様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月 １５ 日現在

機関番号：３４４０３
研究種目：基盤研究(C)（一般）
研究期間：2016～2018
課題番号：１６Ｋ００３１３

研究課題名（和文）省略解析のための漸進的な意味表示の構築に関する研究

研究課題名（英文）An incrementally constructed semantic representation for empty category resolution

研究代表者
大谷　朗（OTANI, Akira）

大阪学院大学・情報学部・教授

研究者番号：５０２８３８１７

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：本研究では動的統語論（Dynamic Syntax: DS）に基づく日本語の形式文法を設計し，DSの派生を利用したコーパスからの意味表示の(半)自動抽出方法を検討した．一般に，ゼロ代名詞や省略などといった言語学的に複雑な現象によって文中の単語や句の省略が多くなると，解析器の効率は低下する．そこで，本研究では，DS文法の枠組みに基づいて，日本語の右方転位と認識動詞構文の統語的・意味的特性といった言語学的な問題を分析し，省略解析に利用できる漸進的な文解析の方略を提案した．

研究成果の学術的意義や社会的意義
くだけた日常会話だけでなく形式を重んじた文書においても，単語や句といった文の構成要素はしばしば省略される．また，文意の解釈では重要であるにも関わらず，省略されているものとして認識されていない要素を，文が含むこともある．本研究は，そうした要素を含むさまざまな文を検討することで，省略が生じる文の性質を言語学的に明らかにするとともに，情報化社会の基盤であるコンピュータを用いた言語処理においても緊要の課題にあげられる省略解析にも応用できるような文法を形式化した．

研究成果の概要（英文）：This research project led to the design of a Japanese formal grammar based on Dynamic Syntax (DS) and considered a way of (semi-)automatically extracting semantic representations from corpora by using DS derivations.　In general, a parsing performance deteriorates as a sentence loses words and phrases because of linguistically complex phenomena such as zero pronoun and ellipsis.　Under the DS framework, we analysed a number of linguistic matters including syntactic and semantic properties of right dislocation and epistemic verb construction, and propose an incremental parsing strategy which is available for empty category resolution.

研究分野：理論言語学・計算言語学

キーワード：省略　empty category　動的統語論　Dynamic Syntax　右方転位　後置文　認識動詞構文　raising to object

１．研究開始当初の背景

(1) 日本語では主語や目的語などの省略が頻繁に起こる．省略の補完は，日本語と他言語との間の機械翻訳や，日本語テキストからのマイニングなどでも重要な技術であり，多くの研究がなされてきた．それらは人手で作成した規則ベースの手法と，コーパスを用いた統計ベースの手法とに大別される．

① 前者は，照応詞と先行詞候補との間の統語的および意味的な制約・選好に着目した規則により，高精度な解析に成功しているが，対象テキストのドメインが限定されている．
② 他方，後者の機械学習や確率モデルを用いた手法は，照応詞やゼロ代名詞と先行詞候補との間の文数などを距離尺度として素性やパラメータの一つにすることで，解析精度の向上を図っている．しかし，いずれの手法も，文・文章の構造は考慮せず，また，手法の一部に根拠の薄いヒューリスティクスを残している点に問題がある．

(2) 日本語以外の言語でも，照応・省略解析の研究は盛んに行われており，コーパスベースの手法が一定の成功を収めている．
　しかし，そうした研究もまた，距離尺度に関しては上記同様の問題がある．照応詞と先行詞候補との間の位置関係を構造的に捉えるべく，構文木中を横断しながら先行詞を探索する手法も提案されてはいるが，探索の優先順序の制御が完全ではなく，議論の余地が残されている．

(3) こうした言語処理手法の開発とは独立に,「狭義」の省略は言語理論研究の主要課題の一つとして，長年分析されてきた．
　中でも Dynamic Syntax (Kempson et al. 2001, Cann et al. 2005, Kempson et al. 2011) に基づく Ellipsis の研究 (Cann et al. 2007) では，理論が規定する漸進的な文の解析・生成と省略の解消があいまって，話し手と聞き手が協調しながら対話を進行させていく compound contribution といった「広義」の省略の研究 (Purver et al. 2014) へと発展している．

２．研究の目的

(1) 本研究は，漸進的に文の意味表示を構築する Dynamic Syntax (DS) に基づいて，日本語の省略解析のための DS 文法を形式化し，コーパスの用例に対して文法解析を行うことで DS 文法の実用性を評価することを目的とする．そのため，以下の段階的な目標を遂行する．

① 言語学および言語処理の先行研究をサーベイし，省略現象の類型と各現象間の理論的関係,また，省略解析におけるそのような現象の重要度および緊要の課題を調査する．
② ①で選択した省略現象に対して， 必要に応じて制約の調整，追加などを行った日本語 DS 文法に基づく理論的分析を提示する．
③ ①で選択した省略現象に応じてコーパスを選定，対応する用例を抽出し，②で提示した分析をそのような例に対して適用することで，本研究の提案する日本語 DS 文法の有効性を評価する．

(2) 言語の多様性を網羅する規則を作ることが難しく，また，新しいテキストを対象とする毎に膨大な作成コストもかかってしまうことから，省略解析に限らず，今日の言語処理研究ではコーパスを用いた統計ベースの手法が主流となっている．そして，そうした研究は一定の成果をあげているものの，対象データを限定するといった条件次第では，統計ベースの解析精度を規則ベースのそれが上回ることは上述の通りである．

(3) さらに付け加えると，統計ベースの手法はコーパスが含むデータの質・量に依存するが,このことは，こうした手法で対象としてきた省略現象が，コーパス内で定義された省略関係の定義に強く依存していることに他ならず，その結果，日本語の研究は「狭義」の省略の扱いに留まっている．

(4) 本研究は言語学の知見をもとに省略現象の類型を整理し，各現象の処理の重要度に応じて省略の類型を DS の汎用的なメカニズムに基づいて順次体系的に捉えていくことを目指す．

(5) コーパスに依存しない規則ベースの手法の確立を策する本研究は，以下の特徴を持つ．

① 省略関係の定義に束縛されないことから，日本語における「広義」の省略解析の基礎研究を推進するという意義を持つ．
② DS に基づく理論分析により，さまざまな省略関係の特徴を明示することで，統計ベースの手法にとっても素性やパラメータを設定する際に参考となる条件などを提示する．

３．研究の方法

(1) 上記の背景および目的をもとに,本研究は DS に基づく日本語の省略現象の理論的説明を明示する．そのため DS の制約を調整し，また，必要に応じて新規の制約を導入することで，実用

日本語 DS 文法を形式化し，コーパスから抽出した実例を解析することで文法を評価する．

(2) この目標を遂行するため，研究期間内には，以下の理論的考察や実験を段階的に行う．

① 言語学の先行研究をサーベイし，省略現象の類型と各現象間の理論的関係を調査する．
② 言語処理の先行研究をサーベイし，①の現象の省略解析における重要度を調査する．
※ ②における緊要な課題から順に，③～⑤を繰り返すことで，実用 DS 文法の射程を拡大しつつ，省略現象の体系的な説明理論を明示する．
③ ②で選択した省略現象に対して，①の調査結果を踏まえつつ，必要に応じて制約の調整，追加などを行った日本語 DS 文法に基づく分析を提示する．
④ ②で選択した省略現象に応じてコーパスを選定し，対応する実例を抽出する．
※ ⑤の検証に十分な実例と必要な情報を含むコーパスが入手できない場合は，手作業で収集した文に対する詳細な事例分析を行い，④と⑤は区別しない．
⑤ ③で提示した分析を，④で抽出した実例に対して適用し，文法の評価と修正を行う．

４．研究成果

(1) 本研究は，漸進的に文の意味表示を構築する Dynamic Syntax の枠組に基づいて，日本語の省略解析のための DS 文法を形式化し，コーパスの用例を解析することで文法の実用性の評価を行った．

先行研究のサーベイから開始し，問題点の洗い出しおよびそれに対する代案，新規の分析を経た新たな理論的形式化を提示，用例の解析を行う流れで計画を遂行し，課題の達成のため，計画年度内は以下の①～③の下位課題を段階的に実施してきた．

① 省略現象の類型の調査と各現象の解析課題としての重要度の考察
② Dynamic Syntax に基づく省略現象の分析と日本語 DS 文法の整備
③ コーパスからの実例の抽出とその解析による DS 文法の評価

(2) 省略解析で必要となる言語情報の形式化は，DS に基づく日本語文法の全体的な骨子の構築(以下の(3)が該当)と，そのような骨子に基づく一言語現象としての省略の分析（以下の(4)が該当)とに大別される．本研究は言語学研究として省略現象の類型および現象間の理論的関係を，また，言語処理研究として省略解析におけるそうした現象の重要度および緊要の課題に注意を払い，先行研究の知見に基づく新たな理論的分析，提案を中心に取り組んだ．

(3) 言語学分野における理論的分析のサーベイにおいて，これまであまり検討されてこなかったものの，コーパス等には現れるような日本語の語順，特に後置文と Wh 語の語順制約について考察した．以下では学術論文として採択されたものおよびその成果の概要をあげる．

① The Word Order Flexibility in Japanese Novels: A Dynamic Syntax Perspective
日本語の基本語順では動詞が文末に現れるが，小説，とりわけ会話部分では後置文という談話現象として，さまざまな要素が動詞よりも後に配置される．この論文ではそうした後置要素の統語タイプを分析し，日本語の語順制約が従来考えられていたものよりも緩いことを Dynamic Syntax の枠組に基づいて説明した．
② The Wh-Licensing in Japanese Right Dislocations: An Incremental Grammar View
日本語の基本語順において動詞より前(左)に現れる要素の中には，後(右)に現れる要素もある．本論文ではそのような要素の一つ，とりわけ Wh 語に関してデータ収集を行った．そして，そのように逸脱した語順の文を後置文とよび，左から右への解析を反映した文法モデルに基づいて，人のリアルタイムな漸進的文処理として後置文の解析を示した．

(4) 言語の理論的分析および言語処理が扱う現象のサーベイにおいて，特に後者では検討されてこなかった構文，目的語への繰り上げ(raising to object: RTO)構文，認識動詞構文について本研究は考察した．以下では学術論文として採択されたものおよびその成果の概要をあげる．

③ Raising to Object in Japanese: An HPSG Analysis
RTO 構文は対角目的語の統語的，意味的性質に関して興味深い問題を提起する構文として研究されてきた．この論文では，構文特有の語順の説明には二種類の主文述語，繰り上げとコントロールが必要であり，また，目的語は補文述語と叙述の関係にあることを論じ，主辞駆動句構造文法の枠組の下でそうした文法制約の厳密な形式化を示した．
④ Note on Japanese Epistemic Verb Constructions: A Surface-Compositional Analysis
認識動詞構文は，変形に基づく統語理論において盛んに研究されてきた．この論文では，英語を対象に分析されてきたこの構文の統語的性質に注意を払いつつ，組合せ範疇文法の枠組に基づいて日本語の意味的性質を検討した．この構文は表層的な主文目的語が補文の意味的な項を統御していると捉えられ，その点で日英語の間に違いがないことを示した．

＜引用文献＞
① Kempson, Ruth, Wilfried Meyer-Viol, and Dov Gabbay (2001) *Dynamic Syntax: The Flow of Language Understanding*, Blackwell.
② Cann, Ronnie, Ruth Kempson, and Lutz Marten (2005) *The Dynamics of Language: An Introduction*, Elsevier
③ Cann, Ronnie, Ruth Kempson, and Matthew Purver (2007) "Context and Well-formedness: the Dynamics of Ellipsis," *Research in Language and Computation*, Vol.5, Issue 3, pp.333-358.
④ Kempson, Ruth, Eleni Gregorimichelaki, and Christine Howes (eds.) (2011) *The Dynamics of Lexical Interface*, CSLI Publications.
⑤ Purver, Matthew, Julian Hough, and Eleni Gregoromichelaki (2014) "Dialogue and Compound Contributions," in Amanda Sten and Srinivas Bangalore (eds.), *Natural Language Generation in Interactive Systems*, pp.63-92, Cambridge University.

５．主な発表論文等

〔雑誌論文〕（計２件）
① Ohtani, Akira (2017) "Raising to Object in Japanese: An HPSG Analysis," *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation*, pp.72-80, 査読有.
https://wwww.aclweb.org/anthology/Y17-1013/
② Seraku, Tohru and Akira Ohtani (2016) "Wh-Licensing in Japanese Right Dislocations: An Incremental Grammar View," *Empirical Issues in Syntax and Semantics*, Vol.11, pp.199-224, 査読有.
http://www.cssp.cnrs.fr/eiss11/

〔学会発表〕（計１件）
① Ohtani, Akira (2017) "Raising to Object in Japanese: An HPSG Analysis," The 31st Pacific Asia Conference on Language, Information and Computation (PACLIC 31), University of the Philippines Cebu, Philippines, November 16-18, 2017.

〔図書〕（計２件）
① Ohtani, Akira and Mark Steedman (to appear) "Note on Japanese Epistemic Verb Constructions: A Surface-Compositional Analysis," in Jong-Bok Kim and Chu-Ren Huang (eds.), *Studies in East Asian Linguistics*, Springer.
② Seraku, Tohru and Akira Ohtani (2016) Chapter 8: "The Word Order Flexibility in Japanese Novels: A Dynamic Syntax Perspective," in Takashi Ogata and Taisuke Akimoto (eds.), *Computational and Cognitive Approaches to Narratology*, IGI Global, pp.213-244 (420p)

６．研究組織

(1)研究分担者
研究分担者氏名：
ローマ字氏名：
所属研究機関名：
部局名：
職名：教授
研究者番号（8 桁）：

(2)研究協力者
研究協力者氏名：
ローマ字氏名：

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。