

CCG による日本語文法記述の進捗と展望 ～活用体系・統語構造・意味合成～*

戸次 大介

bekki@is.ocha.ac.jp

お茶の水女子大学 大学院人間文化創成科学研究科
〒112-8610 東京都文京区大塚 2-1-1

概要

本プロジェクトでは、統語構造の記述に組合せ範疇文法 (Combinatory Categorical Grammar (CCG): Ades and Steedman (1982), Steedman (1996, 2000), Steedman and Baldridge (2007)), 意味表示の記述に高階動的論理 (higher-order dynamic logic: Bekki (2000); 戸次 (2009)) を用いた日本語の形式文法理論を構築している。本文法理論は、以下 1)~3) の要請を同時に満たす理論として完成しつつあり、本稿ではその進捗と展望について述べる。

- (1) a. 日本語の構文に対する網羅性
- b. 記号論としての形式的厳密性
- c. 活用体系・統語構造・意味合成の理論的統合性

1 文科系言語学と工科系自然言語処理の乖離

言語科学には、生成文法、形式意味論、日本語教育学、国語学 (日本語学) などに代表される「文科系言語学」と、音声／音韻解析、形態素解析、構文解析、意味解析などを計算機上で行う「工科系言語処理」の二つの流れがある。

この両者は、自然言語処理の概念が誕生して間もない 1980 年代までは、学問内容的にもコミュニティ的にも、それほど隔たりのないものであった。ところが 1990 年

代にさしかかると、自然言語処理のコミュニティにおいて、文科系言語学、ひいては記号論的アプローチそのものに対する失望感が蔓延するようになり、その後、統計的言語処理が成功を収めてからは、二つのコミュニティ間の溝は埋めがたいものとなってしまった。その結果、同じく自然言語を研究対象とする学問分野でありながら、互いの研究成果がほとんど参照されず、また研究者同士が話し合っても共通言語がほとんどない、という状況が続いている。この状況を打開するためには、文科系言語学と工科系言語処理に共通のプラットフォーム的文法理論が必要である。

2 プラットフォーム的文法理論の不在

しかし、1980 年代までの自然言語処理において明らかとなったことは、自然言語処理に役立つ文法理論というものが存在するとしたら、少なくとも (1) の三つの要請を満たすことが求められる、ということである。

ところがそれらを「同時に満たす」理論を構築することは、文科系言語学の側からすると厳しい要請である。理論言語学におけるこれまでの研究は、説明対象を個別の現象に絞り込んだものが多い。それらの研究は、分野ごとのプラットフォーム理論 (変形生成文法における Minimalism 等) を暗黙に仮定してはいるが、個々の研究を寄せ集めても、必ずしも首尾一貫した文法理論を形成する保証はない、という問題がある。

その結果、これまで日本語文法を、形式的な体系を用いつつ包括的に記述するという試みは、変形生成文法に基づいた「変形文法と日本語」(井上 (1976)), GPSG/HPSG に基づいた “Japanese Phrase Structure Grammar” (Gunji (1987)) など、数えるほどしか存在していない。またそのことが、自然言語処理コミュニティ

* 本研究は、科学研究費補助金 (若手 (B) 20700125 「CCG と高階動的論理による等位接続構文の形式的記述」(研究代表者: 戸次大介) 平成 20 年度~21 年度) および科学研究費補助金 (基盤研究 (C) 20500148 「確実性アノテーション: 『確実性判断を表す意味的文脈』を記述したコーパスの構築」(研究代表者: 川添愛) 平成 20 年度~22 年度) の助成を受けている。

品詞素性値	俗称	品詞素性値	俗称	品詞素性値	俗称
<i>v</i> (<i>verb</i>)	動詞	<i>v :: K</i>	カ行変格活用動詞	<i>n</i> (<i>nominal predicate</i>)	状詞
<i>v :: 5</i>	五段活用動詞	<i>v :: S</i>	サ行変格活用動詞	<i>n :: da</i>	
<i>v :: 5 :: k</i>	カ行五段活用動詞	<i>v :: Z</i>	ザ行変格活用動詞	<i>n :: na</i>	
<i>v :: 5 :: s</i>	サ行五段活用動詞	<i>v :: URU</i>	ウル型活用動詞	<i>n :: no</i>	
<i>v :: 5 :: t</i>	タ行五段活用動詞	<i>a</i> (<i>adjective</i>)	形容詞	<i>n :: tar</i>	
<i>v :: 5 :: n</i>	ナ行五段活用動詞	<i>a :: i</i>	イ形容詞	<i>n :: ni</i>	
<i>v :: 5 :: m</i>	マ行五段活用動詞	<i>a :: i :: auo</i>	アウオ段イ形容詞	<i>n :: ∅</i>	
<i>v :: 5 :: r</i>	ラ行五段活用動詞	<i>a :: i :: i</i>	イ段イ形容詞	<i>n :: to</i>	
<i>v :: 5 :: w</i>	ワ行五段活用動詞	<i>a :: i :: NAS</i>	ナシ型活用形容詞	<i>n :: da ⊔ na</i>	
<i>v :: 5 :: g</i>	ガ行五段活用動詞	<i>a :: BES</i>	ベシ型活用形容詞	<i>n :: da ⊔ no</i>	
<i>v :: 5 :: z</i>	ザ行五段活用動詞		
<i>v :: 5 :: b</i>	バ行五段活用動詞			<i>n :: da ⊔ na ⊔ ni</i>	形容動詞
<i>v :: 5 :: IKU</i>	イク型活用動詞			<i>n :: da ⊔ no ⊔ tar</i>	判定詞
<i>v :: 5 :: YUK</i>	ユク型活用動詞		
<i>v :: 5 :: ARU</i>	アル型活用動詞			<i>n :: da ⊔ na ⊔ no</i>	
<i>v :: 5 :: ARU</i>	アル型活用動詞			<i>⊔tar ⊔ ni ⊔ ∅ ⊔ to</i>	
<i>v :: 5 :: NAS</i>	ナサル型活用動詞				
<i>v :: 5 :: TOW</i>	トウ型活用動詞				
<i>v :: 1</i>	一段活用動詞				

図1 戸次 (2010) における用言の品詞分類

における「言語学者を雇っても一向に役に立つ文法を書かない」という失望につながったと思われる。

したがって、自然言語処理の研究者が言語学の成果を参照しようとする、諸論文の統合方法および扱われていない現象の扱い方については独力で考え出さなければならぬ状況にある。言語学者ではない言語処理研究者が、その方法で一貫性のある体系を作ることにはきわめて困難である。この事情が、近年の言語学と言語処理との乖離を引き起こし、結果的に、理論的・記述的研究の成果が工学的応用研究にほとんど利用されないという事態に至っている。

3 1990年代以降の理論言語学における進展

しかしその後の二十年間において、言語学は、特に数理論理学との境界領域において進展し、自然言語のより本質的な構造を捉えた理論が、統語論・意味論の双方で現れ始めている。

統語論においては、組合せ範疇文法（以下 CCG）が登場した。チョムスキー流の生成文法において「*wh* 移動」等の長距離依存の現象は「移動操作」を必要とする主張されているが、「移動操作」は非局所的操作であるため、言語処理においては計算量上、問題視されている。これに対して CCG の分析は生成文法や HPSG^{*1}と

は本質的に異なり、移動現象が局所的演算として定義できるだけでなく、様々な等位接続構文が導出できるという経験的妥当性と、統語操作が Curry の組合せ論理における組合せ子に（したがって Curry=Howard 同型から、ある種のヒルベルト流証明論に）還元されるという説明的妥当性が得られている。

一方、意味論においては、1990年代初頭に動的述語論理 (Groenendijk and Stokhof (1991)) が登場し、それまでの動的意味論 (Kamp (1981)、Heim (1982) 等) の記述力と、Montague (1973) 以来の合成原理に基づく形式意味論が統合されるというパラダイムシフトが起こった。この「動的論理」に基づく形式意味論はその後順調に発展し、広範囲の言語現象を記述することが可能になりつつある。

4 CCG に基づく形式日本語文法：戸次 (2010)

そこで本プロジェクトでは、統語論として CCG、および意味論として動的述語論理を発展させた^{*2}高階動的論理を採用し、新たな日本語の形式文法理論を提示した (戸次 (2010))。この成果は著書「日本語文法の形式理論」として一冊の本にまとめ、くろしお出版より出版されている。

現象の捉え方としては生成文法を踏襲していると言える。

^{*2} Bekki (2000)、戸次 (2009) を参照のこと。

^{*1} HPSG では移動操作を局所的操作の組合せに分解したものの、

活用種別素性値	五段動詞	一段動詞	カ変動詞	サ変動詞	形容詞	活用種別素性値	状詞	タ	ダ	マス	ズ	ヌ
stem	語幹形	飛	食べ	の	の	な	非丁寧/丁寧	過去	過去	丁寧	否定	否定
ustem	文語連用接続形	—	—	—	—	の	静か	—	—	—	—	—
neg,-l	打消形	飛+ば	食べ+の	の+こ	の+し	—	—	—	—	—	—	—
cont,-l	連用形	飛+び	食べ+の	の+き	の+し	な+く	—	たり	だり	—	—	—
term,-l	終止形	飛+ぶ	食べ+る	の+くる	の+する	な+い	静か+だ/です	た	だ	ます	—	—
attr,-l	連体形	飛+ぶ	食べ+る	の+くる	の+する	な+い	静か+な/—	た	だ	ます	—	—
hyp,-l	条件形	飛+べ	食べ+れ	の+くれ	の+すれ	な+けれ	—	たら	だら	—	—	—
imp,-l	命令形	飛+べ	食べ+ろ	の+こい	の+しろ	—	—	—	—	ませ・まし	—	—
pre,-l	推量形	飛+ば+う	食べ+よ+う	の+こよ+う	の+しよ+う	な+かろ+う	—	たろ+う	だろ+う	ましょ+う	—	—
te,-l	テ形	飛+ん+で	食べ+の+て	の+き+て	の+し+て	な+く+て	静か+で/—	たって	だって	まし+て	—	—
ni,-l	二形	飛+び+に	食べ+の+に	??の+き+に	の+し+に	—	静か+に/—	—	—	—	—	—
neg,+l	文語打消形	飛+ば	食べ+の	の+こ	の+せ	な+から	静か+なら/—	—	—	ませ	—	—
cont,+l	文語連用形	—	—	—	—	の+う	—	—	—	—	—	—
term,+l	文語終止形	—	—	—	の+す	な+し	静か+なり/—	たり	だり	まする	ず	ぬ
attr,+l	文語連体形	—	—	—	—	な+し	静か+なる/—	—	—	まする	ざる	ぬ
hyp,+l	文語条件形	飛+ば	—	—	—	な+き	静か+なれ/—	—	—	ますれ	ざれ	ね
imp,+l	文語命令形	飛+べ	食べ+よ	—	の+せよ	な+かれ	静か+なれ/—	—	—	ませい	—	—
pre,+l	文語推量形	飛+ば+ん	食べ+の+ん	の+こ+ん	の+せ+ん	な+から+ん	静か+なら+ん	—	—	ませ+ん	—	—
te,+l	文語テ形	—	—	—	—	の+う+て	—	—	—	—	—	—
ni,+l	文語二形	—	—	—	—	—	—	—	—	—	—	—
euph	過去接続形	—	—	—	—	—	静か+だっ/でし	—	—	まし	—	—
euph:t	夕接続形	—	食べ+の	の+き	の+し	な+かつ	—	—	—	—	—	—
euph:d	夕接続形	飛+ん	—	—	—	—	—	—	—	—	—	—
mod	様相接続形	—	—	—	—	—	—	—	—	—	—	—
mod:u	ウ接続形	飛+ば	食べ+よ	の+こよ	の+しよ	な+かろ	—	たろ	だろ	ましょ	—	—
mod:d	ダロウ接続形	飛+ぶ+の	食べ+る+の	の+くる+の	の+する+の	な+い+の	静か+の/—	た+の	だ+の	ます+の	—	—
mod:s	ソウ接続形	飛+び	食べ+る	の+き	の+し	な+き	静か+の/—	—	—	—	—	—
vo	態接続形	—	—	—	—	—	—	—	—	—	—	—
vo:r	受身接続形	飛+ば	食べ+ら	の+こ+ら	の+さ	—	—	—	—	—	—	—
vo:s	使役接続形	飛+ば	食べ+さ	の+こ+さ	の+さ	—	—	—	—	—	—	—
vo:e	可能接続形	飛+べ	??食べ+れ	??の+こ+れ	—	—	—	—	—	—	—	—

図2 戸次 (2010) の活用表

この研究は、通常の理論言語学・形式意味論の文脈において、単独の現象を単発的に分析するタイプの研究とは、規模的にも目的論的にも異なる。「日本語文法の形式理論」は、日本語文法の理論的（形態論的・統語論的・意味論的）記述を確立し、その成果を自然言語処理において（(1)で述べた網羅性・形式性・統合性の各面で）利用可能な質で提供することを目指している。^{*3}

第一に「網羅性」については、日本語の

語幹/ 活用語尾/ 助動詞/ 接尾語/ 態（ヴォイス）

といった単文構造から、

連体節（関係節）/ 連用節/ 条件節/ 引用節

といった複文構造に至るまで、広範囲の現象をできるだけ例外のないように捉えている。第二に「形式性」については、論理学レベルの厳密性を確保しており、第三に「統合性」については、これまで音韻論もしくは形態論において説明すべきものとされてきた「活用体系」を、統

^{*3} 今のところ、自然言語処理・理論言語学のいずれのコミュニティにおいても、CCGの文法理論としての知名度は十分に高いとは言えない。現時点で挙げられる日本語 CCG 理論の応用研究・発展研究には小嶋 (2006)、尾崎 (2010) 等がある。

語論で扱うという新たな提案をしており、また、意味表示の記述に独自の高階動的論理を用いることによって、日本語の諸構文における意味合成のプロセスを包括的・明示的に示している。これはおそらく先例のない試みである。

また、理論言語学的な意義としては、先行研究において網羅的記述の妨げとなっていた諸現象（活用、接尾辞の接続、かき混ぜ等）に対して、関数合成規則 (Functional Composition Rule)、関数置換規則 (Functional Substitution Rule) 等、CCG 特有の規則による新たな解決法を示している。^{*4}

5 まとめ

このように、本プロジェクトで提示する日本語文法は、文科系言語学である理論言語学・記述言語学と、工科系言語学である自然言語処理・計算言語学に、共通の理論的プラットフォームを与えうるものである。また、これまで理論言語学・記述言語学において蓄積されてきた知見を再解釈・統合して言語処理に「到達」させ、言語処

^{*4} この点については戸次 (2008) で詳しく述べたので、ここでは割愛する。

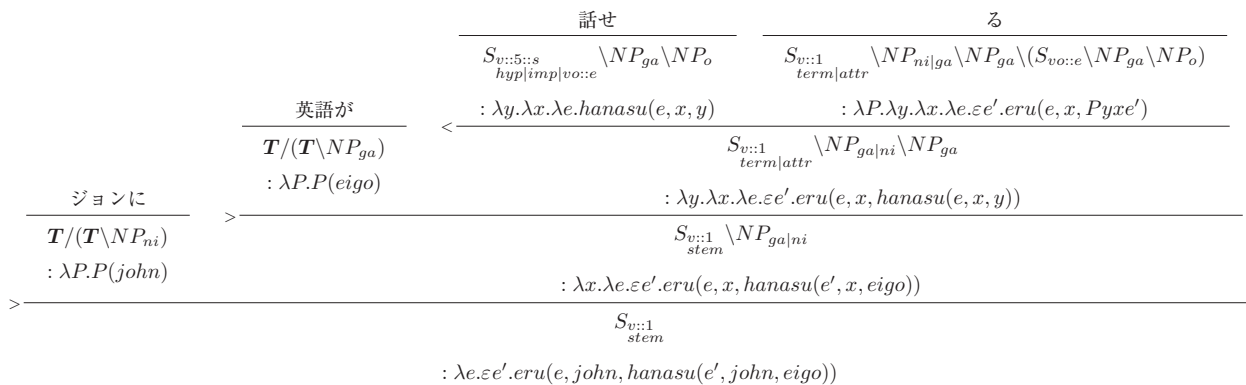


図3 戸次 (2010) による導出例：可能態

理における言語学の「復権」を目指す試みである。

参考文献

Ades, A. E. and M. J. Steedman: 1982, 'On the Order of Words'. *Linguistics and Philosophy* 4(517-558).

Bekki, D.: 2000, 'Typed Dynamic Logic for Compositional Grammar'. Doctoral dissertation, University of Tokyo.

Groenendijk, J. and M. Stokhof: 1991, 'Dynamic Predicate Logic'. *Linguistics and Philosophy* 14, 39-100.

Gunji, T.: 1987, *Japanese Phrase Structure Grammar: A Unification-based Approach*. Dordrecht: D. Reidel.

Heim, I.: 1982, 'The Semantics of Definite and Indefinite Noun Phrases'. Ph.d dissertation, University of Massachusetts. Published 1989 by Garland Press, New York.

Kamp, H.: 1981, 'A Theory of Truth and Semantic Representation'. In: J. Groenendijk, T. M. Janssen, and M. Stokhof (eds.): *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre Tract 135.

Montague, R.: 1973, 'The Proper Treatment of Quantification in Ordinary English'. In: J. Hintikka, J. Moravcsic, and P. Suppes (eds.): *Approaches to Natural Language*. Dordrecht: Reidel, pp. 221-242.

Steedman, M. and J. Baldridge: 2007, 'Combinatory Categorical Grammar'. In: R. Borsley and K. Borsjars (eds.): *Non-Transformational Syntax*. Black-

well.

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.

井上和子. 1976. 「変形文法と日本語(上)(下)」, 大修館書店.

尾崎有梨, 櫻井加奈子, 浅井健一, 戸次大介. 2010. 「証明木作成プログラムを用いた CCG 統語導出の実装」, 言語処理学会第 16 回年次大会発表論文集 (掲載予定), 東京大学.

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」, 情報処理学会研究報告. 自然言語処理研究会報告. 2006(124) pp.75-80.

戸次大介. 2008. 「日本語用言の階層性と接続性: CCG による分析」, 言語処理学会第 14 回年次大会発表論文集, pp.1132-1135, 東京大学.

戸次大介. 2009. 「型付きラムダ計算による自然言語の動的意味論」, 第 23 回人工知能学会全国大会オーガナイズドセッション「意味と理解のコンピューティング」招待講演 (2009/6/19).

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 -」, くろしお出版.