

意味論主導の日本語構文解析について

大堀 淳

2022 年 2 月 12 日

概要

本稿では、三上章の洞察をヒントに、新たな日本語の構文解析手法を提案する。

1 チョムスキー流の文法基規則による構文解析の限界

現在の構文解析の主流は、チョムスキー流の文法主導のアプローチ、すなわち、文が表現する意味構造を、その文の構成要素である句構造の導出規則により決定する、というものである。この手法は、チョムスキーの生成文法の提案以来、西洋語はもちろん、日本語の構文解析の研究においても幅広く採用されている、いはば構文解析理論の暗黙の枠組みとなっていると言える。

たしかに、三上が天秤型言語と呼ぶコーブラ（繫辞）を持つ西洋言語の場合、文の意味構造は、主語・述語構造を核とする種々の規則集合で近似できる。しかし、コーブラを持たず、さらに主語・述語構造をそもそも持っていない日本語にとって、このアプローチの有効性は限られる。例として、以下の文を考えてみよう。

- 「象は、鼻が長い」
- 「象は、皆が好きだ」
- 「象は、早く走る」
- 「象は、話しかけても（答えない）」
- 「象は、荷物を運ぶのに（役立つ）」
- 「象は、一緒に暮らす」

日本語を理解する者は、これらの意味を難なく理解できる。これらの意味に従い、各文の「象は」が述語に対して果たす構文上の機能を分析すると、以下のようになる。

例文	述語	「象は」の機能	作用対象
「象は、鼻が長い」	長い	の格，連体修飾	主格の名詞
「象は、皆が好きだ」	好き	を格，対格	述語
「象は、早く走る」	走る	が格，主格	述語
「象は、話しかけても・・・」	話す	に格，与格	述語
「象は、荷物を運ぶのに・・・」	運ぶ	で格，具格	述語
「象は、一緒に暮らす」	暮らす	と格，共格	述語

しかし、これらの各文が表す意味構造を正しく分析する文法規則の集合を定義するのは、困難とおもわれる。日本語は、これらの意味構造を規定する文法構造を持たないからである。日本語の理解は、助詞の表す意味を頼りに、それぞれの構文単位（文節）の役割を推定し、文全体の意味構造を合成することで成り立っていると思われる。「象は」は、その文脈に応じて、連体修飾（の格）、主格、対格、与格、具格、共格の何

れかの役割を果たしうるが、構文上それを規定する十部な規則はない。日本語の話者や読者は、「象は」が出現する文の意味構造からその機能を選択し、文全体の意味構造を合成しえいる、と推定される。

本稿では、日本語の話者・読者が行っているであろう意味に従った構文の分析と構造の合成を、意味主導の構文解析と呼び、その計算論的なアプローチを提案する。

2 三上章による日本語の分析

本稿で提唱する意味主導の構文解析は、日本語の文法構造に関する三上章の洞察と分析 [1] を基礎とする。三上文法の概要の説明は他 [2] に譲り、ここでは、本稿が提唱する意味主導の構文解析に関連すると思われる要素を略述する。

三上の日本語文法の洞察の柱は、以下の3点と理解できる。

1. 日本語の文の構造は「述語一本建て」である。特に、日本語の文には、構文要素としての主語は存在しない。文は、述語（動詞、形容詞、名詞＋だ）を中心に、その述語に関わる格を表す種々の連用修飾語が、並置されている構造を持つ。
2. 述語に対して並置される連用修飾語の語順は自由である。例えば、「講義する」はその意味から、「～が」（主格）、「～に」与格、「～を」対格の3つをとるが、以下から分かる通り、それら3つの連用修飾部分の語順は自由である。
 - 私は、学生に三上文法を講義する。
 - 私は、三上文法を学生に講義する。
 - 学生には、私が三上文法を講義する。
 - 学生には、三上文法を私が講義する。
 - 三上文法は、私が学生に講義する。
 - 三上文法は、学生に私が講義する。

いずれも、意味の通じる正しい日本語である。これらの例は、堤題の「は」を含むが、「私が学生に三上文法を講義する。」のように無題であっても同様である。

主格を構成する連用修飾語は、文頭に来る傾向をもつが、これは「堤題」が文頭にくる傾向にあり、かつ、堤題の対象は主格である場合が多いことによる。

3. 堤題を表す助詞「は」を有する。「は」は、堤題、即ち、テーマ・トピックを提示・変更する役割を果たす助詞である。この堤題の助詞「は」の存在が、日本語の際立った特徴と言える。

なお、「象は」などは、主格を表す「主語」と誤解されるが、西洋語のような主語の構造と機能を有する文法要素は、日本語には存在しない。「象は」の機能は

- 「象」を堤題する
- 述部の欠けている格を兼ねる

の2つである。上記の例から明らかなように、例えば「象は早く走る」のように述部「早く走る」が主格を欠いていれば主格を兼ね、「象は、皆が好きだ」のように主格は充足しているが対格を欠いていれば、対格を兼ねる。さらに、以下の習慣がある。

- 「～は」が主格（が格）か対格（を格）を兼ねる場合は、格助詞「が」と「を」は省略される。
- 「～は」がその他の格（に格、へ格、と格、で格など）を兼ねる場合は、格助詞を伴うのが普通であるが、省略される場合もある。例えば、
 - － 東京には、新幹線が早く着く。
 - － 東京は、新幹線が早く着く。

はどちらも意味の通る日本語の文と認められる。

3 日本語の意味主導解析戦略

以上の分析から明らかな通り，日本語の解析には，「は格」の文における機能の同定と，その役割に基づく意味構造の合成が必要である．本稿が提案する解析手法では，以下の戦略で，この両者の系統的な達成を目指す．

1. 三上文法を参考に，名詞文，動詞文，形容詞文等の種々の文を表す可能な意味構造集合を決定する．この意味構造は，文字通り，文の意味の構造を表し，語順等の構文論的な構造は含まない．例えば，「象は鼻が長い」と「象は速く走る」の意味構造は，以下のように与えられる．

```
{
  種類 = 形容詞文，
  述語 = {形容詞 = "長い"，主格 = {名詞 = "鼻"，の格 = "象"}},
  堤題 = {名詞 = "鼻"}
}
{
  種類 = 動詞文，
  述語 = {動詞 = "走る"，主格 = {名詞 = "象"}，連用修飾 = "速く"}，
  堤題 = {名詞 = "鼻"}
}
```

2. 各文節について，その文節が担いうる文の中の意味を表す文の部分構造を列挙する．例えば，「象は」は，「象は鼻が長い」では

```
{
  種類 = 形容詞文，
  述語 = {主格 = {の格 = "象"}},
  堤題 = {名詞 = "象"}
}
```

を表し，「象は速く走る」では

```
{
  種類 = 動詞文，
  述語 = {主格 = {名詞 = "象"}},
  堤題 = {名詞 = "象"}
}
```

を表す．列挙され文の可能な部分構造の集合を，文節が表現する可能な意味構造と見做す．

3. 文節の列である文を，各文節が表現する可能な意味構造の集合の列で表す．「象は」は，上記の2つの部分構造を含む，以下のような集合で表現できる．

```
[
  {
    種類 = 形容詞文，
    述語 = {主格 = {の格 = "象"}},
```

```

堤題 = {名詞 = "象"}
},
{
種類 = 動詞文,
述語 = {主格 = {名詞 = "象"}},
堤題 = {名詞 = "象"}
},
...
]

```

この表記例では, $[d_1, d_2, \dots]$ で集合を表現している.

4. 可能な意味構造集合の列を順に意味的に結合し, 文全体を表す可能な意味構造集合を得る.

今, 部分構造 r と s が整合性ある全体の部分を表す場合, その合成を $r \sqcup s$ と書くことにする. この演算は, r と s の共通の属性が一致していることをテストし, 一致している場合は, それ以外の属性の和を取る操作である. 共通の属性が一致していなければ, $r \sqcup s$ は未定義である. $r \sqsubseteq r'$ で, r' が r より情報量の多い部分構造を表すとする, $r \sqcup s$ は, $r \sqsubseteq r \sqcup s$ かつ $s \sqsubseteq r \sqcup s$ かつなる最小の要素である. 例えば, r が

```

{
種類 = 動詞文,
述語 = {主格 = {名詞 = "象"}},
堤題 = {名詞 = "象"}
}

```

を表し, s が

```

{
種類 = 動詞文,
述語 = {動詞 = "走る"}
}

```

を表せば, それらの結合 $r \sqcup s$ は以下ようになる.

```

{
種類 = 動詞文,
述語 = {動詞 = "走る", 主格 = {名詞 = "象"}},
堤題 = {名詞 = "象"}
}

```

本稿の提案の主要なアイデアは,

この意味合成演算 \sqcup を, 可能な部分構造を表す集合演算に拡張し, 整合性ある意味構造の候補集合を計算する

というものである. 今, 集合 R を分節 1 が表現する可能な文の部分構造集合, 集合 S を分節 2 が表現する可能な文の部分構造集合とる. すると, 「文節 1 文節 2」の並びが表す可能な部分構造は, R と S のそれぞれに含まれる要素の可能な結合をすべて集めたものと見なせる. この集合を $R \sqcup S$ と表現することにする. R と S がそれぞれ互いに独立な (即ち $r \sqsubseteq r'$ なる要素を含まない) 要素の集

合であれば、この演算の結果は以下のような集合と理解できる。

$$R \sqcup S = \{d \mid d = r \sqcup s, r \in R, s \in S\}$$

この結合結果の集合に含まれる要素はそれぞれ、「文節 1」と「文節 2」の情報を共に含む可能な部分構造をあらわす。文章部分 S が表す可能な部分構造を $\llbracket S \rrbracket$ と書くことにすると、「文節 1 文節 2」が表す可能な部分構造は、

$$\llbracket \text{文節 1 文節 2} \rrbracket = \llbracket \text{文節 1} \rrbracket \sqcup \llbracket \text{文節 2} \rrbracket$$

と定式化できる。一般に、文節 B_1, B_2, \dots, B_n からなる文の可能な意味構造は、

$$\llbracket B_1, B_2, \dots, B_n \rrbracket = \llbracket B_1 \rrbracket \sqcup \llbracket B_2 \rrbracket \sqcup \dots \sqcup \llbracket B_n \rrbracket$$

と表せる。この演算は、結合律が成り立つため、計算の順序は自由である。

例えば、「象は」の可能な構造が 2 要素からなる集合、

```
[
  {
    種類 = 形容詞文,
    述語 = {主格 = {の格 = "象"}},
    堤題 = {名詞 = "象"}
  },
  {
    種類 = 動詞文,
    述語 = {主格 = {名詞 = "象"}},
    堤題 = {名詞 = "象"}
  },
]
```

であり、「走る」の可能な部分構造が 1 要素のみからなる集合

```
[
  {
    種類 = 動詞文,
    述語 = {動詞 = "走る"}
  },
]
```

であれば、「象は走る」が表す可能な文部分構造は、上記の 2 つの集合を結合して得られる

```
[
  {
    種類 = 動詞文,
    述語 = {主格 = {名詞 = "象"}},
    述語 = {動詞 = "走る"}
  },
]
```

である。

このように、集合に拡張された情報統合演算は、整合性のある可能な構造を選びだし、それぞれの情報を合成する力をもっている。

上記の最小上界演算 \sqcup は、[3] で示された通り、情報の量で順序付けられた集合の Smyth Power Domain の拡張であり、一般化された関係データベースの自然結合演算 [4] と一致する。さらに、この演算は `_join` 演算として、SML# に実装されている。そこで、以下の具体的なプログラム開発戦略が可能である。

1. 形態素解析の結果を、SML# のラベル付きレコードで表現。
2. 形態素解析の結果得られる字句列を、各文節（単語 + 助詞）列に統合。
3. 各種の文（動詞文、名詞文・形容詞文、名詞文・準詞文）毎に、文を表すレコード構造を列挙。
4. 各文節に対して、それが果たす文の中の役割を表現する不完全な文表現レコードのリストを構成。
SML# は型付き言語であるため、型の異なるレコードのリストは構成できない。そこで、可能なレコードを、SML# の機能である Dynamic 型に埋め込み、Dynamic 型のリストで表現する。
5. 文節の列からなる文を、前項で得られる Dynamic 型のリストのリストで表現。
6. SML# の自然結合演算を用いて、文を構成する文節が表現する部分構造リストを順に自然結合し、文全体の可能な構造を表すリストを得る。

4 文脈依存の意味主導解析

以上は、意味文脈に依存しない解析である。本稿で提案する意味主導解析のもう一つの優れている点は、以上の手法が、系統的かつ自然に、文脈依存解析に拡張できる点である。ここでの文脈依存の文脈は、構文論的なものではなく、その文が解釈される意味の文脈である。日本語の「は格」の機能である堤題は、もともと意味的なものであり、構文構造に拘束されない。とくに、文の終わりを超えて累計的に管理・維持され、一部は一時的に変更される、等の構造を持つと考えられる。意味文脈は、この堤題で設定された意味空間（象に関すること、鼻に関すること、等々）を表現するデータ構造で表現できると考えられる。このデータ構造の表現には種々の可能性があるが、その一つは、文の可能な部分構造を表す雛形の集合とするアプローチがありうる。すると、文脈依存の意味主導構文解析は、この雛形を自然結合のベースとして用いることにより、種々の省略がある場合も、意味論的に妥当な構文解釈が系統的に可能と考えられる。

参考文献

- [1] 三上章現代語法序説-シンタクスの試み- くろしお出版、1972。
- [2] 大堀 淳、三上文法の概要、草稿。
- [3] BUNEMAN, P., JUNG, A., AND OHORI, A. 1991. Using powerdomains to generalize relational databases. *Theoretical Computer Science* 91, 1, 23–56.
- [4] OHORI, A., BUNEMAN, P., AND BREAZU-TANNEN, V. 1989. Database programming in Machiavelli – a polymorphic language with static type inference. In *Proc. ACM SIGMOD conference*, Portland, Oregon, pp. 46–57.