

黒橋 禎夫

京都大学 大学院情報学研究科  
教授

知識に基づく構造的言語処理の確立と知識インフラの構築

## § 1. 研究実施体制

### (1) 黒橋グループ

- ① 研究代表者: 黒橋 禎夫 (京都大学大学院情報学研究科、教授)
- ② 研究項目
  - ・ 大規模コーパスからの新たな意味フレーム構築手法の考案
  - ・ 大規模コーパスからの同義表現・基本事態対の獲得とクラウドソーシングによる精錬
  - ・ クラウドソーシングによる談話関係コーパスの構築と、談話関係解析システムプロトタイプ  
の構築
  - ・ 知識インフラ分析用コーパスの作成についての検討

### (2) 戸次グループ

- ① 主たる共同研究者: 戸次 大介 (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)
- ② 研究項目
  - ・ 日本語 CCG パーザの整備
  - ・ 依存型意味論に基づく注釈付与コーパスの作成
  - ・ 依存型意味論に基づく意味の理論

### (3) 乾グループ

- ① 主たる共同研究者: 乾 健太郎 (東北大学大学院情報科学研究科、教授)
- ② 研究項目
  - ・ 言明・知識間の論理関係の設計
  - ・ 大規模知識に基づく推論(1): 推論機構の問題点の洗い出し
  - ・ 大規模知識に基づく推論(2): 問題点への対処に向けたプロトタイプシステムの作成

## § 2. 研究実施の概要

テキストは、専門家によるデータの分析結果や解釈、ステークホルダーの批判・意見、種々の手続きやノウハウなどが表出されたものであり、人間の知識表現の根幹をなすものである。言語の計算機処理はウェブをはじめとする大規模テキストの活用によって長足の進歩を遂げつつあるが、本研究ではこれをさらに発展させ、知識に基づく頑健で高精度な構造的言語処理を実現し、これによって様々なテキストの横断的な関連付け、検索、比較を可能とする知識インフラを構築する。また、構築した注釈付与コーパス、辞書、言語解析システムの公開によって研究コミュニティによる一層の研究の加速を実現するとともに、これらの研究成果を企業のカスタマセンター業務等の社会の実問題に適用し、その有用性を評価する。今年度は本研究プロジェクトの開始年度として、各研究項目について以下の研究を実施した。

### 文の意味の表現・計算モデルの構築

従来の自然言語処理では、文章の意味を表現するための確立されたフレームワークは存在せず、述語項構造や一階述語論理式の利用に留まっている。これを打破する新たな意味の表現・計算モデルの構築を目指して、「組合せ範疇文法(CCG)に基づく日本語パーザ+依存型意味論の証明システムによる計算」という枠組みで研究を推進した。具体的には、日本語 CCG による日本語の敬語表現の分析、モダリティ表現を中心とした機能表現の分類とコーパスアノテーション、談話・時制・因果関係のアノテーションを行った。また、依存型意味論による一般化量化子の理論、および「フォーカス」「逆スコープ」等の言語現象を分析・計算する手法を確立した。

### 知識に基づく文脈解析の実現と因果関係知識の抽出

テキストの各論述から直接的に言明とその間の因果関係等の知識を抽出するためには、ウェブテキスト等から獲得する知識の精練、文間の省略照応関係、談話関係などの文脈解析の高精度化が必要となる。そのために、大規模コーパスから意味フレーム・クラスを獲得する新たな手法を考案し、200億語のウェブテキスト集合から、約1,700個の動詞に対して約62,000個の意味フレーム、840個の意味クラスを獲得した。この成果は、本研究領域のトップカンファレンスであるEACL2014,ACL2014に採択された。さらに、クラウドソーシングによる同義表現辞書、基本事態対知識の精練、談話関係コーパスの構築を行い、クラウドソーシングがこのような高度な言語資源の構築において十分に有効であることを明らかにした。

### テキスト横断的な知識の関係付けによる知識インフラの構築

テキストの各論述から抽出される知識は、様々な抽象度・粒度の知識がばらばらに混在したものであるため、これらの知識を相互に関連づける推論機構が必要となる。文書内・文書外の言明間に等価・一般化・粗粒化・不整合等の関係注釈の試作として、意味的關係を6種類の関係ラベル、19種類の理由ラベルと定義し、全部で約1,000文対にラベルを付与した。この試作を通じて、言明間関係を実テキストに付与する際の問題点と対策を整理した。また、大規模知識に基づく推論問題点の洗い出し、その知見に基づき、知識の獲得文脈を考慮した知識適用の枠組みを試験的に構築し、共参照解析テストセット(300問)を用いて評価したところ正解率が5%向上すること(60%→65%)を確認できた。本研究について言語処理学会第20回年次大会で行った発表が大会最優秀賞を受賞した。

### § 3. 成果発表等

#### (3-1) 原著論文発表

##### 論文詳細情報(国内)

A-1. 叢悠悠, 戸次大介. 限定継続を用いたフォーカスの分析と実装. 第 16 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(PPL2014)論文集, 阿蘇の司 ビラパークホテル (熊本県), (2014.3)

A-2. 山本華子, 戸次大介. Neighborhood-Sheaf Semantics を用いた一階述語条件論理の意味論. 第 16 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(PPL2014)論文集, 阿蘇の司 ビラパークホテル (熊本県), (2014.3)

B-1. 萩行正嗣, 河原大輔, 黒橋禎夫. 多様な文書の書き始めに対する意味関係タグ付きコーパスの構築とその分析. 自然言語処理, Vol.21, No.2, (2014.4, to appear).

C-1. 松林優一郎, 飯田龍, 笹野遼平, 横野光, 松吉俊, 藤田篤, 宮尾祐介, 乾健太郎. 日本語文章に対する述語項構造アノテーション仕様の考察. 自然言語処理, Vol.21 No.2, (2014.4, to appear).

C-2. Naoya Inoue and Kentaro Inui. ILP-based Inference for Cost-based Abduction on First-order Predicate Logic. Journal of Natural Language Processing, Vol.20, No.5, pp. 629-656, (2013.12).

##### 論文詳細情報(国際)

A-3. Ran Tian, Yusuke Miyao, Takuya Matsuzaki. Logical Inference on Dependency-based Compositional Semantics. In Proceedings of Association of Computational Linguistics (ACL), (2014.6, to appear)

A-4. Kaneko Kimi and Daisuke Bekki. Building a Japanese Corpus of Temporal-Causal-Discourse Structures Based on SDRT for Extracting Causal Relations. In Proceedings of the EACL-2014 Workshop on Computational Approaches to Causality in Language (CAtoCL), 26th April, Gothenburg, Sweden. (2014.4, to appear).

A-5. Bekki Daisuke. A Type-theoretic Approach to Double Negation Elimination in Anaphora. In Proceedings of the Tenth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS10), pp.56-68,

ISBN:978-4-915905-57-5 C3004 (JSAI), JSAI International Symposia on AI 2013, Tokyo, Japan. 2013.

A-6. Tanaka Ribeka, Yuki Nakano and Daisuke Bekki. Constructive Generalized Quantifiers Revisited. In Proceedings of the Tenth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS10), pp.69-78, ISBN:978-4-915905-57-5 C3004 (JSAI), JSAI International Symposia on AI 2013, Tokyo, Japan. 2013.

A-7. Yamamoto Hanako and Daisuke Bekki. First-order conditional logic and neighborhood-sheaf semantics for analysis of conditional sentences. In Proceedings of the Tenth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS10), pp.96-106, ISBN:978-4-915905-57-5 C3004 (JSAI), JSAI International Symposia on AI 2013, Tokyo, Japan. 2013.

B-2. Daisuke Kawahara, Daniel W. Peterson and Martha Palmer. A Step-wise Usage-based Method for Inducing Polysemy-aware Verb Classes. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014), Baltimore, USA, (2014.6, to appear).

B-3. Gongye Jin, Daisuke Kawahara and Sadao Kurohashi. A Framework for Compiling High Quality Knowledge Bases From Raw Corpora. Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC2014), Reykjavik, Iceland, (2014.5, to appear).

B-4. Tomohide Shibata, Shotaro Kohama and Sadao Kurohashi. A Large Scale Database of Strongly-related Events in Japanese. Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC2014), Reykjavik, Iceland, (2014.5, to appear).

B-5. Daisuke Kawahara and Martha Palmer. Single Classifier Approach for Verb Sense Disambiguation based on Generalized Features. Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC2014), Reykjavik, Iceland, (2014.5, to appear)

.

B-6. Daisuke Kawahara, Daniel W. Peterson, Octavian Popescu and Martha Palmer. Inducing Example-based Semantic Frames from a Massive Amount of Verb Uses. Proceedings of the 14th Conference of the European Chapter of the Association for

Computational Linguistics (EACL2014), Gothenburg, Sweden, (2014.4, to appear).

B-7. Tomohide Shibata, Yusuke Egashira and Sadao Kurohashi. Chat-like Conversational System based on Selection of Reply Generating Module with Reinforcement Learning. Proceedings of the 5th International Workshop Series on Spoken Dialog Systems (IWSDS2014, short paper), pp.124-129, Napa, California, US, (2014.1).

B-8. Masatsugu Hangyo, Daisuke Kawahara and Sadao Kurohashi. Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions. Proceedings of EMNLP 2013: Conference on Empirical Methods in Natural Language Processing, pp.924-934, (2013.10).

B-9. Ryohei Sasano, Daisuke Kawahara, Sadao Kurohashi and Manabu Okumura. Automatic Knowledge Acquisition for Case Alternation between the Passive and Active Voices in Japanese, Proceedings of EMNLP 2013: Conference on Empirical Methods in Natural Language Processing, pp.1213-1223, Seattle, USA, (2013.10).

B-10. Daisuke Kawahara, Keiji Shinzato, Tomohide Shibata and Sadao Kurohashi. Precise Information Retrieval Exploiting Predicate-Argument Structures. Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP2013), pp.37-45, Nagoya, Japan, (2013.10).

B-11. Mo Shen, Daisuke Kawahara, and Sadao Kurohashi. Chinese Word Segmentation by Mining Maximized Substrings. Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP2013), pp.171-179, Nagoya, Japan, (2013.10).

B-12. Gongye Jin, Daisuke Kawahara and Sadao Kurohashi. High Quality Dependency Selection from Automatic Parses. Proceedings of the 6th International Joint Conference on Natural Language Processing, pp.947-951, (2013.10).

C-3. Kazuya Narita, Junta Mizuno and Kentaro Inui. A Lexicon-based Investigation of Research Issues in Japanese Factuality Analysis. In Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013), pp.587-595, (2013.10)

C-4. Naoaki Okazaki, Keita Nabeshima, Kento Watanabe, Junta Mizuno and Kentaro

Inui. Extracting and Aggregating False Information from Microblogs. In Proceedings of the Workshop on Language Processing and Crisis Information 2013 (LPCI 2013), pp.36-43, (2013.10)