

平成 22 年 6 月 1 日現在

研究種目：若手研究 (B)
研究期間：2008 ～ 2009
課題番号：20720109
研究課題名 (和文) 動的統語論と動的意味論を用いた日本語の逐次的解析の研究
研究課題名 (英文) Research on the Incremental Processing of Japanese Based on Dynamic Syntax and Semantics
研究代表者 小林 昌博 (KOBAYASHI MASAHIRO) 鳥取大学・教育センター・准教授 研究者番号：50361150

研究成果の概要 (和文)：本研究では、動的統語論を用いて日本語の統語構造と意味表示を導出し、モデル理論的に意味を解釈する動的モデルの基盤を構築することを目指している。現象として数量詞や不定詞のスコープの曖昧性をとりあげ、スコーレム化された選択関数を用いて解釈する枠組みを提案した。また、解析をコンピュータ上で行うために、モデルを Prolog を用いて実装する研究を進めた。動的統語論の先行研究では、日本語用パーサーの実装の研究がされていなかった。本研究では、解析状態の推移を駆動する推移規則の適用アルゴリズムを提案した。

研究成果の概要 (英文)：This research project is aimed to establish a basis for a dynamic framework for the interpretation of language, an approach to process sentences in Japanese in an incremental fashion on the basis of Dynamic Syntax and model-theoretic semantics. Skolemized choice functions are applied to the treatment of the scope of quantifiers and indefinites NPs in Japanese in this research. This research also presents the parser implemented in Prolog. Since little research about the implementation of sentence processing in Japanese has been proposed in the Dynamic Syntax literature, this research provides an algorithm for the application of transition rules, which enables the state of parsing to shift to the next step.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	500,000	150,000	650,000
2009 年度	400,000	120,000	520,000
年度			
年度			
年度			
総計	900,000	270,000	1,170,000

研究分野：言語学、形式文法、形式意味論

科研費の分科・細目：言語学・言語学

キーワード：動的統語論、意味論、パーサー、実装、数量詞、不定詞

1. 研究開始当初の背景

我々は文を読む際に、単語（やフレーズ）を逐次的に読んでいって意味を理解している。したがって、人間の文処理過程をモデル化するならば、それは左から右へと逐次的に単語を読み込み、意味を導出していく動的モデルとなろう。本研究では、語順やスコープの作用域などの線形的な問題を対象とし、動的統語論を用いて文を逐次的に解析しながら数量詞や不定詞のスコープを意味表示とともに出力するような枠組みの研究を目指した。本研究で理論的基盤として用いている動的統語論(Kempson et al. 2001, Cann et al. 2005)は、LOFT という二分木上の様相論理に基づく言語理論で、文を逐次的に解析し、意味表示を出力する。まず、動的統語論は比較的新しい言語理論であるため、日本語におけるさまざまな構文を解析する規則を整える必要があった。また、理論の妥当性を科学的に検証するためにも、研究の応用面の可能性を広げるためにも、コンピュータ上にパーサーを実装してシミュレーションする必要性を感じていた。日本語の文処理を動的統語論で実現し、それをコンピュータ上に実装する研究はなされていなかったため、実装にあたって必要となる推移規則を適用するアルゴリズムを研究することも本研究の仕事とした。また、本研究による基礎的な洞察が将来的に e-learning 教材の開発などの外国語教育に応用できれば、教育面にもフィードバックをすることができると考えている。

2. 研究の目的

本研究では、人間の言語能力を記述するモデルとしてふさわしい言語理論とその理論をデザインする際に準拠する適切な論理を研究することが目的であるが、そのファースト・ステップとして言語の線形性に関連する現象に焦点を当てた。具体的には、ヘッド・ファイナル言語である日本語において、人間が左から右へと単語を読み、文を理解する際に、数量詞の遊離現象や数量詞と不定詞などのスコープの曖昧性といった線形的な問題がどのように逐次的解析によって説明されるのかを明らかにすることを目的とした。また、同時にその理論的基盤となる論理、数理的な枠組みを整備することも目指した。動的統語論では、量化詞のスコープはスコープ・ステートメントにより記述される。そのため、逐次的解析により解析木が段階的に大きくなっていく際にスコープ・ステートメントにおいてメタ変数を使ってスコープの曖昧性を表現する方法を提案する必要があった。また、理論的枠組みの精密化には、多量のデータを使った実証的研究方法が必要不可欠であるため、コンピュータ上に本研究の枠組み

のパーサーを実装し、理論の発展にフィードバックすることも目的とした。動的統語論の実装に関しては、日本語を解析するパーサーの研究がされていなかったため、日本語の文を解析するために推移規則を適用するアルゴリズムを研究することとした。

3. 研究の方法

研究の方法としては、まず、動的統語論で日本語の様々な文を解析するために、理論の細部をつめる作業を行った。理論の精密化と並行して、動的統語論を用いたパーサーの実装に着手した。実装は Prolog を用いて行った。システムは、日本語の文の入力に対して逐次解析を行った結果の解析木と論理式の形式をとる意味表示を出力する。動的統語論を使った英語の文を解析するパーサーの研究は、Purver and Otsuka (2003) や Otsuka and Purver (2003) に見られるが、日本語のようなヘッド・ファイナル言語を解析するためには、推移規則が適用されるアルゴリズムが英語などのヘッド・イニシャル言語とは異なるため、推移規則適用のアルゴリズムを研究しシステム開発を行うことにした。

次に、当該のパーサーを使いながら量化詞や不定詞などを含む文の解析とスコープを適切に表示する規則を研究した。同時に複文の従属節の名詞句が省略されている場合の文の曖昧性を説明する枠組みへと拡張を図った。不定詞のスコープに関しては、従属節にある不定詞が主節にある全称量化詞に依存する解釈を示すことが指摘されており、そのスコープを動的統語論のアンダースペシフィケーションを利用して導き出す規則を研究することにした。具体的には、Chierchia (2001) に見られるように、不定詞を選択関数と見直して解釈する。本研究ではスコアレム化された選択関数とする定式化を模索することにした。

4. 研究成果

研究成果は以下の通りである。動的統語論では、ある解析状態から次の解析状態に移る際、単語の語彙情報としての語彙規則が適用されるか、純粋に状態を推移させる推移規則が適用されるかの 2 つの可能性がある。英語などのヘッド・イニシャル言語では、まず推移規則が適用され、語彙規則が適用される状態を準備するのだが、日本語のようなヘッド・ファイナル言語では、単語を読みこむたびに語彙規則が適用され解析状態が推移する。推移規則は単語と単語の読み込みの間に随意に適用されるので、11 の推移規則がアトラダムに適用されると考えるとパーサーとしての効率が悪すぎるし、何より理論の

実装がこのままでは非現実的である。本研究では、推移規則適用のアルゴリズムを提案し効果的な実装を実現した。この成果は Kobayashi (2008) “Incremental Processing and Design of a Parser for Japanese: A Dynamic Approach”に見られるとおりである。さらにこの成果を複文における名詞句省略による曖昧性をも解釈する枠組みへと拡張した。具体的には、従属節の主語が省略されている場合、従属節の主語を主節の主語と解釈することもできるが、その文の外にある名詞句を指すと解釈することもできる。本システムを用いると望まれる数の意味を適切に導出することを Prolog による詞みゆてーションにより示すことができる。

日本語の様々な構文を解析する手段としてのパーサーの準備が整ったので、数量詞や不定詞のスコープ等の線形的な問題、さらに多重主語構文の解釈など理論の構築を進めた。Nakamura, Yoshimoto, Mori and Kobayashi (2008)では、多重主語構文も動的統語論を用いて解釈することができることを示した。Nakamura and Kobayashi (2009)では、不定詞をタイプ e のエプシロン・タームとして解釈し、モデル理論的にはスコーム化された選択関数として意味解釈している。Kobayashi and Nakamura (2009)では、不定詞と量化詞をタイプ e のエプシロン・タームととらえ、動的統語論上でスコープ・ステートメントのメタ変数に関するマージ規則を提案している。Kobayashi and Nakamura (2009)では、従属節内の不定名詞句が一番狭いスコープをとる意味や、従属節よりは広いが主節の全称量化詞よりは狭く全称量化詞に依存するスコープ、さらには一番広いスコープをとる意味も一つのスコープ・ステートメントと2つの completion ruleにより全て表現可能であることを示した。メタ変数がないノードにおける Completion rule は以下のようなになる。計算は上から下の状態に遷移する。

$$\frac{\{\dots\{?Ty(t), \langle \downarrow_0 \rangle Fo(Q), \langle \downarrow_0 \rangle Ty(X), \langle \downarrow_1 \rangle Fo(P), \langle \downarrow_1 \rangle Ty(X \rightarrow Y), \langle \downarrow_0 \rangle SC(z), \langle \downarrow_1 \rangle SC(x), \dots, \diamond\} \dots\}}{\{\dots\{Ty(t), \langle \downarrow_0 \rangle Fo(Q), \langle \downarrow_0 \rangle Ty(X), \langle \downarrow_1 \rangle Fo(P), \langle \downarrow_1 \rangle Ty(X \rightarrow Y), \langle \downarrow_0 \rangle SC(z), \langle \downarrow_1 \rangle SC(x), Fo(P(Q)), SC(z < x), \diamond\} \dots\}}$$

メタ変数があるノードには以下の Completion rule が適用される。

$$\frac{\{\dots\{?Ty(t), \langle \downarrow_0 \rangle Fo(Q), \langle \downarrow_0 \rangle Ty(X), \langle \downarrow_1 \rangle Fo(P), \langle \downarrow_1 \rangle Ty(X \rightarrow Y), \langle \downarrow_0 \rangle SC(z), \langle \downarrow_1 \rangle ?SC(x < U), \diamond\} \dots\}}{\{\dots\{Ty(t), \langle \downarrow_0 \rangle Fo(Q), \langle \downarrow_0 \rangle Ty(X), \langle \downarrow_1 \rangle Fo(P), \langle \downarrow_1 \rangle Ty(X \rightarrow Y), \langle \downarrow_0 \rangle SC(z), \dots\}}$$

$$\langle \downarrow_1 \rangle SC(x < U), Fo(P(Q)), SC(x < z) \ V (SC(z), ?SC(x < U), \diamond) \dots\}$$

メタ変数がない規則では、スコープは基本的に線形的に決定される。つまり、左にある量化詞が右に出る量化詞よりも広いスコープをとることになる。一方、メタ変数のある Completion rule では、線形順序にならって変数がユニファイする可能性と従属節内の不定詞が広いスコープをとる意味も得ることができる。これらの2つの規則を組み込むことにより、モデルは、不定詞の望まれるスコープの全ての組み合わせを導出することができる表現力を持つことができる。またモデル理論的には、量化詞はエプシロン・タームとして定義され、スコーム化された選択関数として解釈される。また、先行する名詞句に解釈が依存する関係名詞なども同じ枠組みで説明できることを示すことができた。

今後は、人工知能の分野で言われる制約が動的統語論の解析にどのようにかかわってくるかを調べる予定である。また、日本語における他の様々な現象も包括的に説明するために理論の精密化を推しすすめる。Prolog を用いたパーサーの開発を理論の発展に対応するためにすすめる予定である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

1. Kobayashi M. and H. Nakamura. “On the Scope Interaction of Japanese Indefinites: An Epsilon Calculus Approach”. *Proceedings of 23rd Pacific Asia Conference on Language, Information and Computation (PACLIC23)*. Volume1. pp.269-277. 2009. 査読有。

2. Nakamura H. and M. Kobayashi. “Japanese Indefinites as Epsilon Terms”. *Proceedings of the Sixth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS2009)*. pp.59-70. 2009. 査読有。

3. Nakamura H., K. Yoshimoto, Y. Mori and M. Kobayashi. “Multiple Subject Constructions in Japanese: A Dynamic Syntax Approach”. *Proceedings of the Fifth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS2008)*. pp.14-29. 2008. 査読有。

4. Kobayashi, Masahiro. “Incremental Processing and Design of a Parser for Japanese: A Dynamic

Approach”. Satoh et al. (eds.) *New Frontiers in Artificial Intelligence (JSAI 2007 Revised Selected Papers)*. pp.174-186. Springer. 2008.
査読有.

〔学会発表〕(計 1 件)

1. Kobayashi M. and H. Nakamura. “On the Scope Interaction of Japanese Indefinites: An Epsilon Calculus Approach”. 23rd Pacific Asia Conference on Language, Information and Computation (PACLIC23). 2009 年 12 月 3 日. City University of Hong Kong.

6. 研究組織

(1) 研究代表者

小林 昌博 (KOBAYASHI MASAHIRO)

鳥取大学・教育センター・准教授

研究者番号 : 50361150