範疇文法と部分方向性組み合わせ論理の Curry-Howard 対応に基づく統語解析

理学専攻・情報科学コース　尾崎 博子

1　はじめに

自然言語処理・理論言語学の分野における形式的な文法理論の一つに組み合わせ範疇文法 (CCG: [1]) がある。CCG は頑健な構文解析に使われている一方で、言語学的にも適切な制約を与えることが知られている。

[2] では CCG と論理体系の関係を明らかにするために、部分方向性組み合わせ論理（Subdirectional Combinatory Logic (以下 SDCL)）という形式体系を提案した。SDCL は、組み合わせ論理 (CL) の拡張であり、ランベック計算のように関数適用の二つの方向性を区別する体系である。[2] では [1] のコンビネータを用いて、CCG を SDCL の一体系として位置付けている。

本研究では、まず始めに依存関係が成立するコンビネータの組み合わせを明らかにしたうえで、型システムの正当性を示すために必要とされている 1) Subject-reduction、2) 合流性、3) 停止性、の計算論的性質を証明する。

次に、SDCL と自然言語の文法の対応について議論する。本研究では英語の Extraction（抽出性）に着目する。抽出性には CCG における型繰り上げ規則の有無が関係している。この規則と対応する演繹定理が SDCL でどのような場合に存在し得るか証明することで、文法について論理の観点から議論することが可能となる。

また、実装の観点においては、CCG を用いた構文解析では簡単な文に対してさえ複数の構文木が導出されることがある。しかし処理効率の観点からは 1 つの意味に対しては導出を一意に決定したい。この上で正規化解析 [5] は構文解析のための効率的で完全性のある手法である。本研究では日本語のかき混ぜ文 (scrambling) と英語について正規化解析の手法を適用する方法を提案し、構文解析器の実装を行う。

2　部分方向性組み合わせ論理

CL の → に方向性を持たせ、/ と \ に分けると、CL における各コンビネータから SDCL における対応するコンビネータが複数生じる。例えば、CL におけるコンビネータ $\mathbf{K} : A \to (B \to A)$ には以下の 4 つが対応する。

$$\mathbf{K}_{//} : (A/B)/A \quad \mathbf{K}_{\backslash/} : (A\backslash B)/A$$
$$\mathbf{K}_{/\backslash} : (A/B)\backslash A \quad \mathbf{K}_{\backslash\backslash} : (A\backslash B)\backslash A$$

(他のコンビネータについても同様)

このように SDCL は CL に方向性を持たせた論理である。

3　構文、型規則、略記法

<u>構文</u>　型 $\tau ::= \quad \gamma \mid \tau/\tau \mid \tau\backslash\tau$　　(γ は型の要素)

　　　式 $\Lambda ::= \quad x \mid c \mid \Lambda^{\triangleright}\Lambda \mid \Lambda^{\triangleleft}\Lambda$　(c はコンビネータ)

<u>型規則</u>

$$(>)\ \frac{\Gamma \vdash M : A/B \quad \Delta \vdash N : B}{\Gamma, \Delta \vdash M^{\triangleright}N : A}$$

$$(<)\ \frac{\Delta \vdash N : B \quad \Gamma \vdash M : A\backslash B}{\Delta, \Gamma \vdash M^{\triangleleft}N : A}$$

<u>略記法</u>

$$\tau{/\!\!\!\backslash}\sigma \quad \stackrel{def}{\equiv} \quad \tau/\sigma \ \text{ or } \ \tau\backslash\sigma$$
$$M^{\triangleleft\!\triangleright}N \quad \stackrel{def}{\equiv} \quad M^{\triangleright}N \ \text{ or } \ M^{\triangleleft}N$$
$$(<>) \quad \stackrel{def}{\equiv} \quad (<) \ \text{ or } \ (>)$$

4　依存関係のあるコンビネータ

CL では $\mathbf{B}$ が $\mathbf{S}, \mathbf{K}$ から導出できることが知られているが、[2] における $\mathbf{B}$ が $\mathbf{S}, \mathbf{K}$ からは導出できないため、次のように $\mathbf{S}$ を設定した。

<u>**S**</u>

$$S_{/} : (A/C)/(B/C){/\!\!\!\backslash}(A/B\backslash C)$$
$$S_{\backslash} : (A\backslash C)\backslash(B\backslash C){/\!\!\!\backslash}(A\backslash B/C)$$

以下ではこの依存関係のある $\mathbf{S}, \mathbf{K}$ を用いることとする。

5　簡約規則

SDCL の簡約規則を次のように定める。

$$\mathbf{B}_{/}^{\triangleleft\!\triangleright} f^{\triangleright} g^{\triangleright} x \ \underset{CL}{\to} \ f^{\triangleright}(g^{\triangleright}x)$$
$$\mathbf{B}_{\backslash}^{\triangleleft\!\triangleright} f^{\triangleleft} g^{\triangleleft} x \ \underset{CL}{\to} \ f^{\triangleleft}(g^{\triangleleft}x)$$
$$\mathbf{S}_{/}^{\triangleleft\!\triangleright} f^{\triangleright} g^{\triangleright} x \ \underset{CL}{\to} \ (f^{\triangleleft}x)^{\triangleright}(g^{\triangleright}x)$$
$$\mathbf{S}_{\backslash}^{\triangleleft\!\triangleright} f^{\triangleleft} g^{\triangleleft} x \ \underset{CL}{\to} \ (f^{\triangleright}x)^{\triangleleft}(g^{\triangleleft}x)$$
$$\mathbf{K}_{/}^{\triangleleft\!\triangleright} x^{\triangleright} y \ \underset{CL}{\to} \ x$$
$$\mathbf{K}_{\backslash}^{\triangleleft\!\triangleright} x^{\triangleleft} y \ \underset{CL}{\to} \ x$$

6　計算論的性質の証明

※以下は SDCL の正当性を示すために必要な計算論的性質である。なお、証明の詳細は [3] を参照。

<u>Subject-Reduction</u>

$\Gamma \vdash X : \tau$ で $X \underset{CL}{\to} X'$ の時、$\Gamma \vdash X' : \tau$

<u>Church-Rosser の定理</u>

$\twoheadrightarrow$ を 1 回以上の簡約とすると

$M \twoheadrightarrow M_1, M \twoheadrightarrow M_2$ ならば、

ある N について、$M_i \twoheadrightarrow N \quad (i = 1, 2)$

<u>強正規化性</u>

$\vdash t : T$ ならば t は停止する

7　抽出性と演繹定理の関係

Extraction とは、wh 移動を指す。"a boy who John thinks that Mary loves" というフレーズが一つの例である。しかし複合名詞句制約のため、"a girl who John met a boy who loved" というフレーズは容認不可能である。しかしこの容認不可能なフレーズを型繰り上げ規則のある CCG では導出できてしまう。

<u>型繰り上げ規則</u>

$$(>T)\ \frac{\Gamma \vdash A : x}{\Gamma \vdash B/(B\backslash A) : \lambda P.Px} \qquad (<T)\ \frac{\Gamma \vdash A : x}{\Gamma \vdash B\backslash(B/A) : \lambda P.Px}$$

そこで、複合名詞句制約と関連のある演繹定理のSDCLにおける成立条件を証明した。

演繹定理

$$(DT_l)\frac{A:x,\Gamma\vdash B:M}{\Gamma\vdash B\backslash A:\lambda^{\triangleleft}x.M}\qquad (DT_r)\frac{\Gamma,A:x\vdash B:M}{\Gamma\vdash B/A:\lambda^{\triangleright}x.M}$$

また、演繹定理の証明に関連する構造規則の証明を行い、SDCLにおける構造規則を決定した。詳細は [4] を参照。

8 正規化解析を用いた構文解析

CCGのように、関数合成規則を持つ範疇文法では、文の単語数に対して指数的に構文解析結果が生じる可能性がある。例えば、"John sees vincent." という文に対しては以下のような複数の異なる構文解析結果が得られる。

(1) John sees vincent

1.
$$(>)\frac{\dfrac{\text{John}}{S/(S\backslash NP)}\qquad (<)\dfrac{\dfrac{\text{sees}}{S\backslash NP/NP}\qquad \dfrac{\text{vincent.}}{T\backslash(T/NP)}}{S\backslash NP}}{S}$$

2.
$$(<)\frac{(>B)\dfrac{\dfrac{\text{John}}{S/(S\backslash NP)}\qquad \dfrac{\text{sees}}{S\backslash NP/NP}}{S/NP}\qquad \dfrac{\text{vincent.}}{T\backslash(T/NP)}}{S}$$

意味が同じである全ての導出結果を計算することは非効率的である。そのため、一つの意味に対して唯一の解析結果（正規形）を見つける正規化解析が必要となる。

8.1 関連研究

[5] では関数合成と関数適用規則に対し、正規化制約を定めた。また、安全性や完全性についても証明されている。しかし、型繰り上げ規則の適用は、辞書内に限られていた。

[6] では [5] では示されていなかった一般合成規則と型繰り上げ規則についても対処できるように正規化制約の拡張を行い、その効果についても検証されている。

8.2 実装

本研究では、複合名詞句制約における言語的な理由、そして解析面での規則の適用可能性の理由から、型繰り上げ規則は規則としては設定せず、固有名詞等は辞書において繰り上げられた型を持つ。この場合、[5], [6] の正規化制約では正規形を決めることができない場合が発生するため、新たな正規化制約を提案し、それに基づいて構文解析器の実装を行った。

新正規化制約

The category which has a semantic representation of the form $\lambda P.P(x)$ cannot be functor in $<$ if the argument is the output of $> B$

8.3 日本語のかき混ぜ文の解析

標準的な語順は言語によって異なる。日本語ではSOVが標準的な語順だが、ある程度自由に語順を変更できる「かき混ぜ操作」があることが知られている。かき混ぜ文は以下のような日本語特有の組み合わせ規則により得られると仮定する。

$$(<\mathbf{C})\frac{X\backslash Y\backslash Z:f}{X\backslash Z\backslash Y:\mathbf{C}f\ (=\lambda y.\lambda z.fzy)}$$

しかし、$(<\mathbf{C})$ を繰り返し適用できてしまう点、単に連続する適用を除去するだけでは十分ではない点、など処理の観点からは大きな問題がある。この問題を解決するために、意味表示を考慮した以下の手法を提案する。

C の正規形

$\mathbf{C}(X(\mathbf{C}Y))$ という形の意味表示が XY と等価（$\alpha\beta$-同値）なら $\mathbf{C}(X(\mathbf{C}Y))$ は正規形ではない

詳細は [7] を参照。

9 まとめと今後の課題

本研究では部分方向性組み合わせ論理（SDCL）におけるコンビネータの依存関係を示し、システムの正当性を示すために必要なSubject-Reduction、合流性、正規化定理という3つの計算論的性質を証明した。

また、SDCLを定義したことで、CCGとCLのCurry-Howard対応が確立でき、これにより抽出性と演繹定理の関係について議論できるようになった。

さらに、実装面では正規化制約を用いたCCGでの英語の構文解析器を実装した。また、日本語のかき混ぜ文に関わる正規化制約の分析を行った。

今後の課題としては、コンビネータ **K** なしでの演繹定理の証明、解析器の精度向上などを考えている。

参考文献

[1] Steedman, M. J. (2000) The Syntactic Process (Language, Speech, and Communication). The MIT Press.

[2] Bekki, D. (2010) "Combinatory Categorial Grammar as a Substructural Logic — Preliminary Remarks —", In the Proceedings of LENLS 7, JSAI International Symposia on AI 2010. Campus Innovation Center, Tokyo, pp.70-83.

[3] 尾崎博子, 戸次大介. (2011) "部分方向性組み合わせ論理の計算論的性質とその証明", Technical Report of Department of Information Science, Ochanomizu University, OCHA-IS 10-2, February 7th, 2011.

[4] Ozaki, H. and Bekki D. (2012) "Extractability as the Deduction Theorem in Subdirectional Combinatory Logic", In the Proceedings of LACL 2012, pp.186-200.

[5] Eisner, J. (1996) "Efficient Normal-Form Parsing for Combiantory Categorial Grammar", In the Proceedings of ACL '96 Proceedings of the 34th annual meeting on Association for Computational Linguistics. pp.79-86.

[6] Hockenmaier, J. and B. Yonatan. (2010) "Normal-form parsing for Combinatory Categorial Grammars with generalized composition and typeraising", In the Proceedings of the Proceeding of COLING 2010. pp.465-473.

[7] 尾崎博子, 戸次大介. (2012) "部分方向性組み合わせ論理における日本語のかき混ぜ文の正規化制約", 第26回人工知能学会 JSAI 2012.