2D5-OS-03b-1

部分型理論による概念表記の展望

Perspective of Aspect Representation with Subtyping Theory

中野悠紀　　戸次大介
Yuki Nakano　　Daisuke Bekki

お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 情報科学コース
Ochanomizu University, Graduate School of Humanities and Sciences, Faculty of Science, Department of Information Science

Synonymity is one of the problems of paraphrase in natural language processing. For example, “The ball is red” equals to “The color of the ball is red”, which is difficult to express in formal semantics. In this paper, we propose a logical treatment of abstract objects by means of aspect functions, such as *color*.

1. はじめに

自然言語処理における「言い換え」の問題の一つとして、「このボールが赤い」と「このボールの色が赤い」という文は同値であるが、形式意味論ではこの二文に対応する論理式の同値性が成り立たない、という問題がある。この問題に対して麻生ら [4] が従来の形式意味論とは異なる意味表現形式を用いた解決法を提案しているが、本研究は Bekki & Asher [2] が提案した Aspect 関数を用い、形式意味論の枠組みの中で解決する方法を提案する。

二章では言い換えの問題について説明し、麻生らの解決法について紹介する。三章では [2] に従い Aspect 関数について説明し、四章では Aspect 関数を用いた言い換えを提案する。

2. 麻生ら (2008) の意味表現形式

自然言語では「このボールが赤い」という文を「このボールの色が赤い」と言い換えることができる。この二文はほぼ同じ意味であり、こうした表現は多様な形で表現することができる。同じ意味の文が多様な形で表現できること、すなわち多様な表層表現が同一の意味を持つことを「同義性」という。

この二文では後者にのみ「色」という語が現れるので、前者において「色」の意味を補う必要があるが、形式意味論ではそのような仕組みがないため、この二文の同義性は成り立たない。この問題に対して麻生らは以下のように解決法を提案している。

2.1 麻生らの言い換え

[4] では、語の統語的特性と意味的特性の両方に基づいて、構成的に説明可能な言い換えを可能にした。麻生らは以下の言い換え処理の手順を取っている。

1. 表層文を語彙分割し、依存構造を抽出するとともに、それを反映した意味表現を生成する。
2. 意味表現に対して構造変換規則を適用する。
3. 変換後の意味表現を語彙分割し、依存構造木を得るとともに、表層文を生成する。

麻生らは意味表現に関して、[5] において提案された意味表現形式を採用している。この意味表現形式は、依存構造をベースとし、同じ意味を持つと近似できる文の意味に対して、ほぼ同一の意味表現で表すことができる形式である。

2.2 麻生らの言い換えの例

[4] に従い、「赤いボール」という名詞句を「赤い色のボール」というほぼ同一な意味の名詞句に変換する例を紹介する。

麻生らは意味表現に使われる記号とその接続規則について、次のように定義している。

- 実体や属性を表す名詞的な概念構成素を○で表す
- 現象を表す動詞的な概念構成素を◎で表す
- 英語の関係代名詞相当の構成素を●で表す
- 任意の名詞句に対応づけられ、その名詞句を節構造の外に出す働きを持つ概念構成素を※で表す
- 終止形形容詞の意味表現を、関係節内部の動詞句を取り出して動詞化する働きを持つ特殊な概念構成素（“be” や「ある」に対応する）を $\diamond$ で表す
- ある●が○を指示することを $=$ で示す
- 格概念を $\rightarrow$（矢印）で表す
- 不定な概念構成素名を示すのに変数 X, Y などを用いる

これらの記号は、次のような制約を満たすように接続される。

- ○, ●, 変数は単一の $\rightarrow$ の始点に接続することができる
- ◎は複数の $\rightarrow$ の終点に接続することができる（ただし、$\rightarrow$ の種類によっては一つしか接続できない）。

この定義に従い [4] では「赤い」と「ボール」の意味は図 1 のように定義されている。「赤い」の意味は「赤に等しい色相を内包する色を内包するところの」と表現することができ、「ボール」の意味は「X に等しい色相を内包する色を内包するところのものであり、球に等しい形を内包するところのものであり、Y がそれを投げるところのものであり、$\cdots$」と表すことができる。「赤い」が「ボール」を修飾した「赤いボール」という文は、二つの語彙の意味表現の共通部分を接続して図 2 のように表現される。

連絡先: 中野悠紀，お茶の水女子大学大学院人間創成科学研究科理学専攻情報科学コース戸次研究室，東京都文京区大塚 2-1-1，nakano.yuki@is.ocha.ac.jp

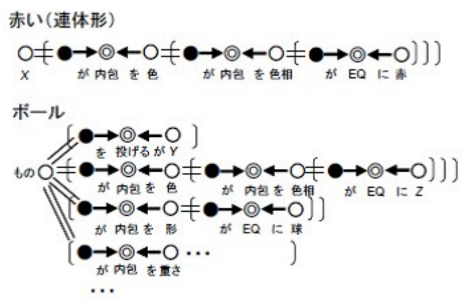

図 1:「赤い（連体形）」と「ボール」の意味表現

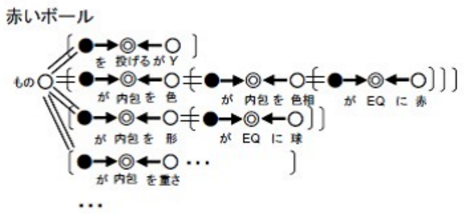

図 2:「赤いボール」の意味表現

また日本語の助詞「の」は [4] では図 3 のように表現されている。

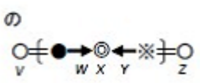

図 3:「の」の意味表現

麻生らは意味構造変換規則を次のように定義している。

- 関係節構造と形容詞終止形の間の変換
  ··· ○ = (● → ◎ ···) ⇔ ··· ○ → ◇ = (● → ◎ ···)
- 関係節構造と叙述形の間の変換
  ··· ○ = (● → ◎ ···) ⇔ ··· ○ → ◎ ···
- 関係節構造と「の」を使った表現の間の変換
  ··· = (● → ◎ ← ○ ···) ⇔ ··· = (● → ◎ ← ※) = ○ ···

この手順に従い、[4] では、図 4 のように「赤いボール」という入力文から「赤い色のボール」を生成している。

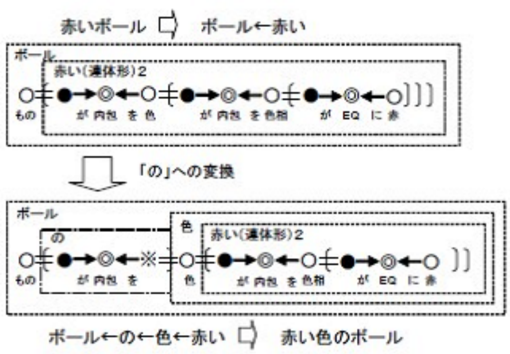

図 4:「赤いボール」から「赤い色のボール」への言い換え

このように [4] では名詞、形容詞の内部構造を柔軟かつ統一的に記述している。一方でグラフ表現による意味表示からでは、推論をどのように行うか、あるいは論理的な（スコープを要するような）概念をどのように表現するかが分からない。そこで本研究では Bekki & Asher [2] によって提案された Aspect 関数により、「色」のような対象アスペクトを論理的に表現する。

3. Bekki & Asher (2012) の型理論

3.1 論理的多義性

[2] では Aspect 関数を用いて自然言語の多義性（Polysemy）と Copredication の説明をしている。

辞書の曖昧性の原因の一つに自然言語の多義性が挙げられる。例として、次のような普通名詞 “book” について考えたい。普通名詞 “book” は少なくとも二つの異なる概念を表しうる。たとえば (1)“John memorized a book.” (2)“John burned a book.” という二つの文を考えてみると、(1) の “book” は *information* の意味であるため、その型を Info と表す。一方 (2) の “book” は *physical object* の意味であるため、その型を PhyObj と表す。しかし次の例のように、異なる型を要求する二つの述語を等位接続の形で用いることができる。

(3)John memorized and burned a book.

このような構文は Copredication と呼ばれている。

[2] では、この問題を部分型理論と Aspect 関数を用いることによって解決している。

3.2 部分型理論と Aspect 関数

S が T の部分型（Subtype, $S \leq T$）だとすると、型 S の任意の項は型 T の項が必要とされる文脈で扱える、つまり $S \leq T$ なら S の全ての要素はまた T の要素となる。

[2] では、この規則を単射 i を使って次のように表している。

$$\frac{t:S \quad i:S\to T}{i(t):T}\,(\to E)$$

この単射 i は任意の部分型 S の要素をその super type T の要素に変換する関数である。

この規則を用いて、(4)“John laughed.” という文について考えてみよう。

$$\dfrac{\dfrac{\dfrac{\text{John}}{j:\mathsf{Man}} \quad \dfrac{}{id:\mathsf{Man}\to\mathsf{Animate}}}{j:\mathsf{Animate}}\,(\to E) \quad \dfrac{\text{laughed}}{\lambda x.laugh(x):\mathsf{Animate}\to\mathsf{Prop}}}{laugh(j):\mathsf{Prop}}\,(\to E)$$

証明木 1:(4) の導出

まず “laughed” は有生な、つまり Animate object を主語に取るので、“laughed” の型は Animate → Prop となる。ここで Animate は Entity の部分型である。また j : Man は “John is a man” を意味し、Man は Animate の部分型である。よって、証明木 1 のように $laugh(j)$ が (4) の意味表示となる。

次に LGB の意味表示を l とし、l : Book とする。そして次のような Aspect 関数 (*aspect function*) を定義する。

$$\begin{array}{ll} asp_I: & \mathsf{Book}\to\mathsf{Info} \\ asp_P: & \mathsf{Book}\to\mathsf{PhyObj} \end{array}$$

book を information aspect に移す関数 asp_I, physical aspect にする関数 asp_P とする。ここで同時に Info と PhyObj の部分型である要素は存在しない。よって Aspect 関数は部分型関係ではないといえる。

Aspect 関数を用いることで (1)“John memorized a book.” と (2)“John burned a book.” の意味表示はそれぞれ $memorize(j, asp_I(l)), burn(j, asp_P(l))$ となる（証明木 2,3）。

このように Aspect 関数とは、オントロジー上のノードからノードへの関数である（図 5）。

$$\dfrac{\dfrac{\text{John}}{j:\mathsf{Animate}} \quad \dfrac{\dfrac{\text{memorized}}{\lambda y.\lambda x.memorize(x,y):\mathsf{Info}\to\mathsf{Animate}\to\mathsf{Prop}} \quad \dfrac{\dfrac{LGB}{l:\mathsf{Book}} \quad \overline{asp_I:\mathsf{Book}\to\mathsf{Info}}}{asp_I(l):\mathsf{Info}}(\to E)}{\lambda x.memorize(x,asp_I(l)):\mathsf{Animate}\to\mathsf{Prop}}(\to E)}{memorize(j,asp_I(l)):\mathsf{Prop}}(\to E)$$

証明木 2：(1) の導出

$$\dfrac{\dfrac{\text{John}}{j:\mathsf{Animate}} \quad \dfrac{\dfrac{\text{burned}}{\lambda y.\lambda x.burn(x,y):\mathsf{PhyObj}\to\mathsf{Animate}\to\mathsf{Prop}} \quad \dfrac{\dfrac{LGB}{l:\mathsf{Book}} \quad \overline{asp_P:\mathsf{Book}\to\mathsf{PhyObj}}}{asp_P(l):\mathsf{PhyObj}}(\to E)}{\lambda x.burn(x,asp_P(l)):\mathsf{Animate}\to\mathsf{Prop}}(\to E)}{burn(j,asp_P(l)):\mathsf{Prop}}(\to E)$$

証明木 3：(2) の導出

$$\dfrac{\dfrac{\dfrac{\text{このボール}}{T/(T\backslash NP_{nc}):\lambda P.Pb:(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}} \quad \dfrac{\text{が}}{(T\backslash NP_{nc})/(T\backslash NP_{ga}):id:\tau\to\tau}}{T/(T\backslash NP_{ga}):\lambda P.Pb:(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}}(>B) \quad \dfrac{\overline{color\_of:\mathsf{PhyObj}\to\mathsf{Color}} \quad \dfrac{\text{赤い}}{S\backslash NP_{ga}:\lambda x.Red(x):\mathsf{Color}\to\mathsf{Prop}}}{S\backslash NP_{ga}:\lambda x.Red(color\_of(x)):\mathsf{PhyObj}\to\mathsf{Prop}}(>B)}{S:Red(color\_of(b)):\mathsf{Prop}}(\to E)$$

証明木 4：(5) の導出

$$\dfrac{\dfrac{\dfrac{\dfrac{\text{このボール}}{\begin{array}{c}T/(T\backslash NP_{nc}):\lambda P.Pb\\ :(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}\end{array}} \quad \dfrac{\dfrac{\text{の}}{\begin{array}{c}(T\backslash NP_{nc})/(T\backslash NP_{nc})/(NP\backslash NP):\lambda f.\lambda P.\lambda x.P(f(x)):\\ (\mathsf{PhyObj}\to\mathsf{Color})\to(\mathsf{Color}\to\mathsf{Prop})\to(\mathsf{PhyObj}\to\mathsf{Prop})\end{array}} \quad \dfrac{\text{色}}{\begin{array}{c}NP\backslash NP:\lambda x.color\_of(x)\\ :\mathsf{PhyObj}\to\mathsf{Color}\end{array}}}{\begin{array}{c}(T\backslash NP_{nc})/(T\backslash NP_{nc}):\lambda P.\lambda x.P(color\_of(x))\\ :(\mathsf{Color}\to\mathsf{Prop})\to(\mathsf{PhyObj}\to\mathsf{Prop})\end{array}}(>)}{\begin{array}{c}T/(T\backslash NP_{nc}):\lambda P.b(color\_of(x))\\ :(\mathsf{Color}\to\mathsf{Prop})\to\mathsf{Prop}\end{array}}(>B) \quad \dfrac{\text{が}}{\begin{array}{c}(T\backslash NP_{nc})/(T\backslash NP_{ga}):id\\ :\tau\to\tau\end{array}}}{\begin{array}{c}T/(T\backslash NP_{ga}):\lambda P.P(color\_of(b))\\ :(\mathsf{Color}\to\mathsf{Prop})\to\mathsf{Prop}\end{array}}(>B) \quad \dfrac{\text{赤い}}{\begin{array}{c}S\backslash NP_{ga}:\lambda x.Red(x)\\ :\mathsf{Color}\to\mathsf{Prop}\end{array}}}{S:Red(color\_of(b)):\mathsf{Prop}}(\to E)$$

証明木 5：(6) の導出

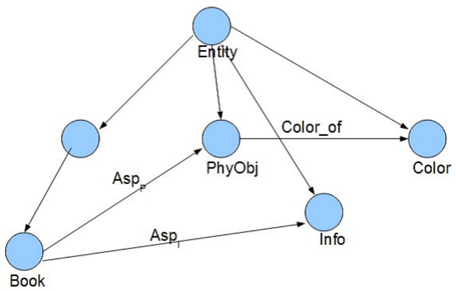

図 5：Aspect 関数

4. 提案

4.1 Aspect 関数を用いた言い換え

ボールなどの Physical Object を、その色相に写す Aspect 関数を次のように定義する。

$$color\_of:\mathsf{PhyObj}\to\mathsf{Color}$$

これを用いて (5)「このボールが赤い」という文が、(6)「このボールの色が赤い」という文と言い換え可能、つまり同じ意味表示を持つことを示す。

「このボール」は本来コ系指示表現であるが、ここでは議論を単純化するため、固有名詞に準じて次のように表す。

$$T/(T\backslash NP_{nc}):\lambda P.Pb:(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}$$

また「赤い」という述語は次のように表す。

$$S\backslash NP_{ga}:\lambda x.Red(x):\mathsf{Color}\to\mathsf{Prop}$$

これらを使い、(5) は証明木 4 から導出できる。同様に (6) は証明木 5 から導出できる。(5),(6) の意味表示は共に $Red(color\_of(b))$ である。したがって、この二つの文は同じ意味表示である。

同様に、(7)「この象は鼻が長い」と (8)「この象の鼻が長い」という二つの文が同じ意味表示になることを示す。

- 「鼻」は次のように表す。
  $NP_{nc}\backslash NP_{no}:\lambda x.nose\_of(x):\mathsf{PhyObj}\to\mathsf{PhyObj}$
- 「この象は」は次のように表す。
  $T/(T\backslash NP_{ga}):\lambda P.Pz:(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}$
- 「この象」は次のように表す。
  $T/(T\backslash NP_{nc}):\lambda P.Pz:(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}$
- 「長い」という述語は次のように表す。
  $S\backslash NP_{ga}:\lambda x.Long(x):\mathsf{PhyObj}\to\mathsf{Prop}$

[6] では「鼻」は存在量化されていたが、「鼻」を「象」の Aspect であると捉えることも可能である。その場合は (7) と (8) の二文の導出は証明木 6,7 に示すように、(5),(6) と並行する構造をもつことになる。

4.2 考察

本分析によると以下の言い換えは全て同義であることが導かれる。しかし、実際には以下に示したようにその容認可能性にはばらつきが見られる。

- 「このボールは赤い」と「このボールの色は赤い」は同じ意味表示である。
- 「この象は長い」と「この象は鼻が長い」は同じ意味表示ではない。
- 「太郎は重い」と「太郎の体重は重い」は同じ意味表示である。

$$\dfrac{\dfrac{\text{この象は}}{\begin{array}{c}T/(T\backslash NP_{ga}):\lambda P.Pz\\ :(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}\end{array}} \quad \dfrac{\dfrac{\dfrac{\dfrac{\phi_{zoo}}{\begin{array}{c}(T\backslash NP_{ga})/(T\backslash NP_{nc})/(NP_{nc}\backslash NP_{no}):\lambda f.\lambda P.\lambda x.\lambda\overrightarrow{x}.P(f(x))\overrightarrow{x}\\ :(\mathsf{PhyObj}\to\mathsf{PhyObj})\to(\mathsf{PhyObj}\to\mathsf{Prop})\to(\mathsf{PhyObj}\to\mathsf{Prop})\end{array}} \quad \dfrac{\text{鼻}}{\begin{array}{c}NP_{nc}\backslash NP_{no}:\lambda x.nose\_of(x)\\ :\mathsf{PhyObj}\to\mathsf{PhyObj}\end{array}}}{\begin{array}{c}(T\backslash NP_{ga})/(T\backslash NP_{nc}):\lambda P.\lambda x.P(nose\_of(x))\\ :(\mathsf{PhyObj}\to\mathsf{Prop})\to(\mathsf{PhyObj}\to\mathsf{Prop})\end{array}}(>) \quad \dfrac{\text{が}}{(T\backslash NP_{nc})/(T\backslash NP_{ga}):id:\tau\to\tau}}{\begin{array}{c}(T\backslash NP_{ga})/(T\backslash NP_{ga}):\lambda P.\lambda x.P(nose\_of(x))\\ :(\mathsf{PhyObj}\to\mathsf{Prop})\to(\mathsf{PhyObj}\to\mathsf{Prop})\end{array}}(>B) \quad \dfrac{\text{長い}}{\begin{array}{c}S\backslash NP_{ga}:\lambda x.Long(x)\\ :\mathsf{PhyObj}\to\mathsf{Prop}\end{array}}}{\begin{array}{c}S\backslash NP_{ga}:\lambda x.Long(nose\_of(x))\\ :\mathsf{PhyObj}\to\mathsf{Prop}\end{array}}(>)}{S:Long(nose\_of(z)):\mathsf{Prop}}(>)$$

証明木 6：(7) の導出

$$\dfrac{\dfrac{\dfrac{\dfrac{\text{この象}}{\begin{array}{c}T/(T\backslash NP_{nc}):\lambda P.Pz\\ :(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}\end{array}} \quad \dfrac{\dfrac{\text{の}}{\begin{array}{c}(T\backslash NP_{nc})/(T\backslash NP_{nc})/(NP_{nc}\backslash NP_{no}):\lambda f.\lambda P.\lambda x.\lambda.P(f(x))\\ :(\mathsf{PhyObj}\to\mathsf{PhyObj})\to(\mathsf{PhyObj}\to\mathsf{Prop})\to(\mathsf{PhyObj}\to\mathsf{Prop})\end{array}} \quad \dfrac{\text{鼻}}{\begin{array}{c}NP_{nc}\backslash NP_{no}:\lambda x.nose\_of(x)\\ :\mathsf{PhyObj}\to\mathsf{PhyObj}\end{array}}}{\begin{array}{c}(T\backslash NP_{nc})/(T\backslash NP_{nc}):\lambda P.\lambda x.P(nose\_of(x))\\ :(\mathsf{PhyObj}\to\mathsf{Prop})\to(\mathsf{PhyObj}\to\mathsf{Prop})\end{array}}(>)}{\begin{array}{c}T/(T\backslash NP_{nc})\\ :\lambda P.P(nose\_of(z))\\ :(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}\end{array}}(>B) \quad \dfrac{\text{が}}{(T\backslash NP_{nc})/(T\backslash NP_{ga}):id:\tau\to\tau}}{\begin{array}{c}T/(T\backslash NP_{ga}):\lambda P.P(nose\_of(z))\\ :(\mathsf{PhyObj}\to\mathsf{Prop})\to\mathsf{Prop}\end{array}}(>B) \quad \dfrac{\text{長い}}{\begin{array}{c}S\backslash NP_{ga}:\lambda x.Long(x)\\ :\mathsf{PhyObj}\to\mathsf{Prop}\end{array}}}{S:Long(nose\_of(z)):\mathsf{Prop}}(>)$$

証明木 7：(8) の導出

- 「太郎は高い」と「太郎の身長は高い」は同じ意味表示ではない。
- 「このビルは高い」と「このビルは高さが高い」は同じ意味表示である。

「このボールは赤い」と「このボールの色は赤い」は同じ意味である。しかし、「この象は長い」という文において、「この象は鼻が長い」という読みの優先度は「この象自体が長い」という読みに比べて遥かに低いため、「この象は長い」と「この象は鼻が長い」は同じ意味には解釈しにくい。また「太郎は重い」という文を、多くの人は「太郎は体重が重い」と解釈するだろう。よって「太郎は重い」と「太郎の体重は重い」の意味表示は同じである。一方「太郎は高い」という文を、「太郎の身長が高い」と解釈することは困難である。太郎の何が「高い」のか一意に定まらないため、文の意味が一意に定まらないからである（「太郎の意識が高い」や「太郎（という奴隷）の値段が高い」など様々な解釈が可能である）。したがって「太郎は高い」と「太郎の身長は高い」の意味は同じではない。しかし、一般的に「このビルは高い」という文は「このビルは高さが高い」として解釈されるので、この二文の言い換えは可能である。

これらの容認可能性の差に対する説明は現時点では明らかではなく、今後の研究課題である。

5. まとめと今後の課題

[2] の Aspect 関数を用いることで、形式意味論では扱えないとされていた「このボールは赤い」と「このボールの色は赤い」のような同義性を、形式意味論の枠組みのなかで解決する分析を提示した。

今後の課題として、容認可能性のばらつきに対して、それらが何故生じるのかを検証し、この容認可能性の差に対する説明を示したい。

参考文献

[1] Asher, N. 2011, *Lexical Meaning in Context — a web of words —*, Cambridge University Press.

[2] Bekki, D. and Asher, N. 2012, *Subtyping in Logical Polysemy and Copredication*, in Proceedings of the Ninth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS9), pp.98-105, International Symposia on AI 2012, Miyazaki, Japan.

[3] Pierce, B.C. 2002, *Types and Programming Languages*, The MIT Press.

[4] 麻生英樹, 伊藤幸宏, 高木朗, 2008, 「言い換えに適した意味表現について」, jsai2008.

[5] 高木朗, 伊東幸宏, 1987, 「 自然言語の処理」, 丸善.

[6] 戸次大介, 2010, 「日本語文法の形式理論—活用体系・統語構造・意味合成—」, 日本語研究叢書 24, くろしお出版.