Osaka University Knowledge Archive

Title	Studies on Information-Based Parsers : A Syntax-Driven Discourse Processing
Author(s)	横川, 博一
Citation	大阪大学, 2000, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/42018
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

Osaka University Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

Osaka University

【1】

氏　　名	横川博一（よこかわひろかず）
博士の専攻分野の名称	博　士（言語文化学）
学位記番号	第　15551　号
学位授与年月日	平成12年3月24日
学位授与の要件	学位規則第4条第1項該当 言語文化研究科言語文化学専攻
学位論文名	Studies on Information-Based Parsers:A Syntax-Driven Discourse Processing （情報に基づく文理解モデルに関する研究－統語論駆動型談話処理）
論文審査委員	（主査） 教授　中西　暉 （副査） 教授　成田　一　　助教授　三藤　博　　助教授　宮本　陽一

論文内容の要旨

本論文は、人間の言語理解メカニズムの普遍性を仮定し、主辞駆動型文法理論（Head-driven Phrase Structure Grammar;Pollard & Sag 1994, Gunji, 1987）に基づき、心理言語学的文理解モデルとして Information-Based Parser を提案したものである。

第1章では、本論文の導入として、これまでの心理言語学における文理解研究を振り返ることによって、人間の言語理解の基本的性質について概観し、本論文の目的について述べた。

人間の言語理解に関する研究は、これまで長い間英語に限られていたが、近年 cross-linguistic な観点から英語とは極めて対照的な構造特性をもつ日本語を扱った研究が盛んになってきている。しかし、心理言語学的実験や考察が数多くなされてきているが、これまでに提案されたモデルやストラテジーの多くは、その都度実験結果に基づいて設定されたアドホックなものが多く見受けられる。人間の言語理解モデルが普遍的なものであると仮定するならば、文理解の諸現象を統一的に説明できるメカニズムが存在するはずであり、何らかの文法理論によって形式化されたモデルを構築する必要がある。

本論文では、制約に基づく文法理論の1つである「主辞駆動型文法理論」を援用して人間の言語理解モデルの構築を試みた。

第2章では、人間の言語理解モデルとして、Information-Based Parser を提案した。まず、2．1節では、本モデルの基本設計として、人間の言語情報処理には制約に基づく設計（constraint-based architecture）が適合していることをみた。2．2節では、制約に基づく文法理論の基本的性質について触れ、文法理論が備えている素性構造、レキシコン、単一化について解説した。しかし、文法理論それ自体は静的な性質を持ったものであり、これを人間の動的な処理に適用するためには、何らかの運用ストラテジーが必要である。本モデルでは、*Lexicon, Constraints, Evaluation* の3つのモジュールを想定した。また、ある種の文（たとえばガーデンパス文）は局所的曖昧性により処理に支障をきたすことがあるが、こうした現象を説明するために、統語・意味情報変更コスト（SYNSEM Information Alternation)、意味情報変更コスト（SEM Information Alternation)、スラッシュ素性設定コスト（(P) SLASH Feature Setting）というコスト素性を設定し、コスト計算を可能にした。

第3章および第4章では、本モデルを文レベルおよび談話レベルの文処理に適用し、モデルの妥当性を検証した。

第3章では、文レベルの文理解に焦点を当て、主として日本語および英語のガーデンパス文を例に取り、モデルを

適用し、処理の局所性（局所的処理の原則)、SYNSEM 情報変更コスト、SEM 情報変更コストなどの性質が心理的に妥当性をもつことを示した。

まず、Pritchett（1992）および Mazuka らをもとに、日本語および英語のガーデンパス文の分類をおこない、これまでに提案されている文理解モデルの基本的性質を概観し、それぞれの問題点を指摘した。

本モデルを日本語および英語のガーデンパス文に適用した結果、ガーデンパス効果、SYNSEM 情報を変更した結果生じるものであり、SEM 情報のみの変更はそれほどコストがかからないことが明らかになった。

また、GP 効果を回避する要因を探ることによって、本モデルの妥当性をさらに示した。レキシコンに含まれていると考えられる SEM 情報の性質について考察し、単に Agent, Theme などの意味役割だけでなく、より直接的に意味役割関係が計算されること、語用論的情報もオンラインで利用されることが心理実験によって明らかになった。次に、日本語では語順の変更（かき混ぜ）によって GP 効果が弱まる例を取り上げ、本モデルではそれを形式的に自然に説明できることを示した。

第 4 章では、談話レベルの文処理に焦点をあて、文脈情報と動詞の処理の相互作用について本モデルの適用可能性を探った。

まず、動詞の処理の基本的性質について心理実験を通して考察した。とりわけ多くの情報を持ち、文の構造を決定するのに大きな役割を果たす動詞は、複数のレキシコンの可能性をもつことが多い。先行文脈情報、意味／語用論的情報によって、特定のレキシコンが優先的に活性化されるメカニズムを明らかにした。

次に、文脈情報による意味的補完現象を通して、動詞の情報処理について明らかにした。その結果に基づき、(P) SLASH feature 設定コストの性質について論議した。(P) SLASH 素性設定コストの性質を明らかにする前に、まず動詞の情報処理の基本的性質について明らかにした。動詞の語彙情報をどう捉えるかについて 2 つの見方を導入し、動詞の情報量に比例してコストがかかることを心理実験を通して示した。また、動詞が要求する必須項とそうではない随意項の違いが心理的に実在することを示した。

さらに、動詞が現れた時点で項が欠落している場合、ただちに (P) SLASH 素性が導入されるわけではないことを示した。つまり、先行する文脈情報により特定のレキシコンが活性化され、(P) SLASH 素性導入が回避されるというメカニズムが存在すると考えることが妥当であることを示した。

先行文脈の情報により、複数の候補の中から特定のレキシコンが活性化される。動詞が要求する項が欠落している場合、(P) SLASH 素性が導入されるが、文脈情報により意味的に補完される場合は、(P) SLASH 素性の導入を回避することができることが心理実験により分かった。また、Yamashita（1997）の実験結果を援用して、かき混ぜによって文処理の難易度は変わらないことから、(P) SLASH 素性が逐次導入されるわけではないことを見た。このことは、ある固定した語順が想定されて文処理が行われるような制約が働いているわけではないことも示している。

第 5 章では、日本語の照応関係理解に関する心理実験を通して、主題（topic）の果たす役割について考察した。日本語の「は」と「が」の違いが、統語的曖昧性文における代名詞の解釈に影響を及ぼすかどうか心理実験を行った。その結果、これまでに提案されているストラテジー（主語割当方略、並行機能方略など）に優先して、主題となっている名詞句の解釈が代名詞に付与されることが分かった。これは、「は」と「が」のもつ語彙的情報の違いがオンラインで利用されていることを示している。つまり、代名詞の解釈プロセスにおいて、主題となっている要素が文脈情報として、レキシコンの素性構造の中に組み込まれることを示しており、本モデルでは素性構造の単一化によって形式的に扱うことができることを示した。

第 6 章では、本論文の要点を総括した後、今後の課題として、以下の点を指摘した。(1)言語処理の制約に基づく基本設計のさまざまなモジュールがどう関連しあっているか。例えば、統語処理に音韻情報がどう影響するかなどについて、広範な考察が必要である。(2)日本語の場合、主要部の助詞が欠落することが多いが、こうした文を人間はどのように理解しているのか、心理実験に基づく考察が必要である。

論文審査の結果の要旨

本論文は、人間の言語理解モデルは言語によらず普遍的であるとの考え方のもとに、文理解の諸現象を統一的に説明できるメカニズムについて主として心理言語学的立場から考察し、そのようなモデルとして制約に基づく文法理論の1つである主辞駆動句構造文法（HPSG）を援用した文理解モデルであるInformation-Based Parserを提案し、本モデルを複雑な構造をもつ文レベルの文処理、文脈情報が利用できる談話レベルの文処理、並びに照応関係理解に適用した場合の本モデルの妥当性を心理言語学的実験を通じて種々の角度から検証したものである。

論文は全6章からなり、第1章ではこれまでの心理言語学における文理解研究を概観し、人間の言語理解の基本的性質および文理解モデルの具備すべき条件について論じ、本論文の立場を明らかにしている。第2章では、制約関係を基本的概念として用い、素性の単一化により構造記述を行うHPSGに代表される制約に基づく文法理論は、その基本的性質としてモジュール間の柔軟な関係と情報の流れの無方向性・同時性を合わせもつなど、人間の文理解モデルとして適切な枠組みであることを指摘し、制約に基づく文法理論を援用した言語理解モデルであるInformation-Based Parserを提案している。本モデルは素性構造、語彙目録、コスト計算等の評価を行う3つのモジュールからなる静的なモデルである。左から右への漸進的処理、柔軟な構成素形成、局所的処理性という3つの運用方略を導入することにより人間の動的な処理に適合させ、統語・意味（SYNSEM）情報変更コスト、意味（SEM）情報変更コスト、スラッシュ（SLASH）素性設定コストを導入することにより文理解の難易度が評価できるようにしている。

第3章では、文レベルの文理解に焦点を当て、主として英語と日本語のガーデンパス文の理解に本モデルを適用した場合の妥当性について心理言語学的実験を交えながら検証している。その結果、(1)ガーデンパス化はSYNSEM情報を変更することにより生じ、その強弱は変更回数によって説明できること、(2)語順の変更、単語の置換等によりガーデンパス化が弱められ、その現象が本モデルで自然に説明できること、(3)局所的処理性、変更コスト等が心理的に妥当性をもつことなどを明らかにし、本モデルの有効性を示している。

第4章では、談話レベルの文処理に焦点を当て、本モデルを適用した場合に、文脈情報が動詞の処理にどのように係わるかについて検討している。まず、文処理において重要な役割を果たす動詞の処理の基本的性質について心理言語学的実験を通して考察し、処理の複雑さは動詞がとる項の数ではなく項構造の可能性のパターン数に依存すること、パターン数が複数個存在する場合、先行文脈情報、意味／語用論的情報によって特定の語彙項目が優先的に活性化されるメカニズムを明らかにしている。さらに、文脈情報による意味的補完現象について考察し、動詞が現れた時点で項が欠落している場合には一般にSLASH素性が導入されるが、文脈情報により意味的に補完が可能な場合にはSLASH素性の導入を回避できることを示している。

第5章では、日本語の照応関係理解に本モデルを適用した場合の妥当性について検証するため、主として心理言語学的実験を通して主題（Topic）の果たす役割について考察している。統語的曖昧文における代名詞の解釈に日本語の「は」と「が」の違いが及ぼす影響について考察し、これまで提案されている主語割当方略、並行機能方略に優先して、主題となっている名詞句の解釈が代名詞に付与される主題割当方略が用いられることを示し、「は」と「が」のもつ語彙的情報の違いがオンラインで利用されていることを実証している。このことより、代名詞の解釈プロセスにおいて、主題となっている要素が文脈情報として語彙の素性構造の中に組み込まれること、すなわち、本モデルで素性構造の単一化によって形式的に扱い得ることを示せたことになり、本モデルの妥当性を実証している。第6章は結論である。

以上、本論文は、文理解の諸現象を統一的に説明できるメカニズムとして、制約に基づく文法理論であるHPSGを援用した文理解モデルを提案し、本モデルを種々のレベルの文理解に適用した場合の妥当性について検証を試みたものである。人間の言語理解に関する研究は、英語から始まり、日本語を含む他の言語へと移行し、どの言語にも共通する普遍的な言語理解モデルの構築が要求され、それに応じて心理言語学的実験や考察が数多くなされているが、これまでに提案されたモデルや方略はいずれも上記の条件を完全には満たしていない。この分野のこのような状況を考えた場合、制約に基づく文法理論を援用した文理解モデルを構築した点、種々の言語現象に本モデルを用いて統一

的な説明を一応与え得た点、およびコストなどの新しい概念を導入し、制約に基づく文法理論の通常の枠組みを拡大した点は高く評価できる。

本論文では、言語処理の制約に基づく基本設計のモジュール間がどのように関連しあっているかについてはほとんど述べられていない。また、考察の対象を英語と日本語に限っているが、理論の枠組みとして普遍的なものを仮定していることから、本モデルがどれだけの普遍性をもっているかなどの問題が残されているが、これらは今後の課題として追求していくべき問題である。

以上の諸点から、本論文はこの分野の研究として高い水準に達しており、博士（言語文化学）の学位論文として十分に価値あるものと認められる。