

機関番号：32642

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500148

研究課題名（和文）確実性アノテーション：『確実性判断を表す意味的文脈』を記述したコーパスの構築

研究課題名（英文）Construction of an annotated corpus for certainty recognition

研究代表者

川添 愛（KAWAZOE AI）

津田塾大学・女性研究者支援センター・特任准教授

研究者番号：20450166

研究成果の概要（和文）：自然言語のテキストには事実のみではなく、書き手にとって真偽が不明な情報や、反事実的な仮定など偽であることが明らかな情報も含まれる。この研究では、機械による情報の確実性判断の基盤とするため、様相・条件・否定表現などの言語学的な分析に基づき、人間が普段情報の確実性を認識するのに利用しているテキストの意味特性をアノテーション（タグ付け）するスキーマを設計し、それに基づいてアノテーション済みコーパスを構築した。

研究成果の概要（英文）：Texts in natural language contain not only factual assertions but also uncertain information, such as speculations, inferences and hypothetical thoughts. With the aim of developing a general automatic system of certainty recognition, we designed an annotation schema to describe semantic features of modal, negative, conditional expressions and their scopes. We constructed an annotated corpus from Japanese news articles based on the schema.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,300,000	390,000	1,690,000
2009年度	1,100,000	330,000	1,430,000
2010年度	1,100,000	330,000	1,430,000
総計	3,500,000	1,050,000	4,550,000

研究分野：自然言語処理

科研費の分科・細目：情報学、知能情報学

キーワード：自然言語処理、アノテーション、理論言語学

## 1. 研究開始当初の背景

自然言語のテキスト中で記述される命題には、次の例に見られるように、事実（ここでは「書き手にとって真である命題」の意）以外にも、書き手にとって真偽が不明なものや、偽であることが明らかである命題がある。

(1) 昨日、県内で新型インフルエンザが発生した。（事実）

○ ○新聞が昨日県内で新型インフルエ

ンザが発生したと報じた。（伝聞）

(2) 昨日県内で新型インフルエンザが発生したようだ。 / (b) 昨日県内で新型インフルエンザが発生したらしい。 / (c) 昨日県内で新型インフルエンザが発生した可能性が高い。（推量）

(3) 県内で新型インフルエンザが発生した場合、どう対応するべきか。（仮定）

(4) 県内で新型インフルエンザが発生したということはない。 / (b) 県内で新型インフルエンザが発生したというのは誤報

だった。(否定)

研究開始当初は、上のような情報の確実性判断に関する自然言語処理分野の研究はまだ少なく、特に日本語についてはまだほとんど着手されていない状況であった。しかし、英語の医学・生物学テキストにおいては、推測や意見を事実の記述から区別する **hedge classification** の研究成果が徐々に増え始めていた。これらの研究の動機の多くは、それまでのキーワード・専門用語の認識による情報抽出では捉えきれない情報の確実性・信憑性に関する判断が自動的にできるようになれば、不要な情報を取り除いてより効率的な情報の獲得が可能になり、確実性判断に関わる人的コストを減らせるという点にあった。

## 2. 研究の目的

本研究では、そのような「確実性を自動的に認識するシステム」構築の基盤として、日本語のテキストに対し「確実性判断に関わる意味的文脈」をアノテーションしたコーパスを構築することを目的とした。具体的には、以下の成果物を得ることを計画した。

- (1) 確実性判断に関わる言語表現(様相・条件・否定表現)のリストと、その統語的・意味的特性に関する理論的な記述
- (2) アノテーション仕様
- (3) アノテーション済みコーパス(日本語のニュース記事を対象)

(1)と(2)については、日本語だけでなく、日本語といくぶん似た特性を持つと思われる韓国語の表現にも対象を広げることを計画した。

アノテーションスキーマおよびアノテーション済みコーパスの構築を計画するにあたっては、理論言語学における様相表現、否定表現・条件表現に関する研究の成果を応用するという点に重きを置いた。理論言語学の知見を用いることにより、テキストのスタイルや分野の違いに依存しない、より一般的に應用可能なシステムが得られると考えたためである。

## 3. 研究の方法

本研究では、1) 様相表現、否定表現、条件表現の振る舞いについての理論的研究と、2) それに基づく実テキストに対するアノテーション実験を、ほぼ並行しながら行った。これにより、双方に対するフィードバックを得て、理論とアノテーション仕様の両方を洗練

させていく効果を狙った。

## 4. 研究成果

20年度から21年度前半は、仮のアノテーションスキーマの設計とそれに基づくアノテーション実験を繰り返し、問題を解決しながらスキーマを洗練させた。特に20年度は、理論と応用の両面からの考察、および英語における先行研究の調査に基づき、アノテーションすべき表現を決定し、これらの表現の特性、特に表現のスコープの広さや表現間の埋め込みにおける意味の変化など、アノテーション仕様に含めるべき項目を列挙した。また、「確実性」という概念の定義など、アノテーションに関わる基本的事項の整備に重点を置き、アノテーションスキーマの第一版を作成した。21年度はさらに、「確実性に関わる文脈」を作り出す表現についての統語的・意味的記述を詳細化し、最終版に近いスキーマ第二版を構築した。このバージョンでは、様相表現、特にその一種である認識的推量表現の、言語学的テストに基づいた下位分類と、その成果を一部利用した条件表現の下位分類を追加した。また上記の理論的成果に基づき、アノテーション対象の命題を分類するオントロジーも新たに作成・追加している。

22年度は、前年度に改訂した「言語情報の確実性に影響する表現およびそのスコープのためのアノテーションガイドライン」に基づき、アノテーション済みコーパスを構築した。同コーパスは、日本語のニュース記事に対し、340種類以上の様相表現、13種類の否定表現、48種類の条件表現とそのスコープをアノテーションしている。また、アノテーション対象テキストの一部は、国立国語研究所の現代日本語書き言葉均衡コーパスのコアデータ(新聞記事)である。アノテーションガイドライン(バージョン2.4、平成23年3月時点での最新版)は、お茶の水女子大学情報科学科のテクニカルレポートとしてWeb上で公開している。

韓国語への応用については、21年度に構築した日本語の表現分類の基準に基づき、韓国語の様相表現の分類を行った。その成果を表した論文は学術誌に採択され、23年12月に掲載される予定である。

当プロジェクトの概要・成果については、人工知能学会および言語処理学会で発表を行った。言語処理学会では、研究代表者(川添)がテーマセッション「複合辞とモダリティ:理論から応用まで」の共同提案者となり、モダリティに関わる表現および認識手法に関心のある研究者とのディスカッションを

行った。また今年度は、話者の態度、信頼性、確実性のアノテーションに関して当プロジェクトと異なるアプローチを採用している研究者グループと連携し、定期的に意見交換をする場を設け、今後の研究協力の可能性を探った。

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 19 件)

- ① 齊藤学、崔榮殊、戸次大介、片岡喜代子、川添愛「言語情報の確実性アノテーションのための韓国語の様相表現の分類」、中華日本研究第3号、2011年12月発行予定(査読有)
- ② 川添愛、齊藤学、片岡喜代子、崔榮殊、戸次大介「様相・否定・条件表現の言語学的分析に基づく確実性アノテーションスキーマの設計」、言語処理学会第17回年次大会発表論文集、pp.143-136、2011(査読無)
- ③ 尾崎有梨・戸次大介「条件論理Cbとそのタブローシステム」、第13回プログラミングおよびプログラミング言語ワークショップ(PPL2011)論文集、pp.33-47、2011(査読有)
- ④ 川添愛・齊藤学・片岡喜代子・崔榮殊・戸次大介「言語情報の確実性に影響する表現およびそのスコープのためのアノテーションガイドライン Ver.2.4」、Technical Report of Department of Information Science, Ochanomizu University, OCHA-IS 10-4、2011(査読無) URL: [http://homepage2.nifty.com/bekki/lab/publications/BKKS\\_annotation%20guidelines\\_0304\\_11.pdf](http://homepage2.nifty.com/bekki/lab/publications/BKKS_annotation%20guidelines_0304_11.pdf)
- ⑤ 片岡喜代子「否定極性と統語的・意味的条件—日本語記述に基づくスペイン語否定現象再考—」、Hispanica54、pp.43-65、2011(査読有)
- ⑥ 川添愛・齊藤学・片岡喜代子・崔榮殊・戸次大介「言語情報の確実性アノテーションのための様相表現の分類」九州大学言語学論集 31、pp.109-129、2010(査読有)
- ⑦ Bekki, Daisuke “Combinatory
- Categorial Grammar as a Substructural Logic --- Preliminary Remarks,” Proceedings of the Seventh International Workshop on Logic and Engineering of Natural Language Semantics (LENLS 7), pp.70-83, 2010. (査読有)
- ⑧ Bekki, Daisuke “Representing Covert Movements by Delimited Continuations (extended version),” Kumiyo NAKAKOJI, Yohei MURAKAMI, Eric MCCREARY (Eds.), New Frontiers in Artificial Intelligence (LNAI 6284), pp.161-180, 2010(査読有)
- ⑨ 川添愛、齊藤学、片岡喜代子、崔榮殊、戸次大介「様相・否定・条件表現の言語学的分析に基づく確実性アノテーションスキーマの設計」、人工知能学会第24回全国大会論文集 2C3-2、2010(査読無)
- ⑩ 片岡喜代子「「N一人」と「Nが一人も」」、KLS 29: Proceedings of the Thirty-third Annual Meeting 29、pp.12-22、2009(査読有)
- ⑪ 片岡喜代子・宮地朝子「日本語の「とりたて」と叙述、その構造条件」、日本語学会予稿集 139、pp.186-191、2009(査読有)
- ⑫ Bekki, Daisuke, Kenichi, Asai “Representing Covert Movements by Delimited Continuations,” Proceedings of the Sixth International Workshop on Logic and Engineering of Natural Language Semantics, pp.71-90, 2009(査読有)
- ⑬ Bekki, Daisuke “Monads and Meta-Lambda Calculus,” New Frontiers in Artificial Intelligence 5447, pp.193-208, 2009(査読有)
- ⑭ Ai Kawazoe (et al) “The development of a schema for semantic annotation: Gain brought by a formal ontological method,” Applied Ontology 4(1), pp.5-20, 2009(査読有)
- ⑮ 川添愛、齊藤学、片岡喜代子、戸次大介「確実性判断に関わる意味的文脈アノテーションの試み」、IPSJ SIG Technical Report 2009-NL-189, pp.77-84, 2009(査読

無)

- ⑯ 川添愛「言語処理から見た前提研究の可能性」、日本言語学会第136回大会予稿集、pp.384-389、2008(査読有)
- ⑰ 川添愛、ナイジェル・コリアー「感染症情報処理と否定現象との関わりー社会的なニーズと言語学の接点」、月刊『言語』2008年8月号、pp.38-45、2008(査読無)
- ⑱ 片岡喜代子「否定関連表現と前提」、日本言語学会第136回大会予稿集、pp.372-377、2008(査読有)
- ⑲ 戸次大介「日本語における前提概念の同定」、日本言語学会第136回大会予稿集、pp.366-371、2008(査読有)

[学会発表](計13件)

- ① 川添愛、齊藤学、片岡喜代子、崔榮殊、戸次大介「様相・条件・否定表現の言語学的分析に基づく確実性判断のためのアノテーション済みコーパスの構築」、言語処理学会第17回年次大会 テーマセッション「複合辞とモダリティ:理論から応用まで」、2011/3/8、豊橋技術科学大学。
- ② 尾崎有梨・戸次大介「条件論理Cbとそのタブローシステム」、第13回プログラミングおよびプログラミング言語ワークショップ(PPL2011)、2011/3/9-11、定山溪ビューホテル(北海道)。
- ③ 石下裕里・戸次大介「前提記述のための動的論理の証明論構築に向けて」、第13回プログラミングおよびプログラミング言語ワークショップ(PPL2011)、2011/3/9-11、定山溪ビューホテル(北海道)。
- ④ Bekki, Daisuke “Combinatory Categorical Grammar as a Substructural Logic ---Preliminary Remarks”, the Seventh International Workshop on Logic and Engineering of Natural Language Semantics (LENLS 7), 2010/11/19, Campus Innovation Center, Tokyo, Japan.
- ⑤ 川添愛、齊藤学、片岡喜代子、崔榮殊、戸次大介「様相・否定・条件表現の言語学的分析に基づく確実性アノテーションスキーマの設計」、人工知能学会第24

会全国大会、2010/6/10、長崎ブリックホール。

- ⑥ 片岡喜代子・宮地朝子「日本語の「とりたて」と叙述、その構造条件」、日本言語学会第139回大会、2009/11/28、神戸大学。
  - ⑦ 戸次大介「CGによる日本語文法記述の進捗と展望~活用体系・統語構造・意味合成~」、言語処理学会第16回年次大会、2010/3/11、東京大学。
  - ⑧ Bekki, Daisuke, Kenichi, Asai “Representing Covert Movements by Delimited Continuations,” the Sixth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS 6), 2009/11/19, Campus Innovation Center, Tokyo, Japan.
  - ⑨ 川添愛、齊藤学、片岡喜代子、戸次大介「確実性判断に関わる意味的文脈アノテーションの試み」、情報処理学会 第189回自然言語処理研究会・第93回情報学基礎研究会 合同研究発表会、2009/1/23、お茶の水女子大学。
  - ⑩ 片岡喜代子「「N一人」と「Nが一人も」」、関西言語学会第33回大会、2008/6/7、大阪樟蔭女子大学。
  - ⑪ 片岡喜代子「否定関連表現と前提」、日本言語学会第136回大会、2008/6/21、学習院大学。
  - ⑫ 川添愛「言語処理から見た前提研究の可能性」、日本言語学会第136回大会、2008/6/21、学習院大学。
  - ⑬ 戸次大介「日本語における前提概念の同定」、日本言語学会第136回大会、2008/6/21、学習院大学。
- [図書](計2件)
- ① 片岡喜代子「否定極性と統語的条件」、加藤泰彦、吉村あき子、今仁生美(編)『否定と言語理論』、pp.118-140、開拓社、2011年。
  - ② 戸次大介『日本語文法の形式理論ー活用体系・統語構造・意味合成ー』、くろしお出版、ページ数356、2010年。

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

ホームページ等 なし

## 6. 研究組織

### (1) 研究代表者

川添 愛 (KAWAZOE AI)  
津田塾大学・女性研究者支援センター・  
特任准教授  
研究者番号 : 20450166

### (2) 研究分担者

戸次 大介 (BEKKI DAISUKE)  
お茶の水女子大学・大学院人間文化創  
成科学研究科・准教授  
研究者番号 : 90431783

片岡 喜代子 (KATAOKA KIYOKO)  
九州大学・大学院人文科学府・  
専門研究員  
研究者番号 : 80462810

### (3) 連携研究者

なし

### (4) 研究協力者

齊藤 学 (SAITO MANABU)  
中華大学・応用日本語学系

崔榮殊 (CHOI YOUNGSOO)  
一橋大学大学院

