# 含意・矛盾認識のための事象間関係知識の整備

# A Database of Knowledge on Relations between Events for Recognizing Entailment and Contradiction in Text

松吉 俊 † 村上 浩司 † 松本 裕治 † 乾 健太郎 ‡
Suguru Matsuyoshi Koji Murakami Yuji Matsumoto Kentaro Inui

## 1. はじめに

ウェブ上には大量のテキスト情報が存在し、そこでは様々なトピックに関して多角的な意見が述べられている。あるトピックに関して、このようなテキスト情報集合の俯瞰図 (言論マップ) を作成するために、現在、我々は、類似や含意、対立など、言論間の論理的関係を解析する基盤技術の開発に取り組んでいる [8]。俯瞰図において個々の言論を他の言論との論理的関係性の中に相対的に位置付けることにより、情報利用者が各言論の信憑性を判断する作業を支援し、情報の偏りや思いこみによる誤信の可能性を抑えることを目指す。

言論間の論理的関係を解析するために必要となる関係知識は、大きく実体(モノ、体言) に関する関係知識と事象(コト、用言) に関する関係知識に分類される。本論文では、我々が整備した、後者の関係知識を含む述語項構造辞書について報告する。

## 2. 既存の事象間関係知識

日本語に関する大規模な事象間関係知識のデータベースとしては、日本語 WordNet[3]、大西らのデータベース [10]、竹内らの辞書 [11] などが挙げられるが、日本語 WordNet は現在利用可能ではない。この章では、上記残り 2 つの言語資源について説明する。

### 2.1 大西らのデータベース

大西ら [2, 10] は、岩波国語辞典 [9] の用言に対する語釈文を利用することにより、人手で事象間関係知識のデータベースを作成した。ここでは、表 1 の一番左の列に示される 10 種類の関係が定義されている。このデータベースに含まれる関係知識の例を以下に示す。

**同義 (言い換え)** 誰か A が 誰か B を取り押さえる → 誰か A が 誰か B をつかまえる

**反義語** 誰か が 何か をあける → 誰か が 何か を締める

**手段** 誰か が 何か を沸かす → 誰か が 何か に熱を加える

### 2.2 竹内らの辞書

竹内ら [11] は、語彙概念構造による動詞意味分析の枠組み [4] に基づいて、Lexeed[5] に存在する高頻度動詞約 4,000 語、約 7,000 語義に対して項構造と意味クラスを記述し、動詞項構造辞書を作成した。この辞書において、事象間の類義関係や反義関係を規定する意味クラス体系は 5 階層からなり、その最下層には約 1,000 の意味クラスが存在する。この辞書における意味クラスとそこに属する動詞の例を以下に示す。

表 1: 事象間関係知識の数

| 関係名 | 英訳 | 知識数 |
|---|---|---|
| 同義 (言い換え) | near synonym | 17,816 |
| 同義・上位 | hypernym | 11,487 |
| 反義語 | antonym | 540 |
| 前提条件 | presupposition | 3,037 |
| 結果 (状態) | effect | 2,163 |
| 付帯状況 | cooccur | 4,274 |
| 不可分 | inseparable | 174 |
| 原因・理由 | cause | 2 |
| 目的 | goal | 882 |
| 手段 | means | 5,532 |
| 計 | | 45,907 |

＜状態変化あり. 位置変化. 位置関係の変化（物理）. 抽出／埋没. 抽出＞

とりだす、抜粋する、つまむ、あばく、···

＜状態変化あり. 位置変化. 位置関係の変化（物理）. 抽出／埋没. 埋没＞ (上記のクラスと反義関係にあるクラス)

埋める、うずめる、もぐる、ひたる、···

## 3. 述語項構造辞書の整備

前章の 2 つの節で述べた 2 つの言語資源を統合・整理し、事象間関係知識を含む述語項構造辞書を編纂した。

### 3.1 辞書の仕様

この述語項構造辞書は 1 つの XML 形式のファイルであり、用言に対して、語義ごとに、以下に示す情報が記述されている。

**ID** この辞書における ID

**見出し語** ひらがなで記述された見出し語

**表記** 送り仮名のゆれなどを含む漢字表記やカタカナ表記

**意味クラス** 2.2 節参照

**出現頻度** SENSEVAL-2 コーパス [6] における出現頻度

**岩波国語辞典 ID** 岩波国語辞典 [9] における ID

**LexeedID** Lexeed[5] における ID

**項構造** 用言がとる必須項のリストであり、それぞれの項に対して、次のような情報が記述されている

†奈良先端科学技術大学院大学, Nara Institute of Science and Technology
‡独立行政法人 情報通信研究機構, National Institute of Information and Communications Technology

表 2: 編纂した辞書におけるエントリーの例

| ID | 04992 |
|---|---|
| 見出し語 | きゅうしゅうする |
| 表記 | 吸収する |
| 意味クラス | 状態変化あり. 位置変化. 位置関係の変化 (物理) . 吸入／排出. 吸入 |
| 出現頻度 | 0 |
| 岩波 ID | 0011538-0-0-0-0-x0 |
| LexeedID | 06027950-4 |
| が格 | 変項: 何か A; 深層格: causer; 項番号: 2; 例: 水が |
| を格 | 変項: 何か B; 深層格: 対象; 項番号: 1; 例: 二酸化炭素を |
| 関係知識 | 同義 (言い換え): ＜何か A＞が＜何か B ＞を＜何か A＞に吸い込む |
| 関係知識 | 同義 (言い換え): ＜何か A＞が＜何か B ＞を吸い取る |
| 関係知識 | 付帯状況: ＜何か A＞が＜何か B ＞を取り入れる |
| 関係知識 | 同義・上位: ＜何か A＞が＜何か B ＞を自分のものとする |

変項 関係知識において、後件の項と対応づけるために用いる

定項 格の前に現れる語が確定している場合に、その語を記述する

深層格 表層格に対する深層格

項番号 半統制意味構造記述 [11] における項番号

格の交替 「に」→「へ」のように、代わりに用いることができる格

例 項の例

事象間関係知識 2.1 節で述べた事象間関係知識を表す、関係と後件の述語項構造の対のリスト

この辞書におけるエントリーの例を表 2 に示す。

### 3.2 編纂手順

2 章で説明した 2 つの言語資源から、次の手順に従い、前節で述べた仕様を満たす述語項構造辞書を編纂した。

1. 岩波国語辞典の語義を 1 エントリーとして、大西らのデータベース (2.1 節) を XML 形式に変換する
2. 各エントリーに対して、SENSEVAL-2 コーパスにおける出現頻度を数える
3. 岩波国語辞典 ID を介して、各エントリーに対して、竹内らの辞書 (2.2 節) から次の情報を抽出する

   意味クラス、深層格、項番号、格の交替、LexeedID

   岩波国語辞典の 1 つの語義に対して、複数の Lexeed の語義が対応する場合、その数だけエントリーを複製し、そのそれぞれに対して、Lexeed の 1 つの語義に対する上記の情報を付与する

### 3.3 現状

現在、編纂した辞書には、29,555 エントリーが登録されており、計 45,907 個の事象間関係知識が含まれている。それぞれの関係に対する知識数を表 1 に示す。

この辞書に存在する事象間関係知識と意味クラスの情報を用いて、述語項構造レベルの言論マップを生成する予備調査を行なった。人手による正解セットと比較した結果、およそ 6 割の精度で言論間の類義関係と対立関係を認識することができた。この調査の詳細については別稿 [7] で述べる。

## 4. おわりに

本研究では、言論マップの作成に必要となる事象間関係知識を含む述語項構造辞書を整備した。この辞書を用いることにより、用言に関して、言論間の類義関係や対立関係を解析することが可能である。

今後は、下位事象を内包する構造を持つ用言について人手で事象間関係知識を整理するとともに、コーパスから獲得された大規模な事象間関係知識 [1] を適切に辞書に組み入れる予定である。

謝辞

本研究は、独立行政法人 情報通信研究機構 委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」の支援の下に実施した。

## 参考文献

[1] 阿部修也, 乾健太郎, 松本裕治. 文内共起パターンと格要素共有情報による事態間関係知識の獲得. 言語処理学会第 14 回年次大会発表論文集, pp. 797–800, 2008.

[2] 青山桜子, 阿部修也, 大西良明, 乾健太郎, 松本裕治. 事態間関係の獲得のための動詞語釈文の構造化. 言語処理学会 第 13 回年次大会発表論文集, pp. 286–289, 2007.

[3] Francis Bond, Hitoshi Isahara, Kyoko Kanzaki, and Kiyotaka Uchimoto. Boot-strapping a WordNet using multiple existing WordNets. In *Proceedings of the 6th International Language Resources and Evaluation (LREC2008)*, 2008.

[4] 影山太郎. 動詞の意味と構文. 大修館書店, 2001.

[5] 笠原要, 佐藤浩史, 田中貴秋, 藤田早苗, 金杉友子, 天野成昭. 「基本語意味データベース: Lexeed」の構築. 情報処理学会研究報告 2004-NL-159, pp. 75–82, 2004.

[6] 黒橋禎夫, 白井清昭. SENSEVAL-2 日本語タスク. 電子情報通信学会技術研究報告 NLC2001-36, pp. 1–8, 2001.

[7] 村上浩司, 松吉俊, 増田祥子, 松本裕治, 乾健太郎. 言論マップ生成のための事象間類似・対立関係の認識. 第 7 回情報科学技術フォーラム (FIT2008) 発表論文集, pp. –, 2008.

[8] 村上浩司, 松吉俊, 隅田飛鳥, 森田啓, 佐尾ちとせ, 増田祥子, 松本裕治, 乾健太郎. 言論マップ生成課題：言説間の類似・対立の構造を捉えるために. 情報処理学会研究報告 2008-NL-186, pp. –, 2008.

[9] 西尾実, 岩淵悦太郎, 水谷静夫（編）. 岩波国語辞典第五版. 岩波書店, 1994.

[10] 大西良明, 乾健太郎, 松本裕治. 事態間関係知識の整備と含意文生成への応用. 言語処理学会 第 14 回年次大会発表論文集, pp. 1152–1155, 2008.

[11] 竹内孔一, 乾健太郎, 竹内奈央, 藤田篤. 意味の包含関係に基づく動詞項構造の細分類. 言語処理学会 第 14 回年次大会発表論文集, pp. 1037–1040, 2008.