# 論理意味論にもとづく日本語文解析文法・辞書の開発

代表研究者 吉本 啓 東北大学高等教育開発推進センター教授 共同研究者 Alastair J. Butler 日本学術振興会外国人特別研究員

#### 1 はじめに

本研究は、共同研究者であるバトラーが考案した、論理意味論にもとづく文法理論であるスコープ制御理論(Scope Control Theory または SCT, Butler 2007, 2010a, 2010b)を用いて、日本語のコンピュータ解析のための基礎的な文法規則体系および辞書を構築することを目的として行った。

形式統語論と形式意味論とを組み合わせて言語処理に応用する試みは80年代に盛んであったが、その後、大量のテキスト・データを収集して統計モデルを構築する手法に取って代わられた。しかしこの数年来、この方法に対して、これ以上の性能の向上がもはや限界に達しており、また当初は画期的と思われたものの、処理性能の評価方法が確立された現在ではそれ程のものではないことが分かるなど、批判も強くなっている。このことから、形式統語論・意味論にもとづく規則主導型の処理システムをプロトタイプとして作り、これを利用した統計処理モデルを開発する方法が提案されている。本研究は上記の研究動向を視野に入れ、情報通信の発達により求められている高度の言語処理システムの開発に利用することを考慮して、日本語の主要な構文を解析するための文法規則および基本辞書を構築し、新たな観点からの文解析の手法を確立することを目的としている。

代表研究者らはこれまでの形式文法理論にもとづく文法開発の過程で、日本語文法研究者である南不二男 (1974)の文構造 4 階層説の重要性に気付いてその基本原理を取り入れ、変項を階層的に導入することによって、複文における主語・主題の一致や時制解釈に関する現象を統一的に説明することに成功してきた。このことを通じて、日本語の文構成のもっとも普遍的な原理を解明すると同時に、最低限度の統語処理によって深い意味解析を行うための展望が開けた。申請者らは南が語と語の共起関係にもとづいて仮定した文階層が、スコープ制御理論が可能にする変項の導入・制御のいくつかのバリエーションの1つに対応することを示した。これによって自説の普遍性を示すとともに、南理論に集積された日本の言語研究の世界における意義を明らかにすることが出来た。我が国における言語データの収集とその分析の蓄積は、質・量ともに他の先進国に決して劣るものではないが、この四半世紀間の世界における言語研究の革新が取り入れられていないために、世界の最先端の研究者にはほとんど注目されてこなかった。これらの文法研究の成果を申請者らの理論にもとづいて解釈、形式化することを通じて、その普遍的な意義を明らかにすることが研究の大きな動機である。また同時に、上記のように言語処理研究者からの論理文法の再認識がなされていることもあり、言語処理への応用を考慮した文解析システム・プロトタイプの開発を行った。

#### 2 スコープ制御理論

スコープ制御理論 (Butler 2007, 2010a, 2010b) は、自然言語の文の意味理解を行うために動態意味論 (Dynamic Semantics, Groenendijk and Stokhof 1991, Kamp and Reyle 1993) を拡張発展させた形式言語 理論である。形式文法研究において一般に文の構造の根幹であるとされる依存関係 (dependencies) に関して、従来の諸理論はこれを統語論的なものとして捉えてきたのに対し、SCT では意味論的な関係として形式化・処理を行う。

従来の形式文法理論では、述語論理によって捉えられた文の意味構造と、言語学的な文の解析構造(文法構造)とが乖離しているという問題を解決し、多様な構文を解析するために、統語構造やそれにもとづく統語素性の複雑な操作を行ってきた。SCT は、この問題に対して根本的に異なるアプローチを取る。すなわち、自然言語の文構造は、意味の基本的な構成要素であるスコープ(変項の値)を導入したり、操作あるいは評価することによって形作られる依存関係を直接反映したものであると考えるのである。単純な表層統語構造のみを仮定し、スコープの操作によって、文の意味解析(評価)を行う。文の解析や解釈において最も重要な、文が文法的(well-formed)であるための条件を、SCT は文の各領域において必要とされるスコープの数

と、割当関数によって供給されるスコープの数とが一致することであると定義する(「表現統合性条件」と呼ばれる)。この条件の下に、文を構成する領域ごとにスコープの導入や、他の変項へのシフトや、新たなスコープの出現による一時的な不活性化を行うことにより自然言語の文の複雑な成り立ちを体系的にシミュレートする。

文法役割(主語・目的語等)を表す束縛名称(binding names)に対してその値であるスコープのリストが与られ、置換え等の制御を実行することで文の意味解析が行われる。この目的のために10個の基本オペレータが用意されている。さらに、基本オペレータにもとづいて糖衣構文(syntactic sugar)を定義する(英語解析用文法では49個)ことによって、文法記述を容易に行い、可読性を高めている。

ごく単純な文の意味解析例を以下に示す。例文(1a)に対応する SCT 式(1b)が意味評価への入力となる。 (2a)は関与する束縛名称を、(2b)は初期値としての、スコープを伴う束縛名称を表す。 (2c, d, e) は糖衣構文による各単語の語義記述である。 (2e) が量化詞(quantifiers)としてスコープ(x)の主語句と述語句への置換えを行い、これと(2c, d)の動詞・名詞の語義記述とが共に評価されて、(3)の述語論理式を生じる。

- (1) a. A girl similes.
  - b. a girl // smiles
- (2) a. 1c = ["h", "x"]
  - b. fh = ["e"]
  - c. smiles = r lc fh ["x"] nil "smiles"
  - d. girl = r lc fh ["h"] nil "girl"
  - e.  $a = \lambda f$ . some lc fh "e" [f]
- (3)  $\exists g: (g, (1b))^{\bigcirc} = girl(x) \land smiles(x)$

SCT は英語、日本語の他に、フランス語、ドイツ語、中国語、Tukang Besi 語(インドネシア)等、多様な言語に適用され、それらの意味解析に有効であることが確認された。特に英語に関しては、Butler (2010a)に示したように、各構文の意味解析法定式化の問題は解決済みである。そこで提案された意味解析規則は全て Standard ML を用いてインプリメントされている。

## 3 英文解析実験

制限の無い英語テクストを意味表示(述語論理式)へと自動解析する実験を行うことによって、SCT による一般的な意味解析の方法を確立した(Butler, Miyao, Yoshimoto and Tsujii 2010, Miyao, Butler, Yoshimoto and Tsujii 投稿中)。表層的な統語解析によって、意味評価の入力となる SCT 式を得る。そのためには近年、統計や機械学習を用いたローバストな手法が開発されている。従来、深い意味解析のために用いられてきた方法では複雑な統語的構造操作が行われ、多数の統語規則のコントロールに関して問題を生じた。これに対し、SCT は、意味解析の前にローバストな統語解析器により表層解析を行うので、より正確な結果を得ることが出来る。これにより得られた SCT 式に対し、意味的依存関係に着目するスコープ操作を行って、意味解析結果(述語論理式)を得る。

統語解析器として、HPSG にもとづく解析器である Enju (Miyao and Tsujii 2008) を使用する。その理由は、Enju がカバー率が高い上にきわめて正確であること、また、SCT が必要とする詳細な統語情報を提供することが出来ることにある。ただし、その出力結果から SCT が入力として想定する形式へと変換するプログラムが必要であり、これは Enju 開発者の宮尾氏と共同で開発した。

以上のシステムにより英文意味解析実験を行った結果、無制限の英文テクストを述語論理式による意味表示へと自動翻訳するにあたって、カバー率の高いシステムを構築出来ることが分かった。開発用の文セット (Penn Treebank の一部、Marcus, Marinkiewicz and Santorini 1993) の84.2% に対して完全な意味表示を与え、また全く新しい生データ (Penn Treebank から選んだ1,518 文) の82.7% に対して完全な論理式を出力することが出来ることが判明した。このことから、自然言語の文の意味の自動処理に関して SCT が有効であると考える。以下に、Wikipedia の英文記事の自動意味解析の結果を挙げる。人手による修正は加えていない。前置詞句が出現する場合は、述語の構造が拡張され、文法項の一種として扱われる。代名詞の先行詞は文脈的要因によって決定される。

英文	述語論理式	
Space Invaders is an arcade video game designed by Tomohiro Nishikado. It was originally manufactured and sold by Taito in Japan, and was later licensed for production in the United States by the Midway division of Bally.	exist x y z (arcade_video_game(x) ^ design-passive +by(x, tomohiro_nishikado) ^ space_invaders = x ^ exist u (u = choose_(space_invaders, x) ^ originally_manufacture-past-passive +by +in_(u, taito, japan) ^ sell-passive +by +in_(u, taito, japan) ^ midway_division +of_(z, bally) ^ production +in_(y, united_state) ^ later_license-past-passive +for_ +by_(u, y, z)))	
Space Invaders is one of the earliest shooting games and features two-dimensional graphics.	exist x y z (early_shooting_game(x) ^ one +of_(y, x) ^ space_invaders = y ^ two-dimensional_graphic(z) ^ feature(space_invaders, z))	
With a laser cannon the aim is to defeat waves of aliens.	exist x y (laser_cannon(y) ^ aim(x) ^ exist z u (alien(z) ^ wave +of_(u, z) ^ defeat +with_(x, u, y)))	
Nishikado drew inspiration from popular media in designing the game. He had to design custom hardware and some new development tools to complete it.	exist x y z u (event(u, game(x) ^ design-prog(nishikado, x)) ^ popular_medium(z) ^ inspiration(y) ^ draw-past +from_ +in_(nishikado, y, z, u) ^ exist v (v = choose_(nishikado, u, z, y, x) ^ have-past(v, exist w x1 (custom_hardware(w) ^ design(v, w) ^ new_development_tool(x1) ^ design(v, x1) ^ exist x2 (x2 = choose_(x1, w, u, z, y, x) ^ complete(v, x2))))))	
First released in 1978, Space Invaders became very successful and popular, and caused a temporary shortage of 100-yen coins in Japan. By 2007, it had earned Taito a total of US 500 million dollars.	exist x y z u (very_successful(space_invaders) ^ popular(space_invaders) ^ 100-yen_coin +in_(x, japan) ^ temporary_shortage +of_(y, x) ^ cause-past(space_invaders, y) ^ exist v (v = space_invaders ^ first_release-passive +in_(v, 1978)) ^ exist v (v = choose_(space_invaders, japan, y, x) ^ us_500_million_dollar(z) ^ total +of_(u, z) ^ earn-perf-past +by_(v, taito, u, 2007)))	

### 4 日本語の解析

#### 4-1 南の4階層と日本語の文構造

従来の形式意味論にもとづく日本語研究は、英語を中心とする欧米語の研究成果をそのまま当てはめるのが普通であった。しかし、これらの研究は、句構造によって捉えられる文構造、特に名詞句と動詞句の文に対する役割が明瞭な欧米語に適応して発展してきたものである。これに対して日本語は、格名詞句の省略が多く、またそれらの間の語順が比較的に自由であり、さらに冠詞の無いことから量化詞に関する理論をそのままでは適用しにくい、という特徴がある。

他方、実際に使用された日本語文のコーパス調査を踏まえて、南(1974)は日本語の文に4段階の階層を認めている。これは文構成素の間の共起関係から導き出されたもので、通常は日本語文の文法的(統語論的)構造であると見なされている。しかし、より正確には、文の述語を中心とするもっとも中核的な層(A層)から文脈に直接関わるもっとも外側の層(D層)まで、文の形式と意味を段階的に構成していくための各ステップであると捉えるべきである。

代表研究者はこれまで、南の文 4 階層節にもとづいた日本語文の用法や意味の分析を行ってきた(吉本 1993, Yoshimoto 1998, 吉本 2004)。2007年に至って、同説は SCT が可能にするスコープの導入・制御のいくつかのバリエーションの 1 つに対応し、後者によって日本語文に関する諸現象が形式意味論の観点から根源的な説明を与えられることに気づき、バトラーと共同研究を行ってきた。その結果、スコープが階層的に導入されると考えることによって、複文における主語・主題の一致と時制解釈とが統一的に説明されることを示した(Butler, Nakamura and Yoshimoto 2009)。また、副助詞「も」を用いた全称量化詞的表現(「どんな・・・も」等)を一般的、構成的観点から説明することにも成功した(Butler 2009)。さらに、南の文階層説に対して反例を指摘して同説が成り立たないことを主張する論者に対して、反例のほとんどは文階層説を保ちながら説明することが可能であり、SCT による形式化になじむことを示した(Yoshimoto, Nakamura and Butler 2009)。

#### 4-2 複文における主語・主題の一致と時制解釈

日本語の従属節もまた南(1974)によって、4 階層のどれかに属するものとされる。このような従属節に属する階層は、主文と従属節の間の主語・主題間の一致や2つの文の時制解釈に大きく関係することが知られている(三上 1970, Kuno 1973)。

A 層に属する従属節「~て」等において、従属節の主語は主題化されているか否かに関わらず、主文の主語に一致する。B 層に属する従属節「~と」等においては、主題化されていない主語「~が」は主文主語と必ずしも一致しない(例文(4a))のに対し、主題化されて「は」を伴う場合、2つの主語は一致する(例文(4b))。

- (4) a. [太郎<sub>i</sub>が上着を脱ぐ]<sub>B</sub> と φ<sub>j</sub> ハンガーに掛けた。
  - b. 太郎  $_{i}$  は  $[\phi_{i}$  上着を脱ぐ]  $_{B}$  とハンガーに掛けた。

C層の従属節「~が」等において、その主語の指示対象は全く主文主語とは独立している。

また、A 層に属する従属節においては、述語は過去の助動詞「た」によりマークされることはない。B 層に属する従属節「 $\sim$ ので」等については、主節述語が過去助動詞「た」によりマークされている場合は、従属節の述語の時制は主節のそれとの関係において解釈される。例文(5)の場合、従属節述語の表す出来事時(E。)は主節のそれ(E<sub>。</sub>)に後続するという解釈を受ける(E<sub>。</sub> < E。)。

### (5) [春子が<u>卒業する</u>]<sub>B</sub> **ので** 一緒に<u>旅行した</u>。

これに対し、主節述語が「た」を伴わない場合、従属節述語の出来事時は主節述語の時制に関係なく、発話時との関係において解釈される。例文(6)では、従属節述語の出来事時は発話時(n)を包含する(n⊆ E<sub>m</sub>)。

#### (6) [春子が休んでいる]<sub>R</sub> ので 悲しい。

これに対して、C層の従属節においては、その述語の時制解釈は主節のそれとはまったく独立して行われる。 SCT を用いて、これらの現象をスコープの階層的導入という観点から統一的に説明することに成功した (Butler, Nakamura and Yoshimoto 2009, Yoshimoto, Nakamura and Butler 2009)。

文の主語に関する情報は束縛名称 "ga"のスコープによって示される。これは B 層において導入され、外側からはアクセスできないものとする。これに対し、主題は束縛名称 "wa"に対して与えられるスコープによって示される。その束縛はすでに文脈において与えられているものと考える。例文 (4a) は SCT により、図 1 の意味評価を与えられる。従属節の主語の指示対象は "ga"のスコープ y であるが、これは従属節の中で閉じている。主節の主語は省略された主題であり、"wa"のスコープ x であるが、これは文脈によって束縛されると考える。x と y が必ずしも同一でないことから、例文 (4a) の読みが説明される。これに対し、図 2 に示すように、例文 (4b) においては、従属節の主語解釈は、主題を表す "wa" から "ga"へとシフトされたスコープ x にもとづいて行われる。主文主語も "wa"のスコープにもとづいて解釈されることから、この例文における主語の一致が説明される。

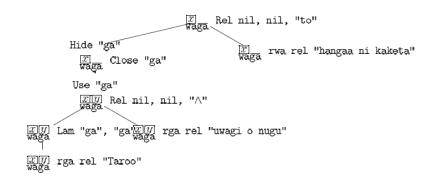


図1: 例文(4a)の意味解釈過程

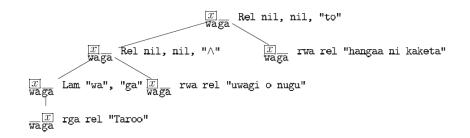


図2: 例文(4b)の意味解釈過程

複文の時制解釈に関しても、主題/主語における "wa"/"ga" と同様、文脈において広域的に束縛される "ev" と、B 層内部のみで束縛される "it" とを想定する。両方とも、出来事時の表示にも用いられる。述語 が過去の助動詞「た」でマークされている時は "ev" の上に新しいスコープが導入され、これにもとづいて 時制解釈が行われる。また、述語が非過去の場合は "it" の上に新しいスコープが導入され、これにもとづ き時制が解釈される。例文(5)の意味解釈結果は(5')のようになる。主節の出来事時は y、従属節のそれは z であり、非過去形の出来事時は "ev" のスコープ以降と解釈されるので、y < z とされる。すなわち、従属節の述語は主節の出来事時との関係において時制解釈を与えられる。

(5')  $\exists y (y \le x \land causal(\exists z (y \le z \land Haruko\_ga\_sotsugyo-suru(z)), isshoni\_ryoko-suru(y)))$ 

これに対し、例文(6)においては主節・従属節とも非過去形であり、両者とも"it"の上に新しいスコープを与えられる。従属節の出来事時を z、"ev"のスコープを x とすると、z は x との関係において時制解釈を与えられる((6')の  $x \subseteq z$ )が、主節は過去形ではないので、x はその出来事時ではなく、文脈によって与えられる基準時、すなわち「現在」である。

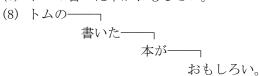
(6')  $\exists y (x \subseteq y \land causal(\exists z (x \subseteq z \land Haruko\_ga\_yasun-de\_iru(z)), kanashii(y)))$ 

#### 4-3 日本語意味解析実験

SCT を用いて日本語文の意味解析を行うシステムのプロトタイプ開発を行った。システムの構成は「3 英文解析実験」で述べたものと類似している。統語解析器としては、日本語の依存/格構造解析を行う KNパーザ (Kurohashi and Nagao 1994) を用いた。同解析器は、下位範疇化フレームを含む豊富な語彙情報を出力し、また出力される依存構造からシステマティックに構成素構造を出力できるという利点がある。しかもきわめて頑強であり、出力される解析結果の精度は 91.1 パーセントとのことである。

例文(7)の意味解析例を以下に示す。KNパーザの出力は(8)のように図示することが出来る。

(7) トムの書いた本がおもしろい。



語彙情報も含めると、(8) はより詳しくは (9) のように示すことが出来る。

(9) -0 1
noun lc fh "トム/とむ"
particle\_case "の"
-1 2
verb lc fh ["ga", "wo"] "書く/かく"
-2 3

```
noun lc fh "本/ほん"
particle_case "が"
-3 0
adjective lc fh [ "ga"] "面白いおもしろい"
```

これは変換規則により、次の SCT 式に書き換えられる。

これが意味評価されることにより、以下の述語論理式が得られる。

(11) exist x y (トム/とむ(x) and 書く/かく(x, y) and 本/ほん(y) and 面白い/おもしろい(y))

現在、京都大学コーパスの解析を通じて、SCT 日本語意味解析システムの拡張を行っている。

## 5 結論

本報告では、形式意味論の手法にもとづく日本語文の意味解析実験について報告した。ダイナミック意味論を文意味解析に応用するスコープ制御理論 (SCT) を用いた英文意味解析について、良好な結果を得ている。同じ手法による日本語文意味解析実験を現在行っており、有望な感触を得ている。伝統的な日本語文法研究の集大成ともいえる南 (1974) の文 4 階層説はそのままで SCT による定式化が行え、このことによって文法・語彙に関する研究の蓄積を生かすことが出来る。

## 【参考文献】

Alastair Butler. 2010a. The Semantics of Grammatical Dependencies. Current Research in the Semantics/Pragmatics Interface, Volume 23, Bingley: Emerald.

Alastair Butler. 2010b. Semantically Restricted Argument Dependencies. Journal of Logic, Language and Information. DOI: 10.1007/s10849-010-9123-8.

Alastair Butler. 2009. Quantification Support with Indeterminate Pronouns. Journal of Cognitive Science, 10: 2, Institute for Cognitive Science, Seoul National University, pp. 113–133.

Alastair Butler. 2007. Scope Control and Grammatical Dependencies. Journal of Logic, Language and Information, Volume 16, Number 3, Springer, The Netherlands, pp. 241–264.

Alastair Butler, Yusuke Miyao, Kei Yoshimoto and Jun'ichi Tsujii. 2010. A Constrained Semantics for Parsed English. 『言語処理学会第 15 回年次大会発表論文集』東京大学, pp. 836-839.

Alastair Butler, Chidori Nakamura and Kei Yoshimoto. 2009. Topic/Subject Coreference in the Hierarchy of Japanese Complex Sentences. In: H. Hattori, T. Kawamura, T. Ide, Makoto Y. and Y. Murakami (eds.) New Frontiers in Artificial Intelligence: JSAI 2008 Conference and Workshops Asahikawa, Japan, June 2008 Revised Selected Papers (Lecture Notes in Artificial Intelligence 5447), pp. 119-132. Springer Verlag.

Jeroen Groenendijk and Martin Stokhof. 1991. Dynamic Predicate Logic. Linguistics and Philosophy 14: 1, 39–100.

Hans Kamp and Uwe Reyle. 1993. From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Dordrecht: Kluwer.

Susumu Kuno. 1973. The Structure of the Japanese Language. Cambridge, Mass.: MIT Press.

Sadao Kurohashi and Makoto Nagao. 1994. KN parser: Japanese dependency/case structure analyzer. In: Proceedings of the Workshop on Sharable Natural Language Resources, pp. 48–55.

Mitchell P. Marcus, Mary Ann Marinkiewicz and Beatrice Santorini. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. Computational Linguistics 19: 2, 313–330.

Yusuke Miyao, Alastair Butler, Kei Yoshimoto and Jun'ichi Tsujii. 投稿中. A Modular Architecture for the Wide-Coverage Translation of Natural Language Texts into Predicate Logic Formulas.

Yusuke Miyao and Jun'ichi Tsujii. 2008. Feature Forest Models for Probabilistic HPSG Parsing. Computational Linguistics 34(1), 35–80.

Kei Yoshimoto. 1998. Tense and Aspect in Japanese and English. Frankfurt am Main: Peter Lang.

Kei Yoshimoto, Chidori Nakamura and Alastair Butler. 2009a. Hierarchy-based Interpretation of Tense in Japanese Complex Sentences. In: Current Issues in Unity and Diversity of Languages: Collection of the Papers Selected from the CIL 18, Held at Korea University in Seoul on July 21-26, 2008, pp. 3417-3435. Seoul: The Linguistic Society of Korea.

Kei Yoshimoto, Chidori Nakamura and Alastair Butler. 2009b. Towards Establishing a Hierarchy in the Japanese Sentence Structure. In: Olivia Kwong (ed.) Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Vol. 2, pp. 875-882, City University of Hong Kong, Hong Kong.

三上章 1970『文法小論集』くろしお出版.

南不二男 1974 『現代日本語の構造』大修館.

吉本啓 2004「日本語複文テンスの構成的意味論:英語との対照の観点から」佐藤滋・堀江薫・中村渉(編) 『対照言語学の新展開』pp. 229-253, ひつじ書房.

吉本啓 1993「日本語の文階層構造と主題・焦点・時制」『言語研究』 103, pp. 141·166, 日本言語学会.

## 〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
Quantification Support with Indeterminate Pronouns	Journal of Cognitive Science, 10:2	2009
Topic/Subject Coreference in the Hierarchy of Japanese Complex Sentences	New Frontiers in Artificial Intelligence	2009
Hierarchy-based Interpretation of Tense in Japanese Complex Sentences	Current Issues in Unity and Diversity of Languages	2009
Towards Establishing a Hierarchy in the Japanese Sentence Structure	Proc. the 23rd Pacific Asia Conference on Language, Information and Computation	2009
Semantically Restricted Argument Dependencies	Journal of Logic, Language and Information	2010
The Semantics of Grammatical Dependencies	Current Research in the Semantics/Pragmatics Interface, Volume 23, Emerald	2010
A Constrained Semantics for Parsed English	言語処理学会第 15 回年次大会 発表論文集	2010
A Modular Architecture for the Wide-Coverage Translation of Natural Language Texts into Predicate Logic Formulas		投稿中