

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 9 日現在

機関番号：13801

研究種目：基盤研究(C)

研究期間：2009～2013

課題番号：21520398

研究課題名(和文) 構文意味解析技術に応用可能な統語論・意味論インタフェースの研究

研究課題名(英文) A Study on Syntax-Semantics Interface for Syntactic/Semantic Parsing

研究代表者

近藤 真 (Kondo, Makoto)

静岡大学・情報学研究科・教授

研究者番号：30225627

交付決定額(研究期間全体)：(直接経費) 3,800,000円、(間接経費) 1,140,000円

研究成果の概要(和文)：ある意味内容を言語で表現する場合、その表現のしかたは何通りも存在する。このように異なる構文構造を持つ文の同義性を判定することは計算機にとって非常に難しい問題の一つである。この問題を解決するためには、構文構造に頼らずに同義性を判定できるかたちで文の意味が表現されなければならない。本研究課題では、Takagi et al. (2006)の意味表現を採用し、入力文が持つ文法的な特徴だけでなく、文脈を参照しながら入力文の意味表現を生成する構文意味解析器を開発し、その構文意味解析器を日本語対話システム、日本語教育システムに組み込むことを通して、計算機による意味理解のための手法を整備した。

研究成果の概要(英文)：Natural language has a number of ways to express the same semantic content. Such synonymous expressions pose a great difficulty to natural language understanding by computers. In order to solve the problem, we need semantic representation which enables the system to judge synonymity of natural language expressions without referring to their grammatical construction. We have adopted the framework for semantic representation proposed by Takagi et al. (2006), and developed a syntactic/semantic parser. The parser generates semantic representation not only by using grammatical features of input sentences but also by referring to the context. We have developed Japanese dialog system and Japanese education systems which take advantage of the parser. We have also developed techniques for natural language understanding based on the semantic representation.

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：統語論 意味論 自然言語処理 構文意味解析

1. 研究開始当初の背景

(1) 従来、言語学における統語論や意味論の研究と、自然言語処理技術の研究(とりわけ構文解析、意味理解に関する研究)は、それぞれ独立して行われてきた。言語学における統語論、意味論の研究は、究極的には人間の言語能力の解明を目指すものであるため、自然言語処理技術との親和性を考慮して研究されることはほとんどなかった。また、自然言語処理の構文解析、意味理解に関する研究は、言語学の知見を応用して行われることはあっても、その研究過程で生じる様々な問題が統語論・意味論研究にフィードバックされることもなかった。

(2) 本研究課題の先行研究である科学研究費補助金・萌芽研究・「自然言語処理技術への応用を前提とした構文意味解析のための言語学的基礎理論の研究」(平成 18-20 年度)において、本研究課題研究代表者と連携研究者らは構文意味解析器の試作・評価を行っていた。先行研究の成果から、自然言語処理の構文意味解析における諸問題を、特定の言語理論の枠組みにとらわれることなく、一つのプロブレムセット(problem set)として検討することで、個々の言語理論の下では見えてこなかった問題の発見や、言語研究全体に対する新しい視点の提供が期待されるとの着想に至った。

2. 研究の目的

(1) 本研究課題では、構文解析と意味解析を融合し文脈を参照しながら、入力文が持つ統語情報に加えて、意味情報・文脈情報を考慮して入力文の構文構造を決定し、構文構造の決定と同時に入力文の意味表現を生成する構文意味解析器の開発と評価を行うとともに、そのような構文意味解析器の開発に資する統語論・意味論インタフェースに関する言語理論研究を目指す。

(2) 統語構造と意味論的項構造とのあいだに一定のある程度安定した関係が見出せるが、その関係は必ずしも計算機が容易に処理できるような一対一の対応関係にはなっていない。そのため、構文解析と意味理解を融合した構文意味解析器の開発においては、統語構造と意味論的項構造との差異を吸収可能な統語論・意味論インタフェースに関する理論が必要となる。本研究課題では、このような統語構造と意味論的項構造との差異を吸収可能な理論を、構文意味解析技術に応用可能なかたちで開発する。

3. 研究の方法

(1) 前述[1-(2)]の先行研究では Takagi et al. (2006) “Semantic Representation for Understanding Meaning Based on Correspondence between Meanings” *Journal of Advanced Computational Intelligence and Intel-*

ligent Informatics, Vol.10 No.6, pp.876-912. で開発された意味表現を採用している。この意味表現では、各実体概念・述語概念の意味内容を属性と属性地のペアのセットとしてフレーム形式で表現する。さらに、概念間の関係は各概念を結ぶポイントで表現される。具体的には、概念 A の属性 a がその属性値として概念 B を取る場合、これらの関係は属性 a の属性値から概念 B へのポイントとして表現される。これは、原則的には名詞、動詞といった内容語の意味がフレームで表され、述語と項の関係や、修飾語と被修飾語の関係がポイントで表現されていると見なすことができる。これにより、異なる統語構造を持つ文が同一の意味内容を表す場合、その同義性を属性・値の単位で比較・参照することが可能になる。

(2) 意味表現の具体例として、「浜松から東京へ行く」「東京へ行く」「出発地は浜松だ」の意味表現をそれぞれ図 1 から図 3 に示す。

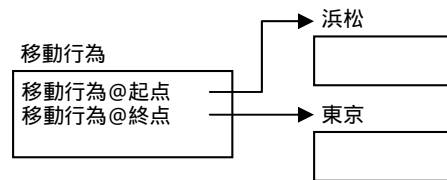


図 1 「浜松から東京へ行く」

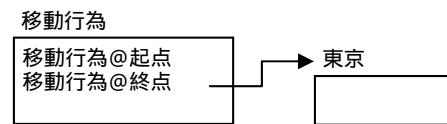


図 2 「東京へ行く」

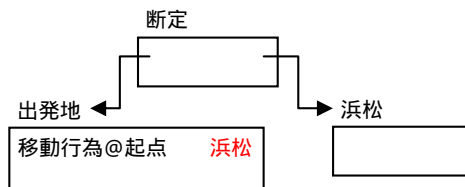


図 3 「出発地は浜松だ」

図 1 では動詞「行く」に対応する「移動行為」フレームが、出発地を表す「移動行為@起点」属性と、到着地を表す「移動行為@終点」属性を持つ。これらの属性の値としてそれぞれ「浜松」フレームへのポイントと「東京」フレームへのポイントが格納されることにより、図 1 の意味表現全体で「浜松から東京へ行く」の意味内容が表現される。同様に図 2 は「東京へ行く」の意味内容を表す意味表現となっている。属性と属性のあいだには「属性は属性値である」という断定の関係が成立しているため、本意味表現では、断定の述語が用いられた場合にのみ、「断定」フレームの属性に断定の対象となる概念フレームへのポイントが格納される。図 3 では断定対象となる「出発地」フレームへのポイントが「断

定」フレームの属性に格納され、その属性値に「浜松」フレームへのポインタが格納されることにより、「出発地は浜松だ」に対応する意味が表現されている。

(3) 「出発地」という名詞自体が「移動行為@起点」を表す属性名詞であるため、これが断定の述語フレームを介して「浜松」フレームと結ばれることにより、「出発地」フレームの値として「浜松」が伝搬される(図3赤字部分)。いっぽう、「移動行為」フレームもまた、図1、図2に示すようにその内部に「移動行為@起点」属性を持つ。そのため、「東京へ行く」という入力文に続けて「出発地は浜松だ」という文が入力された場合、「属性・値」のペアを単位として文脈を参照することにより、前文(「東京へ行く」)の意味表現中の「移動行為@起点」の値として、後続文(「出発地は浜松だ」)の「浜松」を伝搬することが可能になる。これは、図1のように「浜松から東京へ行く」と一文で表現された場合であっても、「東京へ行く。出発地は浜松だ」のように文をわけて表現された場合であっても、それらの統語構造の違いに影響されず、一様な手続きでその同義性を評価できることを意味している。

(4) 本研究課題では、上記の意味表現を採用した構文意味解析器を日本語教育用対話訓練システム、日本語教育用ディクトグロスシステム、日本語対話システムに組み込むことにより、構文意味解析器を評価・改善するとともに、このような統語構造の差異を吸収する意味表現に関する理論を洗練させた。

4. 研究成果

(1) 本研究で開発された構文意味解析器を日本語教育用対話訓練システムに、意味理解部の核として組み込んだ。この対話訓練システムは、ロールプレイにおける学習者の対話相手の役割を果たす。また、システムはロールプレイ中で学習者が発話すべき内容を、先の意味表現形式で保持する(これを「シチュエーション知識」と呼ぶ)。これにより、文法的な入力であれば学習者は定型文に縛られることなく、自由な入力を行うことができる。また、非文が入力された場合には、システムが持つシチュエーション知識と、入力文に含まれる内容語を比較することにより、学習者が言わんとすることを推定し、対話を継続することが可能になる。これらから、学習者の入力文における文法的誤りの有無が判定できると同時に、文法的な発話であっても、それが文脈に即した発話であるか否かが判定できる。

また、本意味表現が統語構造の差異を吸収する性質を持つことから、学習者入力文が定型表現と同じ統語構造を持つものであるか、あるいは異なる統語構造を持つが意味的に同等な入力であるかを区別して判定するこ

とができる。これは、近年、外国語教育で注目されている Focus on Form (FonF) という教授法を教育用システムで実現するために必須の機能である。そこで、日本語教育用対話訓練システムを FonF に対応させるよう拡張し、本意味表現を用いた対話訓練システムが計算機による FonF の実現にとって有用であることを確認した。この成果をまとめた Kondo et al. (2010) [5. 主な発表論文等・雑誌論文・] は、ICCE 2010 において Best Paper Award にノミネートされた。

(2) (1)の日本語教育用対話訓練システムの意味理解部を核として、日本語教育用ディクトグロスシステムを開発した。ディクトグロスとは、学習者が聞き取った内容を復元する外国語学習活動であり、意味的に等価な文が復元されればよいという点が、従来のディクテーションとは異なっている。本研究課題で採用している意味表現は、同義性を判定可能であると同時に、ディクトグロスで復元すべき文の意味内容をシチュエーション知識として保持することにより、学習者入力文の誤りも判定することができる。これらの機能を実装した日本語教育用ディクトグロスシステムを開発し、その成果まとめた Kondo et al. (2012) [5. 主な発表論文等・雑誌論文・] は、ICCE 2012 において Special Mention Award を受賞した。

(3) 非文を含む入力に対応する必要がある日本語教育用システムにおいて、本意味表現を採用したシステムが上記の成果をあげたことは、本意味表現が実用レベルの対話システムにおいても有用であることを強く示唆している。実用レベルの対話システムにおいては、ユーザの入力が常に文法的である保証はなく、むしろ一定程度の非文の入力を想定しなければならない。統語構造によらず、文脈を参照して意味の比較・統合が可能な本意味表現を採用したシステムが、この意味において一定の頑健性を備えていることが明らかになった。

(4) 本意味表現を用いた意味理解においては、入力文の意味を文脈やシステムが持つ知識と比較する際に、入力文中の概念とそれが持つ「属性・値」を先行文脈やシステムの知識から検出している。その検出にあたっては、概念間の上位・下位関係や全体・部分関係に基づいて検出を行っていた。いっぽう、現実の対話においては、先行文脈で言及された概念と、入力文中の概念とのあいだで属性値の伝搬が起こる際に、上位・下位関係や全体・部分関係にない概念間で属性値の伝搬が起こる場合がある。

このような事例に対応するために、ある全体現象が連続する複数の部分現象によって構成される場合に着目し、その部分現象間での属性値の伝搬を可能にさせる連続現象モ

デルを開発した。

連続現象モデルでは、本意味表現形式で表された具体的な値を持たない部分現象概念群と、それらが全体現象を構成する際の生起順序が抽象的にモデル化されている。また、全体現象の中で必ず生起する現象と随意的に生起する現象とが明示的に区別されている。この連続現象モデルはいわば全体現象のひな型のようなものであり、これに基づいて具体的な連続現象のインスタンスが文脈中に展開される。これにより部分現象間での生起順序を決定し、その生起順序に基づいて時間・場所に関する属性値を部分現象間で伝搬することが可能になった。この成果は Noguchi et al. (2013) [5. 主な発表論文等・雑誌論文・] として発表された。

5. 主な発表論文等

[雑誌論文](計7件)

Yasuhiro NOGUCHI, Satoru KOGURE, Makoto KONDO, Ichiro KOBAYASHI, Hideki ASOH, Akira TAKAGI, Tatsuhiro KONISHI, Yukihiro ITOH (2013) "Event Sequence Model for Semantic Analysis of Time and Location in Dialogue System" *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)*, pp.339-348. 査読有

Makoto KONDO, Ryo SANNO, Asanori TASHIRO, Yasuhiro NOGUCHI, Satoru KOGURE, Tatsuhiro KONISHI, Yukihiro ITOH (2012) "Development of a Dictogloss System Oriented for Focus on Form" *Proceedings of the 20th International Conference on Computers in Education (ICCE 2012)*, pp.1-8. 査読有

Makoto KONDO, Yoko DAICHO, Ryo SANNO, Yasuhiro NOGUCHI, Satoru KOGURE, Tatsuhiro KONISHI, Yukihiro ITOH (2010) "Form-Wise Error Detection in a FonF-Based Language Education System" *Proceedings of the 18th International Conference on Computers in Education (ICCE 2010)*, pp.9-16. 査読有

[学会発表](計6件)

田代朝紀、小暮悟、野口靖浩、小西達裕、近藤真、伊東幸宏「FonFを指向する日本語ディクトグロスシステムの開発と評価」教育システム情報学会(JSiSE)2013年度第3回研究会. 2013年9月14日. 静岡大学

西巻孝祥、野口靖浩、小暮悟、小西達裕、小林一郎、麻生英樹、高木朗、近藤真、伊東幸宏「連続現象モデルを用いた意味理解と対話システムへの利用」人工知能

学会全国大会第26回大会「意味と理解のコンピューティング」2012年6月15日. ゆ〜あいプラザ山口県社会福祉会館
佐野良、野口靖浩、小暮悟、小西達裕、近藤真、伊東幸宏「FonFを指向する日本語対話訓練システムにおける学習者の誤り判定機構の拡張」第35回教育システム情報学会全国大会. 2010年8月28日. 北海道大学

[図書](計1件)

小野尚之、近藤真、藏藤健雄、松岡和美、藤本幸治(編著)(2013)『生成文法の軌跡と展望』金星堂. pp.2-25, pp.67-87.

6. 研究組織

(1)研究代表者

近藤 真 (KONDO, Makoto)
静岡大学・情報学研究科・教授
研究者番号: 30225627

(2)連携研究者

小西 達裕 (KONISHI, Tatsuhiro)
静岡大学・情報学研究科・教授
研究者番号: 30234800

小暮 悟 (KOGURE, Satoru)
静岡大学・情報学研究科・講師
研究者番号: 40359758