

日本語用言は階層的か连接的か：CCG による分析

戸次 大介

bekki@ecs.c.u-tokyo.ac.jp

東京大学 21 世紀 COE「心とことば—進化認知科学的展開—」

〒153-8902 東京都目黒区駒場 3-8-1

1 はじめに

筆者が進めている日本語 CCG プロジェクト (戸次 (forthcoming), 小嶋 (2006) 等) は、日本語の活用体系・統語構造・意味合成に対し、組合せ範疇文法 (Combinatory Categorical Grammar: CCG)*¹を用いて、言語学的妥当性、形式的厳密性、現象に対する網羅性を併せ持つ文法理論を構築することを目指している。

本発表では、日本語文法の記述枠組みとして CCG を用いる動機の一つとして、語尾の膠着における「階層性」と「接続性」の対立という問題を取り上げ、CCG では「関数合成規則」により、両者の対立が解消され、統一的な分析が可能となることを示したい。

2 日本語の用言における階層性と接続性

膠着語である日本語の統語・意味計算過程を考えると、以下 (A) のような、用言語幹・活用語尾・助動詞・接尾語等が連続する形式を、どのような構造として捉えるべきか、という問題がある。

(A) 小僧が和尚の饅頭を [食べ][始め][て][い][た][ら][し][い][の][だ]

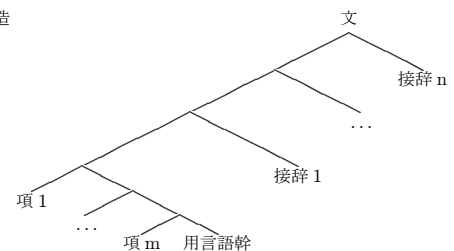
このような音連鎖に対しては、「どこからどこまでが一語であるか」という問題と相まって、一般に「階層性」と「接続性」という二つの相反する性質が見出される。^{*2}

*¹ Ades and Steedman (1982), Steedman (1996, 2000) 等。

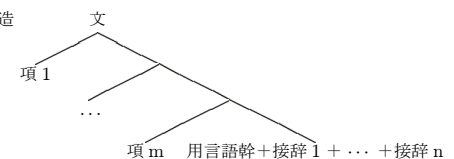
*² 本稿で論じている「論理の階層性」と「形態の接続性」の対立は、金水 (1997) において「論理と形態のミスマッチ」として指摘されている。

「しかし、本格的な統語構造への接近はやはり近代を待たねばならなかった。すなわち、橋本進吉、時枝誠記、渡辺実、北原保雄、水谷静夫らの学説である。これらの学説の間には、容易に克服しがたい鋭い対立が見て取れる。それは、日本語における形態論的構造と、文が表すであろう“論理構造”とのある種の“ミスマッチ”に由来していると見ることができる。橋本進吉のいわく

(1) 階層的構造



(2) 连接的構造



階層的構造を支持する要因には、意味表示におけるスコープの問題がある。たとえば、「らしい」は「小僧が和尚の饅頭を食べ始めていた」までをスコープに取るのに対し、「のだ」は「小僧が和尚の饅頭を食べ始めていたらしい」全体をスコープに取っている、等。一方、连接的構造を支持する要因には、「食べ始めていたらしいのだ」で一語であるという直観、形態論的要因、アクセント核付与規則等^{*3}が挙げられる。

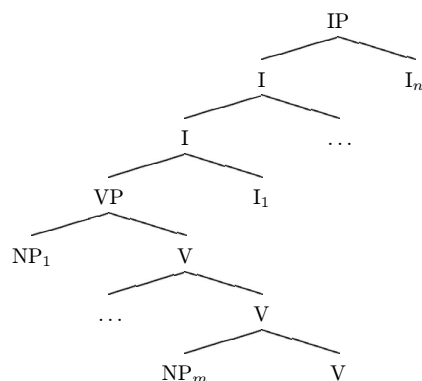
「文節文法」は形態論的構造を最大限に重視した統語分析であるが、文の論理構造を見るには不向きなモデルである。これに対し、時枝誠記のいわゆる「入れ子型」モデルは、簡潔ながら日本語の文構造から見事に論理構造を浮かび上がらせた。しかし、ここでは日本語の自然な形態論的単位 (橋本の「文節」、松下大三郎の「詞」、渡辺実の「成分」、山田孝雄、鈴木重幸の「語」) が壊されている。」(金水 (1997, p.140-141))

*³ Kitagawa (1986), Manning et al. (1999) を参照のこと。形態論的議論としては、「お食べ始めになる」「*お食べ始めていになる」のようなコントラストがあることから、「～始める」までは一語であり、「～ている」まではそうではない、という考え方があがる。アクセント核付与規則の問題については、機会を改めて論じたい。

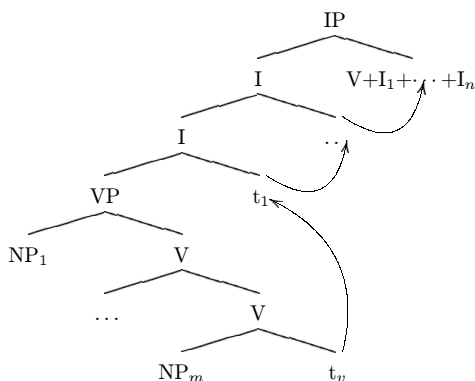
3 生成文法による「二本立て」分析の問題点

生成文法では、この相反する構造を、移動操作によって結びつけている*4。それによって、意味表示では (3) の構造、音韻表示では (4) の構造、という二本立ての分析を取る。

(3)



(4)



しかしこの分析は、文を生成する計算では比較的負荷が小さいが、逆計算、すなわち文から統語構造を解析する計算（パーズング）では計算量が大きくなってしまふことから、我々人間が現実に行う文処理では、違う形で文法が記述されていると考えざるを得ない。

CCG による分析は、この計算論的問題を回避しつつ、活用語尾・助動詞・接尾語等の接続について、新たな視点を提供するものである。

*4 Minimalism 以降の標準的な日本語分析では、(3) の構造が PF と LF に出力され、PF において移動が起こり (4) の構造に至る、というものであるが、この考え方自体は Kuroda (1965) に遡る。Kitagawa (1986) では逆に、連接的な構造を先に派生して PF と LF に出力し、LF において接辞の移動が起こり (3) の構造に至る、という分析であるが、PF と LF で別の構造を取り、その間を移動で結びつけているという点では共通している。

4 組合せ範疇文法 (CCG) による統合

戸次 (forthcoming) の分析では、(5) のように、語幹、活用語尾、接尾語の統語範疇をいずれも $S \backslash S$ として分析する（態のように、項の組み替えがある場合はもう少し複雑になる。）

(5)

太郎が	来
$T / (T \backslash NP_{ga})$: $\lambda P.P(taroo)$	$S \backslash NP_{ga}$: $\lambda x.\lambda e.kuru(e, x)$
なかつ	た
$S \backslash S$: $\lambda P.\lambda e.\sim Pe$	$S \backslash S$: $\lambda P.\lambda e.(ta(e); Pe)$

この統語範疇 $S \backslash S$ は、関数合成規則と組合せられることにより、興味深い二面性を示す。否定のスコープの曖昧性について考えてみよう。(4) のような文に対しては、以下 (B) の意味表示 a、b の両方を導出できる必要がある。

(B) 三人以上の学生が来ないなら、授業は中止だ。

a. $\sim 3(x)[gakusei(x); kuru(e, x)]$

b. $3(x)[gakusei(x); \sim kuru(e, x)]$

CCG による分析では、これらの意味表示の差を統語的な導出過程の差に帰着させることができる。まず、(B)a. に相当する導出は、以下のようになる。

(6)

三人以上の学生が	来	ない
$S / (S \backslash NP_{ga})$: $\lambda P.\lambda e.3(x)[gakusei(x); Pxe]$	$S \backslash NP_{ga}$: $\lambda x.\lambda e.kuru(e, x)$	$S \backslash S$: $\lambda P.\lambda e.\sim Pe$
S : $\lambda e.3(x)[gakusei(x); kuru(e, x)]$		
S : $\lambda e.\sim 3(x)[gakusei(x); kuru(e, x)]$		

一方、(B)b. に相当する導出は、以下のようになる。

(7)

三人以上の学生が	来	ない
$S / (S \backslash NP_{ga})$: $\lambda P.\lambda e.3(x)[gakusei(x); Pxe]$	$S \backslash NP_{ga}$: $\lambda x.\lambda e.kuru(e, x)$	$S \backslash S$: $\lambda P.\lambda e.\sim Pe$
S : $\lambda e.3(x)[gakusei(x); \sim kuru(e, x)]$		

「来」と「ない」には関数適用規則ではなく、関数合成規則が適用されている点に注意する。それによって、「ない」の語彙項目は同じものを用いているにも関わらず、二つの異なる読みを派生することが可能となっている。

Definition 4.1 (関数合成規則). (forward/backward functional composition)

$$\begin{array}{c} X/Y : f \quad Y/Z : g \\ \hline >B \quad X/Z : \lambda x.f(gx) \end{array} \quad \begin{array}{c} Y \setminus Z : g \quad X \setminus Y : f \\ \hline <B \quad X \setminus Z : \lambda x.f(gx) \end{array}$$

このように、統語範疇 $S \setminus S$ の要素と「左から係る量化子」の間では、導出順序が異なるとスコープの上下関係も異なる。一方、統語範疇 $S \setminus S$ の要素同士の間では、導出順序に依らず、常に右側の要素が上のスコープを取る。次の例を見てみよう。

$$(8) \quad \begin{array}{c} \begin{array}{c} \text{来} \\ \hline S \setminus NP_{ga} \\ : \lambda x.\lambda e.kuru(e, x) \end{array} \quad \begin{array}{c} \text{ない} \\ \hline S \setminus S \\ : \lambda P.\lambda e.\sim Pe \end{array} \\ \hline <B \quad \begin{array}{c} S \setminus NP_{ga} \\ : \lambda x.\lambda e.\sim kuru(e, x) \end{array} \quad \begin{array}{c} \text{だろう} \\ \hline S \setminus S \\ : \lambda P.\lambda e.daroo(Pe) \end{array} \\ \hline <B \quad \begin{array}{c} S \setminus NP_{ga} \\ : \lambda x.\lambda e.daroo(\sim kuru(e, x)) \end{array} \end{array}$$

$$(9) \quad \begin{array}{c} \begin{array}{c} \text{来} \\ \hline S \setminus NP_{ga} \\ : \lambda x.\lambda e.kuru(e, x) \end{array} \quad \begin{array}{c} \text{ない} \\ \hline S \setminus S \\ : \lambda P.\lambda e.\sim Pe \end{array} \quad \begin{array}{c} \text{だろう} \\ \hline S \setminus S \\ : \lambda P.\lambda e.daroo(Pe) \end{array} \\ \hline <B \quad \begin{array}{c} S \setminus NP_{ga} \\ : \lambda x.\lambda e.kuru(e, x) \end{array} \quad \begin{array}{c} S \setminus S \\ : \lambda P.\lambda e.daroo(\sim Pe) \end{array} \\ \hline <B \quad \begin{array}{c} S \setminus NP_{ga} \\ : \lambda x.\lambda e.daroo(\sim kuru(e, x)) \end{array} \end{array}$$

「ない」と「だろう」は共に統語範疇 $S \setminus S$ の要素である。(8) に示した二つの導出は、「来ないだろう」という形式について、これらの要素を異なる順序で組合せた場合を示しているが、最終的な統語範疇と意味表示は同一である。このように、統語範疇 $S \setminus S$ の要素同士の間では、関数合成規則をどのような順で適用しようと、常に同じ結果が導出され、スコープに関しては常に右の要素が上になるのである。

5 接続性とモノイド

我々が直観的に感じる「接続性」とは、どのような条件から成り立つ概念であろうか。接続という概念が二つの形態素を接合する操作であるとする、まず、そのような操作は結合則 (associativity) を満たしていると言えそうである。

$$(10) \quad \begin{array}{l} \text{食べ} + (\text{て} + \text{しまう}) = (\text{食べ} + \text{て}) + \text{しまう} \\ \text{食べて} + (\text{しまっ} + \text{た}) = (\text{食べて} + \text{しまっ}) + \text{た} \end{array}$$

また、この操作が交換則 (commutativity) は満たさないことが明らかである。更に、音声的に空の要素が存在

する考えると、少なくとも PF では次の等式が成り立つであろう。

$$(11) \quad \begin{array}{l} \text{食べてしまう} = \text{食べてしまう} + \varepsilon \\ = \varepsilon + \text{食べてしまう} \end{array}$$

これらの事実は、形態素の接続によって生じる形式が、抽象代数 (abstract algebra) でいうモノイド (monoid) と類似している^{*5}ことを示している。モノイドとは単位元 (identity element) を持つ半群 (semigroup) であり、あるいは逆元 (inverse element) を持たない群 (group) である。

Definition 5.1 (モノイド). モノイド (monoid) とは、以下のような性質を満たす集合 M と、二項演算 $+$: $M \times M \rightarrow M$ の組 $\langle M, + \rangle$ である。

- 結合則 (associativity): 全ての $a, b, c \in M$ に対して、 $(a + b) + c = a + (b + c)$ が成り立つ。
- 単位元の存在 (identity element): 全ての $a \in M$ に対して、 $a + \varepsilon = \varepsilon + a = a$ が成り立つような $\varepsilon \in M$ が存在する。

形態素の接続演算を表すのに「+」の記号が用いられることが多いのは、無意識にモノイドを指向しているとも言える^{*6}。

統語範疇が $S \setminus S$ である要素の集合は、関数合成規則を加法とするモノイド (monoid) をなしているからである。証明を簡単に示す。

証明. まず、 $S \setminus S$ の要素間では、関数合成規則に関して結合則 (associative law) が成り立つ。

$$(12) \quad \begin{array}{c} \begin{array}{c} P_1 \\ \hline S_{[2]} \setminus S_{[1]} \\ : f \end{array} \quad \begin{array}{c} \begin{array}{c} P_2 \\ \hline S_{[3]} \setminus S_{[2]} \\ : g \end{array} \quad \begin{array}{c} P_3 \\ \hline S_{[4]} \setminus S_{[3]} \\ : h \end{array} \\ \hline <B \quad \begin{array}{c} S_{[4]} \setminus S_{[2]} \\ : \lambda x.h(gx) \end{array} \\ \hline <B \quad \begin{array}{c} S_{[4]} \setminus S_{[1]} \\ : \lambda x.h(g(fx)) \end{array} \end{array}$$

^{*5} 形態素の接続がモノイドにおける演算と異なるのは、接続演算が $M \times M \rightarrow M$ という全関数ではなく、部分関数になっている点である。この意味で、形態素の接続は文字の接続とは異なる。

^{*6} 自然数は加法演算 $+$ に対して交換モノイド (commutative monoid) をなす。

(13)

$$\begin{array}{c}
\frac{P_1}{S_2 \setminus S_1 : f} \quad \frac{P_2}{S_3 \setminus S_2 : g} \\
<B \frac{\quad}{S_3 \setminus S_1 : \lambda x.g(fx)} \quad \frac{P_3}{S_4 \setminus S_3 : h} \\
<B \frac{\quad}{S_4 \setminus S_1 : \lambda x.h(g(fx))}
\end{array}$$

次に、 PF が空であり、 LF が恒等関数であるような語彙項目が、(統語素性の差を捨象すれば) 単位元として振る舞う。

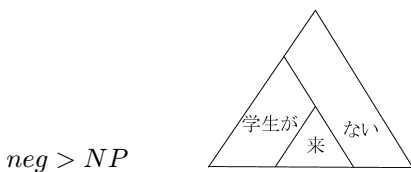
$$\begin{array}{c}
\frac{\emptyset}{S_2 \setminus S_1 : id} \quad \frac{P}{S_3 \setminus S_2 : f} \\
(14) \quad <B \frac{\quad}{S_3 \setminus S_1 : f}
\end{array}$$

$$\begin{array}{c}
\frac{P}{S_2 \setminus S_1 : f} \quad \frac{\emptyset}{S_3 \setminus S_2 : id} \\
(15) \quad <B \frac{\quad}{S_3 \setminus S_1 : f}
\end{array}$$

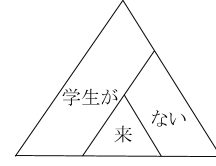
6 まとめ

このように、統語範疇 $S \setminus S$ の要素同士は、逆関数合成規則によって、「连接的に」膠着するのである。しかし、関数合成規則による膠着は、连接的でありながら、スコープの上下関係を保持する。すなわち、左側には「論理的な階層性」を示し、右側には「形態論的な接続性」を示すことになり、階層性と接続性を両立させる構造である。ただし、この分析では自然言語の統語構造における内芯性 (endo-centricity) は否定されることになる。

直観的には、本分析の提示する構造は以下の図のように考えることができる。階層的なスコープの上下関係を表現しうが、同じ向きの要素については连接的に接続しているのである。



$NP > neg$



参考文献

- Ades, A. E. and M. J. Steedman: 1982, 'On the Order of Words'. *Linguistics and Philosophy* 4(517-558).
- Kitagawa, Y.: 1986, 'Subjects in Japanese and English'. Doctoral dissertation, University of Massachusetts. Published by Garland, New York, 1994.
- Kuroda, S.-Y.: 1965, 'Generative Grammatical Studies in the Japanese Language'. Doctoral dissertation, MIT.
- Manning, C. D., I. A. Sag, and M. Iida: 1999, 'The Lexical Integrity of Japanese Causatives'. In: R. D. Levine and G. M. Green (eds.): *Studies in contemporary phrase structure grammar*. Cambridge University Press, pp. p.39-79.
- Steedman, M. J.: 1996, *Surface Structure and Interpretation*. The MIT Press.
- Steedman, M. J.: 2000, *The Syntactic Process (Language, Speech, and Communication)*. The MIT Press.
- 金水敏. 1997 「国文法」 益岡隆志, 仁田義雄, 郡司隆男, 金水敏. 「岩波講座 言語の科学 5 文法」第四章, p.119-158, 岩波書店.
- 小嶋大起, 戸次大介, 宮尾祐介, 辻井潤一. 2006. 「日本語 CCG の語彙項目獲得」, 第 176 回自然言語処理研究会, 鹿児島大学.
- 戸次 大介. (forthcoming) 「日本語文法の形式理論—活用体系・統語構造・意味合成—」, p.1-259, (日本学術振興会平成 20 年度科研費補助金「研究成果公開促進費 (学術図書)」に申請中)