

事象間関係知識の整備と類似・対立認識への応用

松吉 俊[†], 村上 浩司[†], 増田 祥子^{†,††}, 松本 裕治[†], 乾 健太郎[‡]

[†] 奈良先端科学技術大学院大学, ^{††} 大阪府立大学

[‡] 独立行政法人 情報通信研究機構

{matuyosi, kmurakami, shouko, matsu, inui}@is.naist.jp

あるトピックについて書かれた大量のウェブ文書から、そのトピックについての深い理解を得るためには、それらの文書に記述されている個々の言論の信憑性や有効性を適切に判断し、かつ、暗に存在しているかもしれない言論間の対立関係を発見しようとする必要がある。このような作業の支援を目的として、我々は、言論間に存在する類義や対立、根拠などの論理的関係を解析し、それらを言論マップと呼ぶ俯瞰図で明示するための技術の開発に取り組んでいる。本論文では、言論マップの生成に必要な事象間関係知識の整備に関して、既存の2つの言語資源の統合と、下位事象を内包する動詞項構造の記述について報告する。予備実験の結果、整備した事象間関係知識データベースを用いると、約8割の高い精度で言論間の類義関係を認識できることが分かった。

An Application of a Knowledge Base on Relations between Events to Recognizing Textual Entailment and Contradiction

SUGURU MATSUYOSHI[†], KOJI MURAKAMI[†], SHOKO MASUDA^{†,††},
YUJI MATSUMOTO[†], KENTARO INUI[‡]

[†]Nara Institute of Science and Technology, ^{††}Osaka Prefecture University

[‡]National Institute of Information and Communications Technology

We present two language resources for recognizing relations between statements: a database of knowledge on relations between predicate argument structures (PASs) in Japanese and a dictionary of control verbs in Japanese. We have compiled the former database by combining an existing list of relations between two PASs, which were extracted manually from interpretations in a dictionary of Japanese words, with a thesaurus of verb argument structures proposed by Takeuchi et al. This database has 29,555 entries and 45,905 relations between PASs. In a preliminary experiment with this database, a system that recognizes synonymy relation between PASs in Web documents has a performance that its precision is about 80%.

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。情報検索技術の発展により、我々は、あるトピックについて書かれた文書を容易に入手できるようになった。しかしながら、このような情報源から、そのトピックについての深い理解を得ようとするときには、特に、次の2

点について注意する必要がある。

まず第1に、ウェブから入手した文書に記述されている情報は、そのすべてが真実であるとは限らない。そこには、不正確な記述、偏りのある意見、陳腐化した情報などが混在している可能性が非常に高い。そのような情報を含むページの中には、**妥当な根拠**や正確な出典を併記していないにもかかわらず、そこで主張されている命題がさも唯一の事実であるかのように、情報を記述する

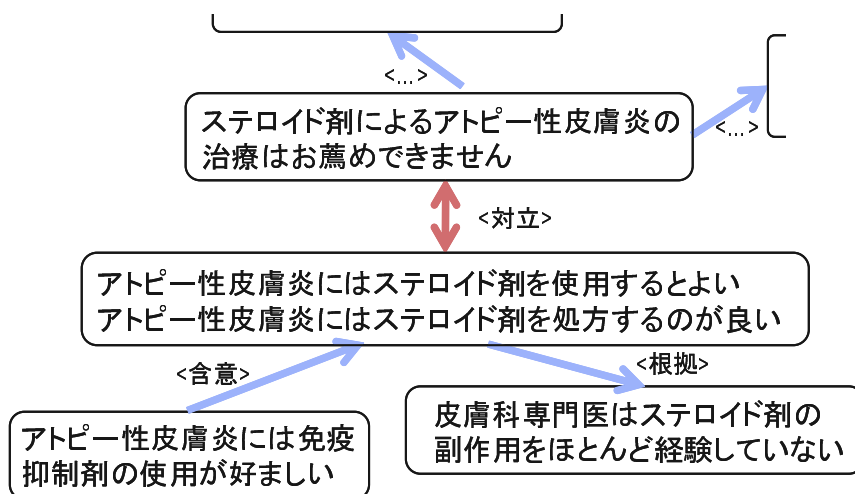


図1 トピック「ステロイド」に関する言論マップ

ページもある。第2に、ウェブは、ハイパーリンクという仕組みによって、情報が有機的に繋がらうという素晴らしい特性を持っているが、あるトピックについて何らかの意見を述べているページに、それと対立するような意見を述べるページへのリンクが必ず存在するとは限らない。例えば、ある化学成分を含む健康食品を肯定的に捉えるサイト(その主だったものは、そのような商品を扱う販売サイト)においては、専らその化学成分の肯定的な効用のみが謳われており、その効用を否定するような、自分たちに都合の悪い情報を記載しているページへのリンクは存在しない。明示的なリンクが存在しないため、このような肯定/否定の対立は、ウェブの読者にとって隠れた関係である。あるトピックについて情報検索システムを用いて検索したときに、得られた結果の上位に肯定的な意見を述べるサイトしか見られない場合、検索ユーザーは、そのトピックに関する否定的な意見は存在しないのだと誤信してしまう可能性がある。

これらの問題点を考慮すると、あるトピックについて書かれた大量のウェブ文書から、そのトピックについての深い理解を得るためには、ウェブの読者は、それらの文書に記述されている個々の言論の信憑性や有効性を適切に判断し、かつ、暗に存在しているかもしれない言論間の対立関係を発見することに努めなければならない。しかしながら、限られた時間で、各言論の信憑性を判断し、言論間の構造を明確に把握することは容易ではない。ウェブの読者がこれらを遂行するのを支援するシステムの実現が望まれる。

このような背景により、現在、我々は、言論間に存在する類義や対立、根拠など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。これらの論理的関係を、俯瞰図のような構造で示すことにより、ウェブの読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している。

論理的関係によって関係付けられた言論の集合は、それぞれの言論をノード、論理的関係をエッジとしたときに1つのグラフを形成する。我々はこのグラフを言論マップと呼ぶ。この言論マップの特長は、個々の言論が他の言論との論理的関係性の中に相対的に位置付けられる点であり、これにより、各言論の信憑性や立場が判断しやすくなると思われる。例として、トピック「ステロイド」に関する言論マップを図1に示す。この図には、以下のような、言論間の論理的関係が含まれている。

- (言論1) アトピー性皮膚炎にはステロイド剤を使用するとよい
- (言論1) ← 類義 ⇒ アトピー性皮膚炎にはステロイド剤を処方するのが良い
- (言論1) ← 含意 ⇒ アトピー性皮膚炎には免疫抑制剤の使用が好ましい
- (言論1) = 根拠 ⇒ 皮膚科専門医はステロイド剤の副作用をほとんど経験していない
- (言論1) ← 対立 ⇒ ステロイド剤によるアトピー性皮膚炎の治療はお薦めできません

図1を見ると、(言論1)に対してその根拠となる事実が存在するため、この言論の信憑性は高いと言えるが、それに対する対立意見も存在することが分かる。このような言論マップは、ウェブ文書における各言論の信憑性を判断し、言論間の構造を明確に把握しようとする際に有用であると思われる。

我々は、前稿¹²⁾において、ウェブ文書集合から言論マップを生成するために達成すべき課題について論じ、それぞれの課題に対しての我々の取り組みについて報告した。これらの課題は、大きく、知識整備に関するものと解析技術の開発に関するものに分けることができる。前者が対象とする、言論間の論理的関係を解析するために必要な関係知識は、主として、**実体**(モノ、体言)に関するものと**事象**(コト、用言)に関するものに分類される。本

論文では、後者の事象間関係知識に関して、我々が行なった既存の言語資源の整備と、下位事象を内包する動詞項構造の記述について報告する。

本論文は、次のように構成される。まず、2章で関連研究について述べる。次に、3章において、既存の2つの言語資源を整備することにより作成した事象間関係知識データベースについて説明する。4章で、下位事象を内包する構造を持つ用言の分類と、それに基づいて行なった、下位事象を内包する動詞項構造の記述について述べる。5章において、3章で作成したデータベースの小規模な評価実験について報告する。この実験では、データベースに含まれる事象間関係知識を、述語項構造レベルの言論間類義・対立関係認識に応用した。最後に、6章で全体をまとめる。

2. 関連研究

言論間の論理的関係解析を行なうには、実体間の関係知識だけではなく、事象間の関係知識を利用することが不可欠である。英語においては、WordNet⁷⁾、FrameNet³⁾、VerbOcean⁵⁾など、事象間関係知識の大規模なデータベースが存在する。これらの言語資源は、事象間の類義関係、上位下位関係、反義関係等を扱っており、近年研究者の関心を集めている含意関係認識 (Recognizing Textual Entailment)⁶⁾の研究において、しばしば利用されている。

日本語においても、事象間関係知識の大規模なデータベース構築に関する研究が行なわれており、人手で関係知識を整理した言語資源として、日本語 WordNet⁴⁾ や大西らの動詞語釈文構造化データ¹⁴⁾などが存在し、大規模コーパスから関係知識を獲得する手法として、Inuiら⁸⁾、Torisawa¹⁸⁾、阿部ら¹⁾の研究などがある。

竹内ら¹⁷⁾は、語彙概念構造による動詞意味分析の枠組み⁹⁾に基づく動詞項構造ソーラスを構築した。後で3.1.2節で見ると、このソーラスは、動詞間の類義関係知識と反義関係知識を提供することができる特長を持っている。彼らのソーラスを、現在我々が利用可能である大西らの動詞語釈文構造化データと統合すると、言論マップの生成に非常に有用である言語資源を構築することができる。この既存の2つの言語資源の統合については、次の3章で詳説する。

英語や日本語には、コントロール構文¹⁶⁾と呼ばれる構文が存在する。例えば、日本語における次の例文(2a)はコントロール構文であり、補文「協力する」の主語は、主節の目的語によって義務的にコントロールを受ける。すなわち、例文(2a)において、「協力する」の主語は、「関係省庁」と解釈され、それ以外の解釈は不可能である。(2a) 政府が関係省庁に協力することを要請した。動詞「要請する」がこのような構文をとるという知識が利用可能であれば、計算機は、例文(2a)が「関係省庁が協力する」という言論を内包していることを認識することができる。言論間の論理的関係認識においては、ある

言論が別の言論を内包していることを認識する技術は必要不可欠であり、それゆえに、補文を持つ用言の構文構造の特徴を把握し、それを知識として蓄えておくことは重要である。日本語に関して、このような知識を集積した言語資源は存在しないため、我々は、このような知識を整理し、個々の動詞にこの情報を付与した。構文構造の情報付与に関しては、4章で詳しく述べる。ガ格が省略された補文においてその主語を同定するという問題は、省略解析が扱う主要問題の1つであり、我々が付与したこの情報は、省略解析においても有用であると言える。

上記の例文(2a)は、次の例文(2b)と同義である。

(2b) 政府が関係省庁に協力を要請した。

構文構造に関する我々のデータベースには、このような「構文構造のゆれ」の情報も記載されており、名詞化された述語「協力」の主語が主節の二格の「関係省庁」であるという知識を外部に提供することができる。名詞化された述語の項構造を同定するという問題は、述語項構造解析において取り扱われている難しい問題であり、我々のデータベースは、その一助になると思われる。

3. 事象に関する既存の2つの言語資源の整備

前章で述べた次の2つの言語資源を整備することにより、言論マップの生成に利用することができる事象間関係知識のデータベースを作成した。

(1) 大西らの動詞語釈文構造化データ

(2) 竹内らの動詞項構造ソーラス

本章では、まず、これらの言語資源について概説する。次に、我々が実際に行なった、これらの言語資源を計算機で利用しやすい形に変換し統合する方法について説明する。そして、作成した事象間関係知識データベースの現状について報告する。

3.1 既存の2つの言語資源

3.1.1 大西らの動詞語釈文構造化データ

大西ら^{2),14)}は、岩波国語辞典¹³⁾における動詞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リストを作成した。例えば、動詞「倒す」にはいくつかの語義が存在するが、そのうちの1つの語義に対して、岩波国語辞典では次に示す語釈文が記述されている。

- 立っている物に力を加えて傾け、横にする。

この語釈文には、「立つ」、「力を加える」、「傾ける」、「横にする」という4つの事象が含まれている。大西らは、見出しである「倒す」とこれらの事象の間に存在する関係を、以下のような関係知識として収集した。

(見出し事象) 誰かが何かを倒す

前提条件 ⇒ 何かが立つ

手段 ⇒ 誰かが何かに力を加える

手段 ⇒ 誰かが何かを傾ける

同義・上位 ⇒ 誰かが何かを横にする

ここで、下線を引いた語は、関係知識における変項であり、計算機による利用上は、X、Y、Zなどの記号と同じ

表 1 意味クラスとそこに属する動詞の例

大分類 1	大分類 2	中分類	小分類 1	小分類 2	属する動詞
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	吸う、吸収する、吹き込む、吸い取る
状態変化あり	位置変化	位置関係の変化 (物理)	吸入/排出	排出	吐く、発散する、漏れる、放流する
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	埋める、うずめる、もぐる、ひたる
状態変化あり	位置変化	位置変化 (物理)	起点からの移動	-	走る、下りる、転居する、飛び出す

機能のみを持つ。関係知識の読みやすさを考慮して、我々は、変項を表す文字列として、「誰か」や「何か」などを用いている。ここで注意すべきは、「誰か」や「何か」などの文字列の使い分けは、格要素に対する選択制限を表しているわけではないということである。その使い分けは、純粹に関係知識の読みやすさのみ依存しており、例えば、上記の関係知識の変項「誰か」には、人間を指し示す名詞しか入らないわけではなく、ブルトナーや鉄球などの名詞も入れることが可能である。

大西らは、岩波国語辞典に収録されている動詞とサ変名詞約 1 万 4 千語 (約 2 万 4 千語義) に対して上記の作業を行ない、約 3 万 4 千個の事象間関係知識を収集した。彼らが知識の収集に用いた 9 種類の事象間関係とそれに対する事象間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

同義・上位 誰か が 何か を使いこなす ⇒ 誰か が 何か を使う

反義語 誰か が 何か をあける ⇔ 誰か が 何か を締める

前提条件 誰か が 何か を言い逃れる ⇒ 誰か が 何か を問いつめられる

結果 (状態) 誰か が 何か を起こす ⇒ 何か が直立する

付帯状況 何か が流行る ⇒ 何か が次々と伝わる

不可分 誰か が 何か を上げ下げする ⇒ 誰か が 何か を上げる

目的 誰か が 何か A を 何か B にメモする ⇒ 誰か が 何か A を忘れない

手段 誰か が 何か を沸かす ⇒ 誰か が 何か に熱を加える

3.1.2 竹内らの動詞項構造シソーラス

竹内ら¹⁷⁾は、語彙概念構造による動詞意味分析の枠組み⁹⁾に基づいて、Lexeed¹⁰⁾に存在する高頻度動詞約 4 千語 (約 7 千語義) を対象とした動詞項構造シソーラスを構築した。このシソーラスの基本単位は動詞の語義であり、各エントリーに対して、項構造、深層格、意味クラス、例文などの情報が記載されている。

彼らのシソーラスの意味クラス体系は 5 階層からなっており、最上層は「状態変化あり」、「状態変化なし (活動)」、「状態変化なし (状態)」の 3 クラスであり、最下層は約千のクラスに分類されている。このシソーラスの意味クラスとそこに属する動詞の例を表 1 に示す。このシソーラスにおいて、最下層の同じクラスに属する動詞は、互いに類義関係にあると見ることが出来る。例えば、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」に属する動詞「吸う」と「吸収する」は、類義関係にある語対として利用することができる。

シソーラスの構築時に、竹内らは、「ある語とその反義語には、何らかの共通する概念が存在する」という観点のもと、そのような概念を小分類 1 における“/”を含むクラス名で表し、ある語とその反義語を小分類 2 において分類している。表 1 の 1 つめと 2 つめのクラス間、および、3 つめと 4 つめのクラス間に、この分類法に基づく反義語の分類例が示されている。例えば、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」は、その中に反義語を内包する共通概念であり、その下位に存在するクラス「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」に属する動詞「吸う」は、それと異なる下位クラス「状態変化あり. 位置変化. 位置関係の変化 (物理). 吸入/排出. 排出」に属する動詞「吐く」と反義関係にあると見ることが出来る。

上の 2 つの特徴をまとめると、竹内らの動詞項構造シソーラスは、動詞間の類義関係知識と反義関係知識を提供することができると言える。これらの関係知識は、言論マップの生成に大いに利用することができる。

竹内らは、半統制意味構造記述を導入することにより、シソーラスの分類間の横断的な対応関係を記述する方法を提案している。この方法の詳細については、竹内らの論文¹⁷⁾を参照されたい。

3.2 事象間関係知識データベースへの統合

前の 2 つの節で説明した 2 つの言語資源を統合・整理し、言論マップ生成に直接利用することができる事象間関係知識のデータベースを編纂した。このデータベースの編纂手順を以下に示す。

- (1) 岩波国語辞典の 1 つの語義を 1 エントリーとして、大西らの動詞語釈文構造化データ (3.1.1 節) を XML 形式に変換する
- (2) SENSEVAL-2 コーパス¹¹⁾における、関係知識の前件の用言と後件の用言の出現頻度を数える
- (3) 岩波国語辞典 ID を介して、各エントリーに対して、竹内らの動詞項構造シソーラス (3.1.2 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、Lexeed ID

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

表 2 事象間関係知識データベース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
岩波国語辞典 ID	0011538-0-0-0-0-x0
Lexeed ID	06027950-4
ガ格	変項: 何か A; 深層格: causer; 項番号: 2; 例: 水が
ヲ格	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

の語義に対する上記の情報を付与する

上記の手順 (2) において、関係知識に含まれる用言の出現頻度を取得した理由は、これらの出現頻度を、関係知識集合のフィルタリングに利用することを想定しているからである。例えば、応用システムに出力する関係知識の絶対数を減らしたいという状況においては、前件の用言も後件の用言もコーパスに全く出現しないような関係知識から優先的に取り除くことが考えられる。出現頻度を数えるコーパスとして SENSEVAL-2 コーパスを利用した理由は、このコーパスには、岩波国語辞典の語義の情報がタグ付けされているからである。

3.3 事象間関係知識データベースの現状

我々が作成した事象間関係知識データベースは、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 データベース 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 竹内らの動詞項構造シソーラスにおける意味クラス

出現頻度 SENSEVAL-2 コーパスにおける出現頻度

岩波国語辞典 ID 岩波国語辞典における ID

Lexeed ID Lexeed 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

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文字列

定項 格の前に現れる語が確定している場合に、変項の文字列の代わりに記述される語

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

関係知識 関係名と後件の述語項構造の対のリストで表現される、大西らの事象間関係知識

このデータベースにおけるエントリーの例を表 2 に示す。現在の事象間関係知識データベースには、動詞の語義

表 3 事象間関係知識データベースに含まれる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
目的	goal	882
手段	means	5,532
計		45,905

単位で 29,555 エントリーが登録されており、計 45,905 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関係知識の数を表 3 に示す。

予備実験として、整備した事象間関係知識データベースを、述語項構造レベルの言論問類義・対立関係認識に応用し、データベースの小規模な評価を行なった。この実験の詳細については、5 章で述べる。

4. 下位事象を内包する動詞項構造の記述

2 章で述べたように、言論間の論理的関係を認識する過程において、補文を持つ用言の構文構造を把握することは重要である。本論文では、補文を持つ主節の用言が表す事象を**上位事象**、補文の用言や補文と同等の構造を持つ名詞句が表す事象を**下位事象**と呼ぶ。我々は、下位事象の標識と下位事象の主格を表す表層格に基づいて、下位事象を内包する構造を持つ用言 (上位事象の用言) を 11 の型に分類した。

以下、本章では、まず、下位事象を内包する構造が持つ 3 つの特徴について述べる。次に、下位事象の標識と下位事象の主格を表す表層格に基づく、上位事象の用言の分類について説明する。そして、最後に、この分類体系に基づいて行なった、下位事象を内包する動詞項構造の記述について報告する。

4.1 下位事象を内包する構造が持つ 3 つの特徴

下位事象を内包する構造は、上位事象の用言と下位事象の標識に関して、次の 3 つの特徴を持つ。

A. 下位事象を内包する構造を持つ用言のほとんどが、

表 4 下位事象を内包する構造を持つ用言 (上位事象の用言) の分類

分類型	ことを /のを	を	ことに /のに	に	-を-と	(基本形)と	(意向形)と	(命令形)と	ように/ ようにと
決意型	上位のガ	上位のガ	-	-	-	上位のガ	上位のガ	-	-
約束型	上位のガ	上位のガ	-	-	-	上位のガ	上位のガ+	-	-
命令型	ニ	ニ	-	-	-	-	-	ニ	ニ
言う型	-	-	-	-	-	下位のガ	上位のガ+	ニ	ニ
許可型	ニ、下位のガ	ニ、ノ	-	-	-	-	-	-	-
提案型	下位のガ	ノ	-	-	-	-	上位のガ+	-	ニ
発言型	下位のガ	ノ	-	-	-	下位のガ	上位のガ+	-	-
感覚型	下位のガ	ノ	-	-	ヲ	下位のガ	上位のガ	-	-
認識型	下位のガ	ノ	-	-	-	下位のガ	-	-	-
感情型	-	-	下位のガ	ノ	-	下位のガ	-	-	-
態度型	下位のガ	ノ	-	-	-	-	-	-	-

複数の下位事象の標識をとることができる

- B. 一般に、上位事象の用言に対して、下位事象の主格を表す表層格は、下位事象の標識により異なる
- C. 下位事象の標識に対して、下位事象の主格を表す表層格が一意に定まっているわけではない

上記の A. に関して、例えば、動詞「命令する」は、次の例文に示されるように、さまざまな下位事象の標識をとることができる。(以下、例文中の下位事象の標識に下線を引く。)

- (3a) 部長が部下にノルマを達成する ことを 命令した。
- (3b) 部長が部下にノルマの達成 を 命令した。
- (3c) 部長が部下にノルマを達成 しろと 命令した。
- (3d) 部長が部下にノルマを達成する ように 命令した。
- (3e) 部長が部下にノルマを達成する ようにと 命令した。

上記の B. に関して、例えば、動詞「言う」を上位事象に持つ以下の例文において、下位事象の主格を表す表層格は、すべて異なる。

- (4a) 太郎が花子に来月京都に 行けと 言った。
- (4b) 太郎が花子に来月京都に 行こうと 言った。
- (4c) 太郎が花子に弟が来月京都に 行くと 言った。
- これらの例文において、下位事象の標識はすべて異なっており、「来月京都に行く」のは、順に、花子、太郎(と花子)、弟であり、その主格を表す表層格は、それぞれ、二格、上位のガ格(と二格)、下位のガ格である。

上記の C. に関して、例えば、下位事象の標識「ことを」を含む以下の例文において、下位事象の主格を表す表層格は、すべて異なる。

- (5a) 太郎が編集者に締め切りまでに原稿を完成させる ことを 約束した。
- (5b) 次郎がカメラマンに作品を撮影する ことを 許可した。
- (5c) 三郎が花子に友達が大学に合格した ことを 報告した。

これらの例文は、順に、主語コントロール構文、目的語コントロール構文[☆]、非コントロール構文と呼ばれるものであり、下位事象の主格を表す表層格は、それぞれ、上

位の下位のガ格、二格、下位のガ格である。

一方、これまでの例文から、上位事象の用言と下位事象の標識が与えられると、下位事象の主格を表す表層格が一意に定まることが推測できる。

4.2 下位事象を内包する構造を持つ用言の分類

前節で述べた3つの特徴を考慮し、「ある下位事象の標識をとったときに、下位事象の主格を表す表層格が何になるのか」に基づいて、下位事象を内包する構造を持つ用言を11の型に分類した。この分類を表4に示す。ここで、「上位のガ+」は、「上位のガ格(と二格や聞き手)」を意味する。許可型の欄における読点は、上位事象が二格を持つときはその二格が、そうでないときは読点の右に示されている格が、下位事象の主格を表す表層格であることを示す。

4.3 分類に基づく記述

我々は、表4の分類体系に従い、岩波国語辞典に存在する動詞979語(1,161語義)に対して、下位事象を含む例文を記述するとともに、その分類型を付与した。

対象とした動詞リストは、3章で説明した事象間関係知識データベースから、下位事象を内包する構造を持っている可能性が高いと思われる動詞の意味クラスを指定して抽出した。分類型付与においては、語義単位の各動詞に対して、以下の作業を行なった。

- (1) 下位事象を内包する構造を持つ動詞であるかどうか判断する
- (2) 下位事象を含む例文を作成する
- (3) 表4に従って、分類型を付与する

対象とした1,161語義のうち、728個を下位事象を内包する構造を持つと判断し、残りの433個をそうでないと判断した。728語義に対して付与された分類型の分布を、動詞の例とともに表5に示す。実際分類作業においては、表4に示される、「とりうる下位事象の標識のリスト」をすべて満たしていなくても、その分類型に分類している場合がある。この点に関して、今後、分類体系を見直す必要がある。

このようにして整理したデータは、表4を参照することにより、言論間の論理的関係認識に応用することができる。例えば、上記の例文(5b)において、このデータを

[☆] 例文(3a)~(3e)、(4a)も目的語コントロール構文である。

表 5 作業を行なった動詞リストにおける分類型の分布

分類型	数	例
決意型	25	決意する、決心する、思い立つ
約束型	28	約束する、協定する、請け負う
命令型	93	命令する、要請する、せがむ
言う型	1	言う
許可型	16	許可する、禁止する、免除する
提案型	3	提案する、発案する (2 語義)
発言型	170	発言する、報告する、申し上げる
感覚型	124	思う、認める、判断する
認識型	54	見る、聞こえる、教わる
感情型	143	驚く、焦る、感心する
態度型	71	歓迎する、叱る、手伝う
計	728	

参照すると、「許可する」は許可型の動詞であり、「作品を撮影する」のは「カメラマン」であることが分かるので、例文 (5b) は、「カメラマンが作品を撮影する。」という言論を含意することが分かる。また、このデータを参照すると、上位事象の動詞がどのような下位事象の標識をとることができるのかが分かるので、例えば、例文 (3a)~(3e) がすべて同義関係にあることを認識することができる。

5. 事象間関係知識データベースの評価実験

3 章で整備した事象間関係知識データベースを、述語項構造レベルの言論間類義・対立関係認識に応用し評価するという小規模な予備実験を行なった。

5.1 述語項構造レベルの言論間類義・対立関係認識

我々は、2つの事象間の関係知識のリストを用いて、述語項構造レベルの言論間の類義・対立関係を認識するシステムを提案した¹²⁾。このシステムは、ウェブを利用することにより、入力されたトピックに対して、まず、その論点集合と関連言論集合を獲得し、次に、論点ごとに、述語項構造レベルの言論間の類義・対立関係を認識し、最終結果を言論マップとして出力する。このとき、事象間の類義関係知識を用いることにより、言論をクラスタリングし、反義関係知識を用いることにより、言論クラスター間に < 対立 > という論理的关系を導入する。

5.2 実験方法

評価実験に用いたトピックは、「ステロイド」と「喫煙」の2つである。前節で述べたシステムを利用することにより、これらのトピックのそれぞれに対して、以下の5個、10個の論点集合を獲得した。

ステロイド 医師、皮膚、かゆみ、ステロイド剤、ステロイド軟膏

喫煙 肺がん、害、ニコチン、健康、危険性、リスク、マナー、女性、受動喫煙、禁煙

論点ごとに、トピックに関連する述語項構造レベルの言論間に、人手で類義関係と対立関係を付与することにより、評価コーパスを作成した。「ステロイド」と「喫煙」の各トピックに対する評価コーパスの規模は、それぞれ4,347 言論、5,989 言論であった。

表 6 トピック「ステロイド」における類義関係認識の結果

		コーパス		計
		○	×	
システム	○	822	204	1,026
	×	478	2,843	3,321
計		1,300	3,047	4,347

表 7 トピック「喫煙」における類義関係認識の結果

		コーパス		計
		○	×	
システム	○	501	129	630
	×	479	4,880	5,359
計		980	5,009	5,989

実験においては、事象間関係知識データベースの類義関係知識として、次の2種類のものを用いた。

- 以下に挙げる7種類の関係知識
 - 同義 (言い換え)、同義・上位、結果 (状態)、付帯状況、不可分、目的、手段
- 3.1.2 節で述べた、動詞項構造シソーラスの意味クラス体系による類義関係知識

反義関係知識としては、次の2種類のものを用いた。

- 反義語の関係知識
- 3.1.2 節で述べた、動詞項構造シソーラスの意味クラス体系による反義関係知識

「評価コーパス内に、類義関係にある言論を少なくとも1つ持っている言論を抽出できるか」という観点から、システムの類義関係認識の評価を行なう。人間の判断と比較することにより、その精度と再現率を求める。「人手による言論クラスタリング結果を与えたときに、人間と同じような対立関係認識ができるか」という観点から、システムの対立関係認識の評価を行なう。

5.3 結果

各トピックにおける類義関係認識の結果を、それぞれ表6と表7に示す。表の「○」は、類義関係にある言論を持つと判断した言論の数を、「×」は、そうでないと判断した言論の数を表す。各欄の値は、すべての論点に対する値を合計したものである。「ステロイド」における類義関係認識の精度は80.1%(822/1026)、再現率は63.2%(822/1300)であり、「喫煙」においては、精度は79.5%(501/630)、再現率は51.1%(501/980)であった。

前節の実験設定において、「ステロイド」における対立関係認識の精度は45%(14/31)、再現率は26%(14/54)であり、「喫煙」における対立関係認識の精度は59%(20/34)、再現率は13%(20/158)であった。

これらの結果から、事象間関係知識データベースを用いると、約8割の高い精度で言論間の類義関係を、5割程度の精度で対立関係を認識できることが分かる。上記の結果において再現率が低い主な理由は、評価コーパスにおいては、項の名詞間の類義・対立関係を考慮して言論間の論理的关系が付与されているのに対し、現在の我々のシステムでは、文字列完全一致を除いて、名詞間の関

係を全く考慮していないからである。

本予備実験の結果を観察することにより、「形容詞の連用形」+「動詞(なる、する、ある)」の形式に関する関係知識を収集する必要があることが分かった。例えば、以下の2つの言論が類義関係にあることを認識するためには、

- 喫煙でニコチン濃度が**高まる**
- 喫煙でニコチン濃度が**高くなる**

次のような事象間関係知識が必要である。

同義(言い換え) 何かが高まる ⇔ 何かが高くなる

今後は、このような関係知識を自動的に生成し整理するとともに、次のような関係知識も取り扱う必要がある。

同義・上位 誰かが何かに畏縮する ⇒ 誰かが**小さくなる**

6. おわりに

本論文では、言論マップの生成に必要な事象間関係知識の整備に関して行なった、既存の2つの言語資源の統合と、下位事象を内包する動詞項構造の記述について報告した。小規模な予備実験の結果ではあるが、このデータベースを用いると、約8割の高い精度で言論間の類義関係を認識できることが分かった。

今後は、2つの異なる手法を用いて、データベースの整理を続ける。1つは、人手による関係知識の記述であり、青山らの手法²⁾を形容詞、形容動詞、述語的慣用句に適用することにより、高品質のデータベースを作成することを目指す。もう1つは、大規模コーパスから獲得された関係知識の体系化であり、阿部らの手法¹⁾によって獲得された事象間関係知識を、例えば、FrameNet^{3),15)}で提案されているような、継承、部分全体、順序などの関係を用いて整理することを検討している。

謝 辞

下位事象を内包する動詞項構造の記述作業を実際に遂行していただきました、(有)ランゲージウェアの衛藤純司氏に深く感謝いたします。

本研究は、独立行政法人 情報通信研究機構の委託研究「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」の一環として実施した。

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積文の構造化. 言語処理学会 第13回年次大会発表論文集, pp. 286-289, 2007.
- 3) Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*,

pp. 86-90, 1998.

- 4) 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.
- 5) Timothy Chklovski and Patrick Pantel. VerbOcean: Mining the web for fine-grained semantic verb relations. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-04)*, pp. 33-40, 2004.
- 6) Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment*, 2005.
- 7) Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- 8) Takashi Inui, Kentaro Inui, and Yuji Matsumoto. Acquiring causal knowledge from text using the connective marker tame. *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. 4, No. 4, pp. 435-474, 2005.
- 9) 影山太郎. 動詞の意味と構文. 大修館書店, 2001.
- 10) 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.
- 11) 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.
- 12) 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. 55-60, 2008.
- 13) 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.
- 14) 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.
- 15) Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. FrameNet II: Extended theory and practice. <http://framenet.icsi.berkeley.edu/>, 2006.
- 16) 佐藤香織. 日本語のコントロール構文の成立条件について. 筑波応用言語学研究, Vol. 8, pp. 29-42, 2001.
- 17) 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.
- 18) Kentaro Torisawa. Acquiring inference rules with temporal constraints by using Japanese coordinated sentences and noun-verb co-occurrences. In *Proceedings of Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL06)*, pp. 57-64, 2006.