

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 6 月 8 日現在

機関番号：33908

研究種目：基盤研究(C)

研究期間：2009～2011

課題番号：21520417

研究課題名（和文） 談話構造に基づく談話理解研究

研究課題名（英文） Research on Discourse Understanding based on Discourse Structure

研究代表者

白井 英俊 (SHIRAI HIDETOSHI)

中京大学・情報理工学部・教授

研究者番号：10134462

研究成果の概要（和文）：

日本語の談話を主たる題材として、談話理解、特に対話における情報のやりとりに関する理論を構築することと、計算機と人間との対話のモデルの構築を目的とした。理論的な枠組はSDRT理論(分節談話表示理論)のアイデアに基づきながら、言語学的な理論研究と、計算言語学的な応用研究の両面からアプローチした。日本語に特化してSDRT理論の提唱する談話関係を拡充し、談話構造の記述の可能性を明らかにした。ツイッターのような短い談話には特に有効であることが示唆された。

研究成果の概要（英文）：

This research was intended to develop a theory of discourse understanding, especially how the information is flowed in dialogue, and to build a computational model of dialogue between computer and human. Adopting SDRT (Segmented Discourse Representation Theory) as its theoretical framework, we took the theoretical research perspective from linguistics and the application perspective from computational linguistics. As the main target was Japanese discourse, we had to extend the discourse relations to include various aspects of Japanese discourse structure, and found that it is effective to analyze and explain the relationship among the information described in the discourse, especially short one such as tweets.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,300,000	390,000	1,690,000
2010年度	800,000	240,000	1,040,000
2011年度	1,100,000	330,000	1,430,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：人文学

科研費の分科・細目：言語学・言語学

キーワード：意味論、談話構造、談話理解

1. 研究開始当初の背景

本研究は自然言語研究において、現在主流となっている「浅い処理」に対する二つの意味でのアンチテーゼであった。実際、自然言

語処理においては、論理や「意味」からのアプローチは過去のものになった感がある。しかし、「浅い処理」、主として統計的言語処理においては、質の高いコーパスを必要とする。

対象を文とするコーパスは存在するが、文章や対話に対しては、質の高いコーパスは皆無に等しい。これは文章構造に対する理論がまだ十分成熟していないことが主たる原因である。文章構造を抽出することは、語義の曖昧さ解消、照応解消、情報の新旧の弁別、主題共有による文章の分節化などに役立ち、応用としては文章要約や情報抽出などに役立つ。しかし文章構造を抽出するには談話関係（修辞関係）を具体的に設定する必要がある。SDRT（分節談話表示）理論では、談話の内容と形式に基づいた談話関係の定義が行われており、文章や対話に適用され、コーパスの作成も行われている。これらは十分参考になるとはいえ、たとえば「手がかり語」においては言語依存であり、また「手がかり語」自体はひとつの談話関係ではなく一般に複数の談話関係を示唆するものである。したがって、日本語に特化した研究が必要と考えた。

Lascarides らの研究からは、文章構造に対して従来の自然言語処理が行っているような機械学習が適用できないという知見が得られていた。つまり、文章構造は表層的な手がかりだけから決定することは困難であり、そこには語義や構文、文章が担う情報と相互依存した関係があることが明らかとなってきた。このことは従来表層的な手がかりによってのみ行われてきた現在の自然言語処理の研究に対して問題を提示するものと考えた。また一方で、文章構造が語義や構文などに依存するということから、辞書に書くべき情報（特にいわゆる辞書的知識と世界知識との境界）の問題に対する知見や文から得られる情報の表現に対する知見が得られることが期待された。

2. 研究の目的

本研究の主目的は以下の二つであった：

- (1) 談話理解、特に対話における情報のやりとりに関する理論を構築すること、
- (2) 談話、特に計算機と人間との対話のモデルを構築することにより、より深い談話の理解に関する知見を深めること

いいかえれば、本研究は、談話の理解に関して、言語学および計算言語学的な観点から研究を行い、談話の理解の過程を理論化することを主目的とし、その理論の実証のためのモデルを構築することを副次的な目的とした。

3. 研究の方法

日本語の談話を主たる題材として、理論的な枠組はSDRT理論（分節談話表示理論）のアイデアに基づきながら、談話の理解に焦点をあて、言語学的な理論研究と、計算言語学的な応用研究の両面からアプローチするものである。

4. 研究成果

言語におけるコミュニケーションは、普通一文によってはなされず、意味的に関係する複数の文の集まりによって行われる。この文の集まりを「談話」と呼ぶ。談話が単なる文の集まりと異なる特徴は、意味的につながりがあるということである。談話の解釈は、語義や文構造の曖昧さ解消、代名詞の専攻し決定など、自然言語研究で取り込まれる問題の多くを含むテーマである。談話の解釈では個々の文の意味解釈に加えて、談話を構成する文が記述する事象の間の関係などが加わり一見、個々の文の解釈よりも難しさがあがるように見える。しかしながら、談話に意味的なまとまりがあることから、実際には個々の文を独立に解釈するよりも容易になるのであり、まさに人間はそれを利用して情報のやり取りを行なっている。

談話の解釈に対する論理的アプローチとして、談話表示理論（DRT, Discourse Representation Theory）が提案されていた。DRTでは文の論理的な表示である談話表示構造を用いて、アクセス可能性を規定することである種的前提や照応現象の解明に貢献した。しかしながら談話を構成する文や発話の関係（談話関係）には踏み込んでいなかった。本研究でベースにするSDRT（分節談話表示理論）は、DRTを次の点で拡張したものである：不完全記述の導入、談話関係を意味論に組み込み談話の構造化を導入、談話関係の推論により談話更新を定義、など。この理論は主として英語の談話において発展、検証されているが、日本語の談話への適用可能性については十分検討されていなかった。

本研究では収集した解説文（教科書など）、新聞記事（特に社説）、物語（教科書など）、および対話（学生間や母子間の会話）を対象として、談話関係を付与したコーパスを作成した。これにより談話における情報の流れを捉え、それによって談話理解の過程を明確にする道が開かれた。さらに計算機上でのモデル構築や、省略および照応の解消のための精緻なメカニズムの構築に寄与することが期待された。これは主として日本語の談話に対して行われたが、その過程で、本来は英語の談話に対して提案されていた談話関係の追加や修正、精密化が必要となった。そのため、修辞構造理論（RST, Rhetorical Structure Theory）なども参考にしながら、コーパスに基づいて談話関係の修正・追加を行った。その主なものは以下のとおりである：

- (1) 記述の関係
- (2) 人の行動の原因や根拠
- (3) 対話イベント

記述の関係とは、文内の要素の緩やかな結びつきを捉えるためのもので、時間的な関係を含まない、結合する文に対し従属空間を作

る、デフォルトとして記述対象のものに対し照応が存在するという特徴がある。人の行動の原因や根拠とは、「原因結果」関係ほど物理的な因果関係はないが、それにより人の行動の原因や根拠となりうる事象と人の行動記述との関係を捉えるものである。対話イベントとは、物語文章によく出てくる、登場人物同士の対話において基本的な性質を推論するための関係であり、対話に現れる発話の状況は同一で、発話間に時間の推移があり、それぞれの発話者が交代するという特徴がある。

このコーパスの作成過程および分析において、主としてキューフレーズによって合図される話者の主張と根拠の間の談話関係と、因果関係のような事象の内容によって成り立つ談話関係との間に質的な差があり、これらの扱いが談話理解における主要な問題となることが認識された。この観点から談話関係の分類を試みたものが以下である（主なものだけ）。それぞれに結びつきの強さの違う関係がいくつかあり、「強い」関係ほど意味的な結びつきが強く解釈され、関係する要素の省略が可能になるなどの副次効果がある：

- (1) 時制が関係するもの
 - (a) 原因・理由－結果。この弱バージョンが「可能化」
 - (b) 目的－行為。類似な関係として、手段－目的、条件－行為、判断・理由－行為、行為－理由・判断がある
 - (c) 状況：時間経過・主体同一。この弱バージョンとして「状況同一」、さらにこの弱バージョンとして「時間経過」と「時空同一」がある。
- (2) 時制が関係しないもの
 - (a) 根拠－帰結
 - (b) 帰結・判断・主張－根拠。この弱バージョンとして「付加」がある
 - (c) 期待破り（これは、キューフレーズが必要）。この弱バージョンとして「適用外・例外規定」がある
 - (d) まとめ、言い換え／同一事象
 - (e) 詳細化：例示化
 - (f) 詳細化：事象。この弱バージョンとして、「事物詳細化」と「行為詳細化」がある
 - (g) 一般化。この弱バージョンとして「相似」と「選択」（キューフレーズが必要）がある
 - (h) 記述：主題同一。この弱バージョンとして「対象同一」がある。さらにこの弱バージョンとして「主題化」「状況同一」がある。
- (3) 構造によるもの
 - (a) 構造並列（同義的な意味関係を示唆）
 - (b) 構造並列－語義対照（対照的な意

味関係を示唆)

- (c) 対話における関係：疑問、命令、QAP, SARG, Divergence など

このように談話関係を分類し相互の関係を解明することで、談話におけるキューフレーズの役割や、省略などの談話現象の生起要因などがより詳しく特定できるようになった。例えば、逆接の接続詞とされる「が」は、時制が関係しない場合、「記述：主題同一」、すなわち同じ状況についての主題が同一である記述関係と「期待破り」の関係を示唆する。そして、記述関係が第一義であるが、この両方が成り立つならばより強い意味関係となる。そのためには、「が」によって接続される二つの事象（発話内容）が「期待破り」の関係を満たす必要がある（そしてそのような解釈が可能ならば、それが認定される）、という妥当な分析が可能となった。

もっとも談話関係による分析での限界もみえてきた。コーパスの分析では、いわゆる段落をはさんだ文の関係はこれでは捉えられないものが多かった。もっとも最近の文章では段落がかなり恣意的に設定されており、段落の認定そのものも問題であった。

またこのことは逆に、一つの段落内の文や発話の関係をかなりよく捉えられることを含意する。近年ではツイッターのような短くリアルタイムでやりとりされるメッセージが注目されている。そこでの文章は基本的に一つの段落とみなせる。そこで、どのような談話構造が現れるかを、データを収集しつつ分析を行った。そして、すでに分析済みの新聞記事や日常会話などの談話構造との比較を行った。その結果、談話関係ごとに見られる特徴の分析（特に省略現象に注目して）、メディアによる違いの分析（特に修辭関係の出現に注目して）、談話関係を手がかりに談話構造の抽出を自動化するという課題が得られた。さらに文献「論理トレーニング」でも短い文章（一つの段落で収まる内容）がとりあげられている。最終年度では検証材料として「論理トレーニング」の題材をSDRTに基づく分析を行い、それと著者による分析とを比較し、両者の構造の一致度が高いこと、およびSDRTによる分析結果の方が、談話構造の特徴がよりよく捉えられることを確かめた。

さらに談話関係を手がかりに談話構造の抽出を自動化する課題に取り組んだが、これについて分析が中途であり、自動化までには至らなかった。

一方で、談話構造を時制とのかかわりから分析を行なっている。具体的には、英語の大規模コーパス（Corpus of Contemporary American English; COCA）をデータとして利用し、構文パターンと談話構造との相互作用を研究対象にした。その結果、語彙レベルで

は不自然な事例であると想定されたとしても、時制や様相などに関連するより複合的な上位レベルでは実際にライセンスされうる場合が生じることが認められた。さらに、(従来の) 文レベルでは文法許容性が低い場合でも、文を超えた談話構造レベルからライセンスされうるという知見が得られた。この点は、談話構造と文文法とのダイナミックな双方向的関連の現れと考えられる。一般的な言語モデルの観点からいえば、実際の言語データでは、いわゆる「文法性」は文法の階層間の相互作用に依拠するといえる。

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 山添直樹、白井賢一郎、意味論における「イベント」、中京大学国際教養学部論叢、査読無、2012、4巻、57-90
- ② 丹羽敏行、河宮信郎、白井英俊、キーボードの文字配列の改良研究、中京大学現代社会学部紀要、査読無、2012、5巻、165-183
- ③ 有田節子、現代日本語の複文におけるテンスとモダリティ、本語文法学会第12回大会発表予稿集、査読有、2011、1-10
- ④ 白井賢一郎、形式意味論と認知言語学、日本認知言語学会論文集、査読有、9巻、2009、551-561

[学会発表] (計3件)

- ① 有田節子、現代日本語の複文におけるテンスとモダリティ、日本語文法学会第12回大会、2011年12月3日、東京外国語大学
- ② 有田節子、省略と残存一発話冒頭に出現する提題助詞「は」の分布と機能一、Seventh International Conference on Practical Linguistics of Japanese (ICPLJ7)、2011年3月6日、サンフランシスコ州立大学
- ③ 奥村泰章、白井英俊、談話構造コーパスの提案、認知科学会第26回大会、2009年9月12日、慶応義塾大学

[図書] (計2件)

- ① 有田節子、開拓社、ことばとコミュニケーションのフォーラム、2011、265
- ② 白井英俊、他、朝倉書店、計量国語学事典、2009、423

6. 研究組織

(1) 研究代表者

白井 英俊 (SHIRAI HIDETOSHI)

中京大学・情報理工学部・教授

研究者番号：10134462

(2) 研究分担者

白井 賢一郎 (SHIRAI KENICHIROU)

中京大学・国際教養学部・教授

研究者番号：20162753

有田 節子 (ARITA SETSUKO)

大阪樟蔭女子大学・学芸学部・教授

研究者番号：70263994

(3) 連携研究者

なし