促している。

## 4.非線形な表現構造を記述する方法

前章までの議論に基づいて、本章では、言語表現の非線形な構造を表現するための文型パターンを定義し、

<sup>1)</sup> 意味的に代替可能な要素は、0個(値域が NULL)でも良い。そのような要素は、日本語表現にあってもなくても英語側の表現は変化しないような要素である。

<sup>2)</sup> プラトンの問題は、「子供はなぜ幼少にして言葉をマスターできるか」と言う問題で、「人間はわずかの経験(少ない入力)でなぜこんなに賢くなれる(沢山の出力が出せるか)か」という問題とも解釈される。ここでは、この問題を「有限の能力と無限の表現能力の関係の問題」に置き換えているが、デカルトの問題(言語能力は1か1かにして使用されるか)と考えることもできる。古い問題ではあるが本質的であり、計算機による言語理解を実現する上で、避けて通れない問題の一つである。

<sup>3)</sup> チョムスキーは、プラトンの問題に答えるため、生得的な言語能力の存在を仮定し、深層構造と生成規則の組み合わせで説明しようとしたが、その後、生成規則代わる方法として普遍文法の考えを打ち出している。

現実の言語表現の用例の持つ非線形な構造を文型パターンで記述する方法を示す。

#### 4.1 文型パターン化と変数化のレベル

##### (1)文型パターンの定義

原言語表現の意味を目的言語表現によって記述することについてはすでに述べた。この場合、前章の議論から、原言語表現のある要素が線形要素であるのは、以下の2つの場合である。

一つは、原言語側の表現に対して、目的言語表現側にそれに対応する要素がある場合で、もう一つは、目的言語表現側にそれに対応する要素はないが、原言語側のその要素を削除しても対応する目的言語表現は変化しない場合である。このことに着目して、文型パターンを定義3を設ける。

##### 定義3:日本語文型パターン

日英言語の表現対において、線形要素を変数記号に書き換えた表現を文型パターンと呼ぶ。

但し、変数記号化される要素は原言語側、目的言語側で異なる文法的属性を持つ要素でも良い。なお、目的言語側に対応する要素のない場合は、元の目的言語の表現がそのまま文型パターンになる。

##### (2)線形要素の3レベルの変数化

ところで 3.3(2)で述べたように、表現要素の線形性と非線形性を判断するには、あらかじめ表現要素の選び方を決める必要がある。そこで、本稿では、言語表現の文法的な構成単位に着目して、単語、句、節を表現要素とした3種類の変数化を考える。

##### <単語レベル>

表現に含まれる名詞、動詞、形容詞、副詞などの自立語を表現要素とし、そのうちの線形なものを変数化する。助詞や助動詞は変数化されず字面のまま残されるか、もしくは、字面グループを表す関数を設けそれによって記述する。これらの要素を論理結合子と見ると、文型パターンは論理式に類似した表現と見ることができる。

##### <句レベル>

名詞句、形容詞句、動詞句、副詞句など句を要素としそのうちの線形なものを変数化する。

##### <節レベル>

連体節、連用節等を要素と考え、そのうちの線形なものを変数化する。

#### 4.2 線形要素と非線形要素の記述

文型パターンの記述においては、線形要素を記述する変数と非線形要素を記述する字面が基本となるが、例えば、テンス、アスペクト、モダリティ、格変化など述部の語尾や格助詞相当語などに関する字面など、特定の字面を指定するための関数(形式指定関数)や記号などを設けそれを使用しても良い。そこで、文型パターンは、表1で示すような字面、記号、変数、関数の4種類の要素によって記述することとし、以下、それらを用いた記述方法の原則を示す。

##### (1)線形要素の変数化の原則

文型パターンにおいて表現構造の中の線形要素は、「変数」、「任意要素」、もしくは表記の揺らぎを表すための「関数」のいずれかで表現される。ここでは、線形要素を変数化する際の注意点を3点示す。

表1. 文型パターンを記述するための文字や記号

| # | 分類 | 説明   |   |
|---|----|--|---|
| 1 | 字面 | 通常の日本語文、英文で使用されるすべての文字、または文字列で記述する。  |   |
| 2 | 記号 | スペース記号とプラス記号   | 「離散記号」"/"と「連鎖記号」"+"を使用する。両者は、それぞれ、をその位置にゼロ個以上の原文任意要素が存在しても良いこと、存在してはいけないことことを意味する。  |
|   |    | 要素選択記号   | 当該位置に使用可能なパターン要素が複数存在することを表す記号で、例えば、表現要素、...が意味的に交代可能な表現要素であるとするとき、日本語文型パターンの該当する位置でこのいずれが使用されても良いことを、(    ... )によって表現する。         |
|   |    | 任意要素記号   | 省略可能な要素を指示する記号で、"# n [    ... ]"の形式で記述する。表現、...は省略可能(交替可能)であることを意味し、# nは、パターン内での省略可能な要素の番号を表す。、...には、字面、変数、関数、及び、それらの組み合わせが使用できる。 |
|   |    | 順序任意要素指定記号   | 指定された複数の要素の順序が任意であることを示す。<br>例)彼女を   彼は   駅まで )送る。  |
|   |    | 位置変更可能要素指定記号   | 位置を変更して良い要素及び変更可能な位置を示す。<br>例)#1)彼の作品は#1(誰が見ても)第1級のものだ。   |
| 3 | 変数 | 変数は、線形要素を表す記号である。表される部分的表現の文法的属性に対応して、単語、句、節を表す変数に大別される。文法的属性をアルファベット大文字の字種で表し、数字は、日本語文型パターン内で出現する変数の順番を数字で表す。例えば、N3(名詞)、V2(動詞)、NP5(名詞句)、CL2(節)など。 |   |
| 4 | 関数 | 変数関数   | 英語文型パターンの記述で使用される関数で、変数名が関数名として使用される。動詞、形容詞、形容動詞を名詞に変換するなど、文法属性の変換機能をもつ。例)N(V1) 動詞 V1 を名詞化した表現を意味する。                              |
|   |    | 字面関数   | 関数名として字面を使用した関数である。引数には、変数が使用され、それによって表される表現が関数名で示される字面を持つことを意味する。例)外は大変暑かったので「past(大変(CL 1))ので」                                  |
|   |    | 要素抽出関数   | 英語文型パターンのみで使用される関数で、引数にバインドされた表現から、主語、目的語などの特定の文要素を取り出す。例)CL1 = Weeatanapple のとき、subj(CL1) = we                                   |
|   |    | 形式指定関数   | 単一の引数を持つ関数で、動詞、動詞句、節を過去形、可能形、否定形、受身形などに変形するなど、引数で指定された表現に対して関数名で指定された構造の表現を示す。例)「登っている」 teiru(V1)                                 |
|   |    | 構文合成関数   | 英語文型パターンのみで使用される関数で、複数の引数によって指定された表現を組み合わせて、so_that 構文、too_for_to 構文など関数名で指定された表現を合成する。   |
|   |    | マクロ関数  | 変数名を関数名とする関数で、変数に対応する表現の構造を定義する。  |

第1は、3.3の(3)の議論から変数化する対象は、あくまで線形要素として取り出せる部分で、変数化した要素の内部が線形構造を持つか否かとは無関係である点である。

第2は、変数化する要素と対応する英語表現の要素が必ずしも同一の文法的属性を持つ必要はなく、日本語表現とその意味を記述した英語表現で、要素の対応関係がとれればよいと言うことである。例えば、名詞Nの日本語要素に対応する英語要素が動詞Vだとすると、英語側の表現構造は、動詞を名詞化する関数N(V)を使用して記述することができるから、その要素は線形要素である。

第3は、あってもなくても意味の変わらない要素も線形要素(原文任意要素)と呼ぶのである。

(2)制約条件の付与

3.3の(3)の議論に基づき、日本語文型パターンの変数に対して、意味的に置き換え可能な範囲を名詞意味属性、動詞意味属性、副詞意味属性などを用いて指定する。これは、論理式においても、変数の値には変域(外延)が存在することに相当する。

なお、与えられた言語表現の意味解析で文型パターンを使用する場合は、意味的に間違っただ言語表現は解析の対象としなくて良いから、通常はこのような制約条件は不要である。しかし、同型異義の文型パターンが存在するときは、この制約条件が必要となる。

4.3 任意化とグループ化

(1)必須要素と任意要素

文型パターンは、「必須要素」と「任意要素」から構成さ

れる。このうち、「必須要素」は、日本語文型パターン内にその要素がないと対応する英語文型パターンが決定できない(すなわち全体が意味をなさない)要素を言う

これに対して、「任意要素」は日本語文型パターン内にその要素がなくても英語文型パターンが記述できるもので、パターン定義に使用するか否かによって、さらに「原文任意要素」と「パターン任意要素」に分類される。「原文任意要素」は、それが削除されても表現全体の意味が変化しないものである。また、「パターン任意要素」は、それを削除すると訳語や訳語挿入位置が不明になるなど、英語による意味の記述が困難になるものを言う

(2)表現要素のグループ化

必須要素とパターン任意要素のいずれかを問わず、助詞、助詞相当語などの表現要素でも、意味的の同等で置き換え可能な文字列が存在する場合がある。そのような要素は、選択記号を使用して、(| |...)の形式で記述する。

4.4 線形とみなせる語順の扱い

言語表現では、語順や出現する位置を変更しても表現構造の持つ意味は変わらないような要素が存在する。線形要素の定義の意味を考えると、このような要素も線形要素と考えることができる。そこで、以下の2種類の記号を導入し、線形要素としての指定を可能とする。

(1)順序任意要素指定記号

格要素など語順を変更しても文型パターンの意味は変化しない要素をグループとして指定する記号で、特に日本語文型パターンの汎用性の向上を狙っている。

表2.文型パターンによる非線形構造の記述例

| 区別     | 日本語文型パターン又は日本文                      | 英語文型パターン又は英文                                   |  |
|--------|-------------------------------------|--|--|
| 単語レベル  | 文型パターン                              | それはN1にあるまじきN2.da。                              | SuchN2 beunseemly for N1.  |
|        | 言語表現例                               | それは学生にあるまじき行為だ。                                | Suchbehaviorisunseemlyforstudents.   |
|        | 文型パターン                              | N1を失ってN2はN3.kako。                              | WithallN2.poss N1dashed,N2V3.past  |
|        | 言語表現例                               | 希望を失って彼は荒れ狂った。                                 | Withallhishopesdashed,hewent wild.   |
|        | 文型パターン                              | N1はN2のN3/ことも#4[ある程度は]V5.                       | (N1 I)canV5whatN2V3#4[osomeexten]t.  |
|        | 言語表現例                               | 次郎の言うこともある程度はわかる。                              | Icanunderstandwhat Jiro says to some extent.                               |
| 句レベル   | 文型パターン                              | あれこれN1.temiru.kakoがN2がNP3.da。                  | All things V1.past, N2's NP3.  |
|        | 言語表現例                               | あれこれ考えてみたがそれがいちばんいい解決策だ。                       | All things considered, that's thebestsolution.                             |
|        | 文型パターン                              | N1の(あと 後)NP2がNP3のためにNP4.kako。                  | After N1 NP2VP4.past toVP3.  |
|        | 言語表現例                               | 洪水のあと何百人もの人が被災者を助けるために力を合わせた。                  | After the flood hundreds ofpeople worked together to help the victims.     |
|        | 文型パターン                              | N1はNP2もV3.hiteiとはNP4.gimuことだ。                  | It is AJP(VP4) that (N1 you) shouldV3.not NP2.                             |
|        | 言語表現例                               | 総理大臣の名前も知らないとはまことに哀れむべきことだ。                    | It is really pitiable that you should not know the Prime Minister's name.  |
| 節レベル   | 文型パターン                              | NP1にN2.rareruてNP3てしまった。                        | NP1 V2.past meintoVP3.ing.   |
|        | 言語表現例                               | あのセールスマンに言いくるめられてむだな物を買ってしまった。                 | That salesman smooth-talked me into buying a useless thing..               |
|        | 文型パターン                              | N1はCL2とはV3.hitei.kako。                         | (N1 I) did V3.not CL2.past.  |
|        | 適用例                                 | 彼があれほど英語が話せるとは思わなかった。                          | I didn't know he couldspeak English so well.                               |
|        | 文型パターン                              | 「CL1.da」とV2/reruてN3はI泡を食った。                    | (N3 I) be.past thrown into confusionwhenV2.past, "CL1 "                    |
|        | 言語表現例                               | 「あす試験だ」と言われて泡を食った。                             | We were thrown into confusion when told, "There will be an exam tomorrow." |
| 文型パターン | CL1.teiru.hitei.danteiとN2はNP3.kako。 | N2VP3.past thatCL1.not.                        |  |
| 言語表現例  | 彼女はもうぼくを愛していないのだとぼくは自分に言い聞かせた。      | I convinced myself thatshedidnotlovemeanymore. |  |

## (2)位置変更可能要素指定記号

副詞 (副詞的表現を含む) など出現する位置が変わっても文型パターンとしての意味が変わらない要素について、それが出現できる位置を指定するものである。

## 4.5 文型パターンの記述例

以上の基準に従って、日英対訳文から、日本語表現の非線形な表現構造と、その意味を記述するような英語の文型パターンを対応させて取り出した例を表2に示す。表では、単語、句、節の3つのレベルで得られた文型パターンを示す。

得られた文型パターンはこの順に汎用性が高い。しかし、要素が線形要素となるか否かは構成要素の選び方に依存するため、単語レベルのパターン化ができてても句レベルのパターン化できない場合があるなど、パターン化可能な範囲は、元の表現によって様々である。

文型パターンの記述で使用する変数や関数の記述方法については、様々な方法が考えられるが、表2では人間が理解しやすいことを狙って以下のように記述した。

### (1)変数の記述

線形な自立語は、品詞属性によってN1 (名詞)、V2 (動詞) 等のように変数化し、同様、句は、NP1 (名詞句)、VP1 (動詞句)、節はCL1のように記述した。数字は、文型パターン内の変数の通番で、英語文型パターンで使用された変数との対応関係を意味する。

なお、単語変数 (自立語の変数) を持つが、句変数と節変数は持たないものが単語レベルの文型パターンであり、句変数を持つもの、節変数を持つものはそれぞれ句レベル、節レベルの文型パターンである。

### (2)関数の記述

関数は、「引数・関数名」の形式で記述した。例えば、「N2.da」では、引数は「N2」、関数名は「da」で、名詞N2の後に助動詞「た」又は「だ」が接続することを意味する。なお、複数の関数の埋め込みは、ピリオドの後に関数名を連続して書くことによって表現した。

### (3)その他記号類の記述

まず、記号「/」は離散記号で、その位置に「原文任意要素」として任意の文節が現れても良いことを意味する。

(N1は)のように括弧でくられた要素は、「パターン任意要素」である。あってもなくても良い要素を指定するのに使用されるが、内部に含まれる変数には、該当する入力文の要素がバインドされるため、省略された主語の補完処理を助けるためにも使用される。なお、補完処理で使用される場合は、同一番号の変数が、文型パターン内のどこかで使用されていないといけない。

また、(あと|後)のように、複数の表現要素を含む括弧は、要素選択記号を表し、どちらの要素が現れても良いことを意味する。

## 5.あしがき

与えられた言語表現が線形であるか非線形であるかを判断する基準を明確にすると共に、表現の線形要素を変数化することなどにより、非線形な表現構造を文型パターンによって記述する方法を提案した。

具体的には、まず、言語表現が持つ本来の意味と言語規範としての意味を定義した後、言語表現とその表す意味の関係において、重ね合わせの原理が成り立つかどうかを基準に、言語表現の構造と意味の関係に対する線形、非線形概念を定義した。

次に、工学的立場から、日本語表現の意味を英語表現で表すことを前提に、与えられた日本語表現の線形性と非線形性を判定する基準を提案した。

最後に、表現構造の線形性と非線形性の区別が、表現要素の選び方に依存することに着目して、非線形な言語表現の意味を文型パターンによって表現する方法を提案し、文型パターンの記述例を示した。

本稿では、非線形な言語表現の構造を文型パターンで表現し、その意味を英語の同様の文型パターンによって記述するための原則について検討したが、非線形な言語表現構造の数は有限を推定される。従って、抽出した非線形な構造をいかに汎化することができるか、また、どれだけの文型パターンがあれば、実用上必要な被覆率が得られるかを知ることが大変重要である。現在、複文 (埋め込みのある文)、重文 (接続のある文) を対象に語レベル、句レベル、節レベルの文型パターンをそれぞれ12万件、10万件、1万件を試作したところであり、今後、これらの文型パターンの被覆率を調べると共に文型パターン記述の改良方法についても検討する予定である。

## 謝辞

本研究は、科学技術振興事業団、戦略的基礎研究推進事業(CREST)の一環として行われているものである。御議論頂いた佐良木昌氏 (長崎純心大)、柴田勝征氏 (福岡大)、新田義彦氏 (日本大学) を初め、関係各位に感謝する。

## 参考文献

- 1) 池原悟 : 自然言語処理の基本問題への挑戦, 人工知能学会誌, Vol.16, No.3, pp.522-430 (2001)
- 2) 長尾真 : 『自然言語処理』岩波書店(1996)
- 3) 長尾真, 黒橋貞夫, 佐藤, 池原悟, 中尾洋 : 『言語の科学』第9巻『言語情報処理』, 岩波書店 (1998)
- 4) 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, pp.1269-1279 (1987)
- 5) 池原悟 : 『自然言語処理における意味解析と意味理解』, 第28回情報学基礎研究会 18-5, pp.31-40 (1992)
- 6) オールウッド, アンデソン, ダール著, 公平, 野家訳 : 『日常言語の論理学』産業図書, 1979
- 7) 三浦つとむ : 『言語と認識の理論』第1~3巻, 勁草書房(1967)
- 8) 池原悟 : 『言語で表現される概念と翻訳の原理』, 電子情報多宇宙新学会, 思考と言語研究会, (2003.12 予定)
- 9) 橋田浩一, 大津由起雄, 今西典子, Yosef Grodzinsky, 錦見美貴子 : 『岩波講座・言語の科学 10 言語の獲得と喪失』, 岩波書店, 1999
- 10) 酒井邦嘉 : 『言語の脳科学』, 中公新書, 2002
- 11) 大津由起雄, 池内正幸, 今西典子, 水光雅則 : 『言語研究入門 - 生成文法を学ぶ人のために』, 研究社, 2002