

令和 2 年 6 月 4 日現在

機関番号：17102

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K02631

研究課題名(和文) 文法理論の精度を評価するためのデータセットの作成

研究課題名(英文) Creating data sets for the evaluation of theories of syntax

研究代表者

上山 あゆみ (UEYAMA, Ayumi)

九州大学・人文科学研究院・教授

研究者番号：70221801

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本プロジェクトは、文法理論の精度を評価するためのデータセットの作成を目標とするものであった。文法性の判断は個人差が大きいと従来、このようなデータセットの作成は現実味が乏しいと考えられてきたが、本プロジェクトでは、文法構造と解釈の関係についての明確な予測体系を背景に、頑健なデータをよりわけ、より広い範囲の研究者にその成果を供することを目指していた。ただし、研究を進める中で、十分な予測を成り立たせるためには、動詞の項構造および名詞の解釈について、より踏み込んだ理論化が必要であるということが明らかとなり、データセット作成のためのレキシコン(脳内辞書)の整備を精力的に行なった。

研究成果の学術的意義や社会的意義

当初の目標とは少し異なるが、動詞の項構造をどのように記述し、それがどのように文構造に反映されるかというプロセスが具体的になりつつあることは、理論言語学にとって非常に意義のあることだと考えている。これまで文法システムの整備ばかりが注目されてきたが、実用に結びつけるためには、辞書の整備が不可欠である。本プロジェクトで進めている理論のように、文法システムと辞書の両方を持つものは、将来的にはコンピュータに実装しうる形式であり、AIへの貢献も十分に考えられる。

研究成果の概要(英文)：The aim of this project has been the construction of the data set which can evaluate the preciseness of the predictions made by a theory of grammar. We have a theory which can derive a semantic representation based on the syntactic structure, and we strived for identifying the crucial example sentences which will distinguish an adequate theory of grammar from inadequate ones. However, it turned out that a much more specific theory of argument structures and the interpretation of nominal expressions is necessary in order to provide a full range of predictions, we devoted to develop the description of the mental Lexicon so that we can proceed to the construction of the data set in the next step.

研究分野：理論言語学

キーワード：生成文法の経験的基盤 容認性判断調査 文法理論の精度の評価 統語意味論 動詞辞書

1. 研究開始当初の背景

(1) 言語学分野および工学分野での研究の現状

どのような単語列(いわゆる文)が提示されたときに、私たちがそれに対してどのような意味を与えるのか、という課題は、言語に関して、枠組をこえて共有される最も基本的な問題である。当然、言語学の文法理論においては、単語列と意味のペアを正しく予測できるかということが議論の対象となる。しかし、文によっては、その観察の妥当性のほうが問題視される場合も多く、そもそも、評価尺度となるデータセットが共有されていないのが現状である。

文における音と意味の対応関係は、単語における音と意味の対応関係とは比較にならないほど複雑である。特に問題になるのは、同じ文であっても、人によって・文脈によって、さまざまな意味解釈がありうるということであり、だからこそ、安易に「正しい文」と「間違った文」を列挙するわけにはいかない。どのようにデータセットを作成するかということこそ、これまでの統語論研究の知見の蓄積を生かして取り組むべき課題である。

昨今では、工学系の自然言語処理の分野において、人間が辞書を書かずとも単語を自動学習し、構文解析もビッグデータに基づいて処理できるシステムが開発されつつある。すでに、新聞などのコーパスに含まれる文章のほとんどについて処理結果が出力されるころまで来ているが、その出力結果は、人間の理解からは程遠いことも多いと聞いている。これも、文の解析結果の精度を評価するための尺度が存在していないことが問題である。

(2) 統語論と「意味」

従来の統語論では、意味の問題には立ち入らず、単に「文法的な単語列と非文法的な単語列を区別する」という課題を掲げている場合が多かった。しかし、特に日本語のような言語の場合、かなり破格の文であっても文脈によっては容認可能な場合が多々あり、形態論的な不具合を除いてしまうと、「真に非文法的な単語列」がそもそも存在するのかすら疑わしい。

しかし、近年、「**その単語列に対して、指定した解釈が可能かどうか**」に注目すれば、個人の中では「**可能性のあるもの**」と「**不可能なもの**」とが峻別できること、さらに、「**文S1でこのような解釈を許さない人は、必ず文S2においてその解釈を許さない**」というような**相関関係**に着目することで、個人差を越えて頑健(robust)な観察を取り出すことができるということが示された(cf. Hoji (2010) "Hypothesis Testing in Generative Grammar: Evaluation of Predicted Schematic Asymmetries," *Journal of Japanese Linguistics* 26, pp. 25-52, Hoji (2015) *Language Faculty Science*, Cambridge University Press.)。文の解釈全体を問題にしようとする観点が多すぎて判断がぶれてしまうので、文解析のカギとなる1点だけを指定するのである。その作業の遂行のために考案されてきた意味表示の理論が申請者が提案する**統語意味論**である。これを用いれば、統語関係に左右される意味の部分だけを取り出し、その文法理論の予測を計算によって導き出すことが可能になる。

2. 研究の目的

本プロジェクトでは、統語意味論により、意味解釈の予測をすることができるシステムを構築した上で、実際に利用が可能な形で「文法理論の精度を評価するためのデータセット」を提出することを目的とする。最初に取り組むのは、これまでの研究の蓄積が十分ある現象(照応関係、スコープ関係、否定呼応表現、省略構文、遊離数量詞構文)である。これらは、これまで理論言語学の中で扱われてきた現象であるため、必ずしも、コーパスなどの中で頻出する構文ではないが、だからこそ、工学分野での文解析器についても、その「応用力」を測ることができるものである。

3. 研究の方法

(1) 特色1：予測の質を評価するためのデータセット

従来、工学分野の文解析器については、実際に使用されている文に対して、量的にどれだけ解析結果を出力することができるかで評価されてきた。それに対して、本プロジェクトで提出するデータセットでは、その解析結果において、必ず達成されていなければならないポイントを指定することによって、その解析が「正解」かどうかを判断することができる。特に、指定された解釈が**不可能であると解析できなければ「正解」とならない**文も含まれているため、量と質の両方向から解析器を評価することができるようになる。

(2) 特色2：相関関係についての観察

さらに、前ページでも述べたように、文の解釈可能性には個人差が大きく、たいていの場合、「指定された解釈ができる人もいれば、できない人もいる」という結果になりがちである。しかし、そこで**相関関係**に注目することによって、「必ず成り立つ関係」が抽出できる。

このような観察を可能にしているのが、申請者が中心となって構築した**EPSA (Evaluation of Predicted Schematic Asymmetry) システム**と呼ばれる仕組みである。基本的には、指定された解釈が可能かどうか、話者が判断しWeb上で入力するアンケートのシステムであるが、特徴的なのは、各話者ごとにcodenameが割り当てられており、同一話者が他の文に対してどのような判断をしたかが追跡調査できるようになっている点、および、各例文が互いに綿密に関係づけられている点である。このシステムのおかげで、様々な相関関係を割り出すことが可能になっている。

(3) 予想される結果と意義

これまでも、EPSAシステムでは、頑健 (robust) な観察がいろいろ発見されてきたが、その結果は、生成文法理論内の用語で述べられていたため、限られた研究者にしか、その成果は知られていない。本プロジェクトでは、その知見をデータセットという形に集約させることによって、より広い範囲の研究者に、この研究の蓄積を伝えようとするものである。

言語というものは実に複雑な体系なので、本来、歴史的な変遷や地理的な変異、他言語との接触による影響など、様々な面からの考察が不可欠である。従来は、形式的な理論を用いる研究と記述を中心とした研究は、なかなか実質的な交流が難しかったが、データセットという具体的な形にすることによって、文解析器を作っている工学系の研究者のみならず、言語の記述的な研究を行なっている研究者とも、深いレベルでの意見の交換が期待できる。

また、この取り組みは、統語論の理論的研究そのものにとっても意義のあるものである。従来、生成文法では、「音と意味の結びつきについてモデルを作り、理論的に検証していく」という考え方は述べられてきたものの、個々の論文では、ごく一部の現象の特定の側面だけが取り上げられるため、最終的に全体としてどのようなシステムが作られつつあるのか、誰にもよくわからない状況になってしまった。体系的にデータセットを作っていくためには、語彙部門での規定やそれぞれの規則の相互作用を十分に計算する必要があり、その作業を通じて、統語論の理論の全体像を再構築することができるだろう。人間の言語能力は、未知の語彙や構文にも対応できる可変性を持ちながら、人間言語としての普遍性も持った奥深いシステムである。本プロジェクトは、そのような人間の言語能力の本質に理論的にアプローチする取り組みの一環である。

4. 研究成果

本プロジェクトは、文法理論の精度を評価するためのデータセットの作成を目標とするものであった。文法性の判断は個人差が大きいため、従来、このようなデータセットの作成は現実味が乏しいと考えられてきたが、本プロジェクトでは、文法構造と解釈の関係についての明確な予測体系を背景に、頑健なデータをよりわけ、より広い範囲の研究者にその成果を供することを目指していた。ただし、研究を進める中で、十分な予測を成り立たせるためには、動詞の項構造および名詞の解釈について、より踏み込んだ理論化が必要であるということが明らかとなり、データセット作成のためのレキシコン（脳内辞書）の整備を精力的に行なった。

結果としては、期間中に当初の目標を達成することができなかったが、動詞の項構造をどのように記述し、それがどのように文構造に反映されるかというプロセスが具体的になりつつあることは、理論言語学にとって非常に意義のあることだと考えている。本プロジェクトで進めている理論は、将来的にはコンピュータに実装しうる形式であり、辞書がさらに整備されていけば、AI への貢献も十分に考えられる。

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 5件／うち国際共著 5件／うちオープンアクセス 6件）

1. 著者名 Hayashishita, J.-R., Daiki Tanaka, & Ayumi Ueyama	4. 巻 36-1
2. 論文標題 A viable way of introducing Japanese verbs to second language learners	5. 発行年 2020年
3. 雑誌名 Journal of Japanese Linguistics	6. 最初と最後の頁 29-72
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 上山あゆみ, 藏藤健夫, 井本亮	4. 巻 19
2. 論文標題 日本語文法学会の展望：理論的研究	5. 発行年 2019年
3. 雑誌名 日本語文法	6. 最初と最後の頁 98-104
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 上山あゆみ, 傍土元	4. 巻 17-2
2. 論文標題 容認可能性と言語理論の説明対象	5. 発行年 2017年
3. 雑誌名 日本語文法	6. 最初と最後の頁 20-36
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 Ueyama, Ayumi & J.-R. Hayashishita	4. 巻 印刷中
2. 論文標題 Sentence Structure and Quantifier Scope in Japanese: A retrospective and Reanalysis	5. 発行年 2020年
3. 雑誌名 Mouton Handbooks on Japanese Language and Linguistics, Handbook of Japanese Semantics and Pragmatics	6. 最初と最後の頁 267-308
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 田中大輝, 林下淳一, 上山あゆみ	4. 巻 32
2. 論文標題 正しく使えるための日本語動詞学習辞典の作成に向けて	5. 発行年 2017年
3. 雑誌名 鳴門教育大学研究紀要	6. 最初と最後の頁 333-347
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 上山あゆみ	4. 巻 20-2
2. 論文標題 これからの生成文法研究: 20世紀の生成文法研究の遺産をどう生かすべきか	5. 発行年 2020年
3. 雑誌名 日本語文法	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Ueyama, Ayumi, J.-R. Hayashishita, and Iwao Takai	4. 巻 117
2. 論文標題 A review of the arguments for the direct/indirect passive distinction	5. 発行年 2020年
3. 雑誌名 文学研究	6. 最初と最後の頁 1-19
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

〔学会発表〕 計13件 (うち招待講演 5件 / うち国際学会 3件)

1. 発表者名 上山あゆみ
2. 発表標題 意味と構造を同時に決めるシステムを目指して
3. 学会等名 関東日本語談話会ワークショップ (招待講演)
4. 発表年 2018年

1. 発表者名 Ayumi Ueyama & J.-R. Hayashishita
2. 発表標題 Incorporating semantic features into Merge: A case study of Japanese
3. 学会等名 Linguistics Colloquium, University of Venice, Italy (招待講演) (国際学会)
4. 発表年 2017年

1. 発表者名 宮岡 大・内野宮 美月・芳賀 湧・吉野 雅耶・井上 美由紀・立山 芽衣・前野 一喜・山本 菜月・楊 佳銘・上山あゆみ
2. 発表標題 助動詞ダロウのアノテーションとガイドライン
3. 学会等名 言語処理学会 第24回年次大会
4. 発表年 2018年

1. 発表者名 松岡 葵・原 縁・宮本 ひかる・中村 允紀・岩下 千咲・栗林 真生・三井 桃子・村田 葵・上山あゆみ
2. 発表標題 文末表現ジャンイのアノテーションとガイドライン
3. 学会等名 言語処理学会 第24回年次大会
4. 発表年 2018年

1. 発表者名 上山あゆみ
2. 発表標題 容認可能性と言語理論の説明対象
3. 学会等名 日本語文法学会 シンポジウム (招待講演)
4. 発表年 2016年

1. 発表者名 宮岡大、上山あゆみ
2. 発表標題 多品詞語カギリのアノテーションとガイドライン
3. 学会等名 言語処理学会
4. 発表年 2017年

1. 発表者名 上山あゆみ
2. 発表標題 日本語統語論：構造構築と意味
3. 学会等名 上山あゆみさんに『統語意味論』の手ほどきをしてもらう会（リベンジ編）（招待講演）
4. 発表年 2016年

1. 発表者名 上山あゆみ、東寺祐亮、張晨迪
2. 発表標題 長距離依存と統語意味論
3. 学会等名 上山あゆみさんに『統語意味論』の手ほどきをしてもらう会（応用編）（招待講演）
4. 発表年 2016年

1. 発表者名 Hayashishita, J.-R., Ayumi Ueyama
2. 発表標題 A theory of character representations based on layered objects and two types of the speaker's introspective judgment: quantifier scope in Japanese
3. 学会等名 Logic and Engineering of Natural Language Semantics 16 (LENLS16) (国際学会)
4. 発表年 2019年

1. 発表者名 上山あゆみ・林下淳一
2. 発表標題 間接受身文の単文構造分析
3. 学会等名 日本言語学会第159回大会発表
4. 発表年 2019年

1. 発表者名 Hayashishita, J.-R., Iwao Takai, Ayumi Ueyama
2. 発表標題 Two types of passives in Japanese
3. 学会等名 7th Japanese Linguistics Symposium, 2019.11.28-29, University of Auckland, New Zealand (国際学会)
4. 発表年 2019年

1. 発表者名 上野聡子・上山あゆみ
2. 発表標題 謙遜の相槌とほめの対象
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 井上恵利佳・原田祐介・上山あゆみ
2. 発表標題 機能による相槌の分類
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

統語意味論デモプログラム (2017年度版)
<http://www.gges.org/syncsemphone/index.cgi>
統語意味論デモプログラム
<http://www.gges.org/syncsem/index.cgi>

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
--	---------------------------	-----------------------	----