

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 9 日現在

機関番号：32612

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24520437

研究課題名(和文) 意味フレームとコーパスに基づく文意理解のための言語資源の構築

研究課題名(英文) Constructing a frame-based language resource for sentence-meaning understanding

研究代表者

小原 京子 (OHARA, Kyoko)

慶應義塾大学・理工学部・教授

研究者番号：00286650

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：本研究の目的は、研究代表者が構築中の日本語フレームネット(JFN)を、語彙の意味情報に加え、構文の意味情報をも含む語彙・構文複合言語資源へと発展させることであった。文全体の意味に、語彙素の意味以外のどのような文要素の意味が関与しているかを明らかにし、それらを記述して言語資源とする方策を探った。コーパス上に現れる文に用いられる構文の洗い出しを行ない、フレーム意味論の手法を用いてそれぞれの構文の意味を分析した。それらの結果をアノテーションとして蓄積するための、構文アノテーションの範囲を定義することができた。さらに、構文の意味情報と意味フレーム情報との関係が明確化できた。

研究成果の概要(英文)：On the basis of analyses of Japanese constructions for the purpose of building a construction in the Japanese FrameNet (JFN) project, it became clear, through this research project, that grammatical constructions can be classified based on whether they evoke frames or not. Recognizing such a distinction among constructions enabled FrameNet-style linguistic resources to have a proper division of labor between frame annotations and construction annotations. Moreover, in addition to the three kinds of “meaningless” constructions which have been proposed already, this research project suggested that there is yet another subtype of constructions without meanings. Furthermore, the research project adds supported to the claim that there may be constructions without meanings (Fillmore, Lee-Goldman & Rhomieux, 2012) in a current debate concerning whether all constructions should be seen as meaning-bearing (Goldberg, 2006: 166-182).

研究分野：言語学

キーワード：言語資源 フレーム意味論 構文文法 コーパス言語学 語彙意味論 認知言語学 日本語構文

1. 研究開始当初の背景

本研究に関連する国内・国外の研究動向は以下の通りであった。

(1) コーパス言語学の浸透

現代日本語の書き言葉に関する初の大規模均衡コーパス「現代日本語書き言葉均衡コーパス」(BCCWJ)が公開されたことにより、コーパス日本語学の研究が今後一層盛んになると期待されていた。認知言語学のアプローチによる日本語分析においてもコーパス言語学の手法が浸透しつつあった。中でもフレーム意味論や構文文法による日本語分析にはコーパスデータを用いた例があったが、実際に分析結果をコーパスアノテーションとして蓄積し言語資源とする試みとしては、本研究代表者が研究協力者たちと構築中の日本語フレームネットがほぼ唯一のものであった。

(2) フレーム意味論の精緻化とフレームネットの発展

日本語フレームネットとは、ある語を用いる際に話者が想起する背景知識に着目しそれらを意味フレームとして定義することで、その語の意味を詳細に定義した言語資源である。日本語フレームネットの理論的基盤であるフレーム意味論は、項構造、格フレーム、結合価パターンに関する枠組みとして広く知られているが、本来は文全体の意味と関連づけて文中に現れる語彙素の意味を理解しようとする理論的枠組みである。しかし、文の意味はそこに現れる語彙素の意味の和ではなく、語彙素以外の様々なレベルの「構文」の意味も文全体の意味に関与している。そこで、英語フレームネット・プロジェクトでは語彙の意味と構文の意味の統一的記述のための枠組みを検討中であった。このような背景から、科研費特定領域「日本語コーパス」に公募研究班として参加した際には、日本語フレームネットを語彙の意味情報に加え構文の意味情報をも含む語彙・構文複合資源へと発展させていくためのパイロットスタディを行った。

(3) 言語資源構築の重要性の高まり

近年様々な日本語資源が構築され公開されてきているが、若干の例外を除き、言語処理研究者が自然言語処理システムでの利用を念頭に作成したものが多く、言語学的枠組みに則っているものは少ない。したがって、言語研究者が言語データ分析時に参照できる言語資源、日本語学習者が利用できる言語資源というものはほとんどなかった。

上記のような学術的背景に基づき開始した本研究の学術的な特色・独創的な点は以下の3点に集約できる。

(4) コーパス言語学：通常言語分析者はコーパスデータを分析するのみであるが、コーパスデータに対するアノテーションをも行う点に特色があった。

(5) フレーム意味論：項構造、格フレーム、結合価パターンを見るためのみならず、文全

体の意味への寄与という観点からコロケーションや複合辞の意味分析にフレーム意味論を用いる点。

(6) 言語資源構築を将来的に目指す：多くの言語資源が言語処理研究者中心に構築されているのに対し、言語学研究者が言語学的見地に基づき、言語処理研究者と連携しながら構築しようとしていたこと、である。

そして、当時予想された結果と意義としては、

(7) 言語学的意義：語彙・構文複合言語資源とすることで、日本語フレームネットがフレーム意味論と構文文法の集大成の場となる。

(8) 日本語学習者用資源として：語彙素以外の様々なレベルの構文の意味は一对一に他言語の表現形式と対応しないことが多く、日本語母語話者以外には理解や運用が困難である。これらの意味情報がアノテーションされたコーパスを、中上級日本語学習用言語資源として活用してもらうことを将来的に目指す。

(9) 自然言語処理用資源として：意味フレームに基づく文の意味分析結果の蓄積は、文書要約、文書の意味構造解析、文書からの知識獲得、評判分析などの意味処理が関わる自然言語処理システムに応用することを目指す。の、3点であった。

2. 研究の目的

本研究の目的は、研究代表者が構築中の日本語フレームネット(JFN)を、語彙の意味情報に加え、構文の意味情報をも含む語彙・構文複合言語資源へと発展させることであった。具体的には、文全体の意味に、語彙素の意味以外にどのような文要素の意味が関与しているかを明らかにし、それらを記述して言語資源とする方策を探ることであった。将来的には自然言語処理技術や日本語学習者にとって有益な言語資源とすることを目指した。

3. 研究の方法

(1) 意味記述の枠組みとしてフレーム意味論を用いた。

(2) 統語的特徴を記述、分析するための枠組みとして構文文法を用いた。フレーム意味論と構文文法を用いることで、語彙素以外の文要素の意味も語彙素の意味と統一的に記述することができるからである。

(3) 文の意味の分析対象として、BCCWJ コーパス上の文を用いた。

4. 研究成果

(1) BCCWJ コーパスから抽出した文を分析し、構文の洗い出しを行ない、それぞれの構文の意味情報を分析した。

(2) 構文の意味情報の分析結果を、「コンストラクティコン」(構文データベース)として蓄積していくために、そしてそれらを自然言語処理技術や日本語学習者にとって有益

な情報とするために、どのような情報を記述すれば良いのかを検討し、「構文アノテーション」の範囲を明確化した。すなわち、これまでの英語・ポルトガル語・スウェーデン語などの他言語での「コンストラクティコン」構築プロジェクトでは、必ずしも「コンストラクティコン」に含めるべき情報の種類について統一見解があった訳ではなかったため、「構文アノテーション」の範囲について新たな提案を行なったということである。具体的には、各々の構文の統語的特徴と意味的特徴の両方を記述する、ということである。その結果、日本語フレームネットの「コンストラクティコン」では、認知言語学における重要なテーゼである「文法と語彙の連続性」を保持できることになる。

(3) 構文の意味情報と、意味フレーム情報との関係を明らかにすることができた。つまり、構文は、意味フレームを喚起するかどうかで分類することができることが明らかになった。さらに、従来の研究で「意味を持たない構文」と分類されていた構文は、1) 意味フレームを全く喚起しないか、2) 構文の階層構造において、下位のレベル(構文のサブタイプレベル)で意味フレームを喚起するか、のいずれかである、ということを指摘した。

(4) 従来の研究では、構文の種類は、1) 意味フレームを喚起する構文、2) 意味フレームを喚起しない構文、3) 構文のサブタイプレベルで意味フレームを喚起する構文、の3種類があることがわかっていたが、新たに4種類目の構文のタイプとして、4) 意味フレームは喚起しないが、その意味が「インタラクションフレーム」に関わる構文があることを指摘した。

(5) 本研究を通じて、日本語の語彙素の意味だけでなく、構文の意味をも記述し、文全体の意味を分析するために、フレーム意味論と構文文法の理論的枠組みが有効であることを示した。

(6) 日本語研究において、コーパス、特にBCCWJコーパスを使うことは研究者の間でかなり浸透して来ているが、コーパスへのアノテーションとして、従来のような述語項構造や、意味役割付与だけでなく、構文に対してもアノテーションができる、という可能性を追求することができた。

(7) 「コンストラクティコン」という概念を認知言語学の分野において浸透させることに、日本語分析の立場から貢献することができた。

(8) 英語フレームネットのみならず、スウェーデン語、ポルトガル語など、同じ枠組みを用いて「コンストラクティコン」を構築している海外の研究プロジェクトや研究者と、共同研究を行なうことができた。

(9) 言語分析だけでなく、日本語の「構文アノテーション」を行ない、「コンストラクティコン」を構築するための、オンラインのツールを整備することができた。

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 1件)

Ohara, Kyoko Hirose "Toward Constructicon Building for Japanese in Japanese FrameNet", Veredas On-Line - Frame Semantics and its technological applications. Vol. 17, No. 1. 2013, pp. 11-27. PPG LINGUISTICA/UFJF-JUIZ DE FORA. 査読有

[学会発表](計 13件)

Kabbach, Alexandre & Ohara, Kyoko Hirose "Japanese FrameNet Data Quality Overview: Toward Full Data Integrity." MAPLEX2015, February 9, 2015, Takinoyu, Tendo, Yamagata, Japan.

小原 京子、「日本語フレームネットに見る文法と語彙の連続性」,日本認知言語学会第15回年次大会、2014年9月21日、神奈川県、慶應義塾大学日吉キャンパス

Ohara, Kyoko Hirose "Relations between Frames and Constructions in Japanese FrameNet: The Syntax-Lexicon Continuum Revisited", 8th International Conference on Construction Grammar (ICCG8), September 4, 2014, University of Osnabrueck, Germany

Ohara, Kyoko Hirose "Relating Frames and Constructions in Japanese FrameNet." LREC '14, May 29, 2014, Reykjavik, Iceland

小原 京子「日本語フレームネットの構文アノテーション」,言語処理学会第20回年次大会、2014年3月18日、北海道、北海道大学

Ohara, Kyoko Hirose. "Constructicon Building in Japanese FrameNet: What it reveals about English and Japanese frames and constructions." The 12th International Cognitive Linguistics Conference (ICLC-12), June 24, 2013, University of Alberta, Canada.

Ohara, Kyoko Hirose, "Semantic Annotations in Japanese FrameNet: Comparing Frames in Japanese and English", The Eighth International Conference on Language Resources and Evaluation (LREC'12), May 24th 2013, Istanbul, Turkey

Ohara, Kyoko Hirose. "Frame-based analysis of polysemous words in Japanese FrameNet." International FrameNet Workshop 2013, April 19, 2013. The Shattuck Plaza Hotel, Berkeley, California, USA.

小原 京子 「日本語フレームネット：文意理解のためのコーパスアノテーション」言語処理学会第19回年次大会、2013年3月13日、愛知県、名古屋大学東山キャンパス

Ohara, Kyoko Hirose "Japanese FrameNet: Lexicon and Constructicon Building for Japanese", The 13th Korea-Japan Workshop on Linguistics and Language Processing: Corpora, Annotation and Human Language Processing, December 1st, 2012, Waseda University, Tokyo, Japan

Ohara, Kyoko Hirose "Japanese FrameNet: Toward Constructicon Building for Japanese", 7th International Conference on Construction Grammar (ICCG-7), August 11th, 2012, Hankuk University of Foreign Studies, Seoul, South Korea

Hasegawa, Yoko, Russell Lee-Goldman, Kyoko Hirose Ohara, Michael Ellsworth, and Charles J. Fillmore "The Frames-and-Constructions Approach to Paraphrase", ICCG7, August 11th, 2012, Hankuk University of Foreign Studies, Seoul, South Korea

Ohara, Kyoko Hirose "Toward a Constructicon Building for Japanese", FRIAS Workshop on Constructions across Grammars. June 18th, 2012, Freiburg Institute for Advanced Studies, Germany

〔図書〕(計 1件)

小原 京子 朝倉書店、「フレームと構文：文の意味記述のためのアノテーション」、田野村忠温編『コーパスと日本語学』、2014、102-118、177

〔産業財産権〕

出願状況(計 0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計 0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
取得年月日：
国内外の別：

〔その他〕
ホームページ等
<http://jfn.st.hc.keio.ac.jp/ja/>

6. 研究組織

(1) 研究代表者

小原 京子 (OHARA, Kyoko)
慶應義塾大学・理工学部・教授
研究者番号：00286650

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：