

令和 2 年 6 月 11 日現在

機関番号：13901

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K00303

研究課題名(和文)形式意味論的アプローチに基づく同時的言語解析システムの開発

研究課題名(英文)Developing incremental interpretation system based on a formal semantic approach

研究代表者

加藤 芳秀(Kato, Yoshihide)

名古屋大学・情報連携統括本部・准教授

研究者番号：20362220

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：自然言語の同時的解析モデルとは、文や発話を途中まで読んだり聴いたりした段階で、その内容を捉えるモデルである。本研究では、形式意味論的アプローチに基づき同時的意味解析システムを開発する。本研究の成果は、以下のようにまとめられる。

(1) 談話表示理論の意味論を拡張し、文の断片に対する意味解釈を定式化した。2つの相補的な意味解釈を定義することにより、文を読んでいくに従って段階的にその意味内容が定まる過程をモデル意味論的に定式化できた。

(2) 同時的構文解析において、深いレベルの構文構造(グラフ構造)を扱うための要素技術として、グラフ構造を近似的に木構造として表現する手法を開発した。

研究成果の学術的意義や社会的意義

同時的意味解析に関するこれまでの研究においては、意味構造を同時的に構成する方法に主眼が置かれており、その意味内容を形式意味論に基づいて解釈する方法についてほとんど検討されていなかった。本研究はそのような方向性の研究の端緒を開くものである。また、グラフとして表現される構文構造を木構造で近似的に表現する手法は、同時的構文解析だけでなく、一文単位の通常の構文解析にも利用可能であり、解析精度向上に有効であることも実験により確認されている。そのため、自然言語解析の分野における学術的意義を有する。

研究成果の概要(英文)：This research project aims to develop an incremental interpretation system based on formal semantic approach. The contribution of the research is summarized as follows: Firstly, the project proposed a model-theoretic approach to incremental interpretation where sentence prefixes have semantic values. The proposed semantics is based on Discourse Representation Theory. In the semantics, two complementary interpretations stipulate a semantic content of each sentence prefix. Secondly, the project developed the method that approximates graph-structured syntactic representations by trees. This approximation method enables tree-based incremental parsers to handle graph-structured representations.

研究分野：自然言語処理

キーワード：漸進的解釈 意味解析 形式意味論 談話表示理論 構文解析 非局所的依存関係 空所化構文

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

自然言語の同時的解析モデルとは、文や発話を途中まで読んだり聴いたりした段階で、その内容を捉えるモデルである。このようなモデルは、リアルタイム音声言語処理システム（例えば、同時通訳システムなど）の言語理解モジュールとしてその実現が期待されている。特に話し言葉においては、文が完結せずに中断されることもたびたび生じるが、文を処理単位として想定した解析モデルではそのような中断文の内容を捉えることができない。一方、同時的解析モデルは中断文の解析にも適している。また、心理言語学の分野では、人間の言語理解プロセスに見られる同時性をモデル化するために、このようなモデルの開発が進められている。

言語解析には構文解析、意味解析、談話解析など様々なレベルの解析が存在し、同時的構文解析については、これまでに計算言語学の分野において精力的に研究が進められている。実際の英語新聞記事を対象として、一文単位の構文解析に近い性能を持つ同時的な構文解析が実現されている。

一方、同時的意味解析についてはその検討が不十分な状況であった。意味構造を同時的に構成する方法はさまざま提案されていたが、同時的な解析処理の過程で構成される意味構造をどのように解釈するかという点については明らかになっていなかった。このような状況から脱却するためには、同時的な言語理解の観点に立って、文の断片を対象とした意味論を確立する必要があった。また、同時的な意味解析システムを実現するためには、その前段階にあたる同時的構文解析において、深いレベルの解析処理を実現する必要があった。

2. 研究の目的

本研究の目的は、自然言語文の同時的理解を実現するための解析技術の開発である。意味解析をその中心課題とする。具体的には次の要請を満たす意味解析システムの実現を目指す。

- 文を途中まで読んだ段階で、途中まで読んだ文の断片の意味内容を捉える（同時性）
- 捉えた意味内容は、文全体の意味内容と矛盾しない（整合性）
- 実際のテキストに適用できる

本研究では、同時的な言語理解の観点に立ち、形式意味論的アプローチに基づき文の断片の意味解釈を定める意味論を設計する。また、この意味論に基づいた意味解析システムを開発する。加えて、意味解析の前段階にあたる構文解析において、意味解析に適した構文構造を生成できる同時的構文解析を開発する。

3. 研究の方法

(1) 文の断片に対する意味構造を設計し、本研究で開発する同時的意味解析システムのベースとする。アドホックな意味解析システムとならないようにするために、形式意味論的なアプローチを採用する。文の断片に対する意味構造は解析の同時性を達成するために次の特徴を持つように設計する。

- 文を先頭から順に読んでいくにつれて、その意味構造は段階的に更新され、最終的に文全体に対する意味構造を得ることが可能である
- 未入力部分の意味構造を不確定なものとして表現できる

設計した文の断片に対する意味構造に対して、その意味解釈を形式意味論的に定式化する。

(2) 意味解析を実現するために、その前段階にあたる構文解析において深いレベルの解析を実現する。同時的構文解析においては、それが生成する構文構造として木構造が想定されているが、深いレベルの情報を扱う構文構造は一般にグラフ構造となるため、同時的構文解析において扱うことができない。本研究では、グラフ構造を木構造により近似的に表現する方法を開発し、木構造ベースの同時的構文解析においてもそのような情報を処理できるようにすることを目指す。具体的には以下のような構文解析を開発する。

(A) 非局所的依存関係を同定できる構文解析を開発する。非局所的依存関係は、ゼロ代名詞や構成素の移動といった言語現象を表現している。非局所的依存関係の情報は意味解析において重要な役割を果たすため、構文解析においてこれを同定しなければならない。

(B) 空所化構文を解析可能な構文解析を開発する。空所化とは、等位接続された句や節といった構成素に共通する要素が、片方の構成素から省略される現象である。空所化構文を含む文を意味解析するためには、省略された要素を補う処理が必要となるが、これを実現する構文解析を開発する。

非局所的依存関係や空所化に関わる情報を句構造のような構文構造において表現すると、その構造はグラフとなるが、(A)及び(B)は、構文構造を木構造としたままで処理するものである。すなわち、木構造に基づく同時的構文解析はこれを扱うことができる。

4. 研究成果

(1) 形式意味論においては、文に対して命題的な解釈（例えば真理条件）を与えることが多いが、同時的な意味解析における困難の一つは、文の断片は基本的に命題を表しておらず、命題と

しての解釈を与えることができない点である。本研究では、文の断片に対する意味構造をモデル理論的意味論に基づき解釈する方法として、談話表示理論 (Discourse Representation Theory, DRT) をベースにした意味論を開発した。DRT において意味構造は談話表示構造 (Discourse Representation Structure, DRS) と呼ばれるが、その意味解釈は割当 (DRS 中に出現する談話指示子と呼ばれる変項に個体を与える関数) の非決定的な更新として定義される。本研究では、文の断片に対する意味構造である部分 DRS の解釈を、割当の更新を規定する二つの集合として定義する。一方の集合は割当の更新の可能性を表現し、もう一方の集合は割当の更新の必然性を表現する。本研究において開発した意味解釈は、以下の性質を持っており、同時性をうまく捉えた意味論となっている。

- 文を同時に意味解析する過程を情報の単調性として説明できる。
- 文に対する DRS の意味解釈とその文の断片に対する部分 DRS の意味解釈との整合性が保たれている

Word	Partial DRS	$[\cdot]^\square$	$[\cdot]^\circ$	Truth value
A	$S_1 \equiv U_2(\lambda Q. (([x_1] \]; U_1x_1); Qx_1))$	$[S_1]^\square(\Gamma) = \emptyset$	$[S_1]^\circ(\Gamma) = \{\{\}\}$	-
student	$S_2 \equiv U_2(\lambda Q. ([x_1]stu(x_1)]; Qx_1))$	$[S_2]^\square(\Gamma) = \emptyset$	$[S_2]^\circ(\Gamma) = \{\{\}\}$	-
uses	$S_3 \equiv [x_1]stu(x_1); (U_3(\lambda Y. [\]use(x_1, Y))))$	$[S_3]^\square(\Gamma) = \emptyset$	$[S_3]^\circ(\Gamma) = \{[x_1 \rightarrow d_1], [x_1 \rightarrow d_2]\}$	-
every	$S_4 \equiv [x_1]stu(x_1), ([x_4] \]; U_4x_4) \Rightarrow [\]use(x_1, x_4)]$	$[S_4]^\square(\Gamma) = \emptyset$	$[S_4]^\circ(\Gamma) = \{[x_1 \rightarrow d_1], [x_1 \rightarrow d_2]\}$	-
red	$S_5 \equiv [x_1]stu(x_1), ([x_4] \]; ([\]red(x_4)]; U_5x_4) \Rightarrow [\]use(x_1, x_4)]$	$[S_5]^\square(\Gamma) = \{[x_1 \rightarrow d_1]\}$	$[S_5]^\circ(\Gamma) = \{[x_1 \rightarrow d_1], [x_1 \rightarrow d_2]\}$	T
laptop	$S_6 \equiv [x_1]stu(x_1), [x_4]red(x_4), laptop(x_4) \Rightarrow [\]use(x_1, x_4)]$	$[S_6]^\square(\Gamma) = \{[x_1 \rightarrow d_1]\}$	$[S_6]^\circ(\Gamma) = \{[x_1 \rightarrow d_1]\}$	T

Model

$D = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7\}$
 $I(student) = \{<d_1, <d_2, >\}$
 $I(laptop) = \{<d_5, <d_6, >\}$
 $I(tablet) = \{<d_3, <d_4, >\}$
 $I(red) = \{<d_5, <d_6, <d_7, >\}$
 $I(blue) = \{<d_1, <d_2, >\}$
 $I(uses) = \{<d_1, d_2, >, <d_1, d_5, >, <d_1, d_6, >, <d_2, d_7, >\}$

図 1：文の断片の意味構造とその解釈

図 1 は、文 “A student uses every red laptop” に対する同時的な意味解析の過程を示したものである。表の 2 列目は、文の断片に対して構成される部分 DRS であり、3, 4 列目は、図に示したモデルの下での部分 DRS の 2 種類の意味解釈である。例えば、3 行目の意味解釈は、文の断片 “A student uses ...” において “student” はモデルにおける個体 d_1 もしくは d_2 を指示する可能性が残されているが、その他の個体を指示する可能性はないことを表現している。五行目は、文の断片 “A student uses every red ...” に続く単語が何であっても (“laptop” であっても “tablet” であっても) その文は、このモデルの下では真となることが保証されることを表現している。この例に示したように、本研究で開発した意味論は、文の断片に対してモデル理論的な意味解釈を与えることができる。

(2A) 非局所的依存関係は、字面上あらわれない空要素とフィラーと呼ばれる構成素との対として表現される (これは、空要素の位置にフィラーにあたる構成素が存在するものとして解釈すべきであることを表している)。本研究では、非局所的依存関係自体は構文構造から削除し、これを復元するための情報を構文構造にアノテーションとして埋め込む手法を開発した。結果として得られる構文構造は元のグラフ構造を近似的に表現した木構造となっており、木構造をベースとした構文解析で取り扱うことができる。アノテーション処理は 2 つのステップからなる。最初のステップは、フィラーの情報をアノテーションするステップである (図 2 左)。次のステップは、空要素を構文構造のノードとして埋め込むステップである (図 2 右)。非局所的依存関係の復元は逆向きの処理にあたるが、アノテーション処理は非可逆であるため、言語学的な知見に基づいたヒューリスティックなルールを用いて復元する。構文木コーパス Penn Treebank を用いた実験により、グラフ構造を近似的に表現した木構造から、99.5%の精度で元のグラフが復元できることを確認した。また、開発した手法は、非局所的依存関係同定手法 (F 値: 78.4%) と比べて、大幅に性能で上回ることを確認した (F 値: 89.0%)。

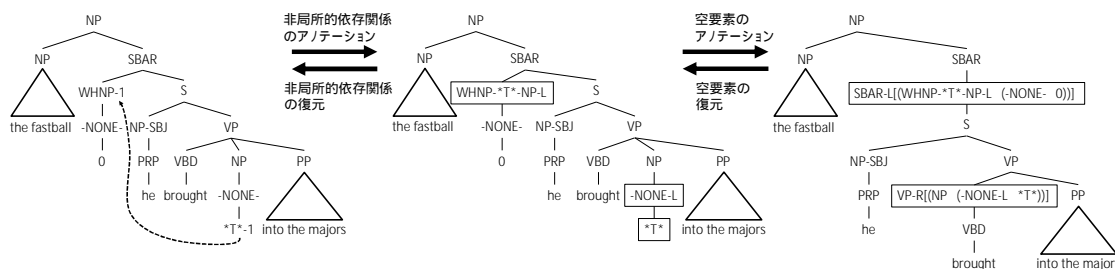


図 2：構文構造 (グラフ) の木による近似, 及びその復元

(2B) 空所化構文を解析する手法として、構成素の文法・意味的役割に着目し、役割の一致に

基づき対応付けを行う手法を開発した。空所化構文において省略された要素は、等位接続されている「省略がなされた構成素」と「省略がなされていない構成素」の間のある種の対応関係に基づいて特定することができる。従来の手法においては、構文構造においてこの対応関係を表現し、直接的に構文解析モデルとして学習するアプローチがとられていた。しかし、空所化構文は出現頻度の低い言語現象であるため、学習のためのデータが十分ではなく、モデルの構築が困難であるという問題があった。本研究では、対応関係を直接的に学習するのではなく、対応関係にある要素は同一の役割を担うという点に着目し、文法・意味的役割を同定する構文解析モデルを構築する。文法・意味的役割に関する情報は通常の文にも多数含まれるものであるため、十分な学習データを用意することができる。すなわち、空所化構文に現れる対応関係と比較して学習が容易である。続く処理として、文法・意味的役割の一致に基づくシーケンスアラインメントにより対応関係を同定する。Penn Treebank を用いた実験により提案手法の性能を評価したところ、従来の空所化構文解析手法（F 値：12.9%）よりも大幅に性能で上回ることを確認できた（提案手法の F 値：70.8%）。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計7件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Yoshihide Kato, Shigeki Matsubara
2. 発表標題 PTB Graph Parsing with Tree Approximation
3. 学会等名 The 57th Annual Meeting of the Association for Computational Linguistics (国際学会)
4. 発表年 2019年

1. 発表者名 加藤 芳秀, 松原 茂樹
2. 発表標題 文法・意味的役割に基づく空所化構文の解析
3. 学会等名 言語処理学会第26回年次大会発表論文集
4. 発表年 2020年

1. 発表者名 加藤 芳秀, 松原 茂樹
2. 発表標題 非局所的依存関係同定の木構造を用いた近似解法
3. 学会等名 言語処理学会第25回年次大会発表論文集
4. 発表年 2019年

1. 発表者名 Yoshihide Kato, Shigeki Matsubara
2. 発表標題 Model-Theoretic Incremental Interpretation Based on Discourse Representation Theory
3. 学会等名 The 32nd Pacific Asia Conference on Language, Information and Computation (国際学会)
4. 発表年 2018年

1. 発表者名 生駒 流季, 加藤 芳秀, 松原 茂樹
2. 発表標題 構文解析のためのRNNモデルの学習における構造化損失関数に関する考察
3. 学会等名 第16回情報学ワークショップ
4. 発表年 2018年

1. 発表者名 加藤 芳秀, 松原 茂樹
2. 発表標題 モデル理論的意味論に基づく漸進的意味解釈の定式化
3. 学会等名 言語処理学会第24回年次大会発表論文集
4. 発表年 2018年

1. 発表者名 高尾 大樹, 加藤 芳秀, 松原 茂樹
2. 発表標題 助動詞を考慮した自然論理タブロー法
3. 学会等名 言語処理学会第24回年次大会発表論文集
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----