

2段階のクラウドソーシングによる 談話関係タグ付きコーパスの構築

河原 大輔^{1,2,a)} 町田 雄一郎^{1,b)} 柴田 知秀^{1,2,c)} 黒橋 禎夫^{1,2,d)} 小林 隼人^{3,e)} 颯々野 学^{3,f)}

概要：本稿では、クラウドソーシングを活用することによって、談話関係タグ付きコーパスを構築する手法を提案する。談話関係のタグ付けは従来は非常に長い時間とコストがかかるものであったが、提案手法ではこれを2段階のクラウドソーシングのタスクに分けることによって短時間で構築する。1文書が3文からなる短い日本語テキストを対象としてクラウドソーシングを行い、1万文書(3万文)からなるコーパスが8時間弱で構築できることを示す。また、構築したコーパスを利用して談話関係解析器を開発し、コーパスの有用性を示す。

Rapid Development of a Corpus with Discourse Annotations using Two-stage Crowdsourcing

DAISUKE KAWAHARA^{1,2,a)} YUICHIRO MACHIDA^{1,b)} TOMOHIDE SHIBATA^{1,2,c)} SADA0 KUROHASHI^{1,2,d)}
HAYATO KOBAYASHI^{3,e)} MANABU SASSANO^{3,f)}

Abstract: We present a novel approach for rapidly developing a corpus with discourse annotations using crowdsourcing. Although discourse annotations typically require much time and cost owing to their complex nature, we realize discourse annotations in an extremely short time while retaining good quality of the annotations by crowdsourcing two annotation subtasks. In fact, our experiment to create a corpus comprising 30,000 Japanese sentences took less than eight hours to run. Based on this corpus, we also develop a supervised discourse parser and evaluate its performance to verify the usefulness of the acquired corpus.

1. はじめに

我々はテキストを読んで解釈するとき、節や文などのそれぞれを単独で解釈するわけではなく、他の節や文などとの関係を理解しながらテキスト全体を解釈する。計算機によるテキスト理解を実現するためには、節・文間の関係を高精度に解析するシステムを構築する必要がある。このよ

うなシステムは談話関係解析や談話構造解析と呼ばれ、自然言語処理の重要かつ基盤的な解析の一つであるが、タグ付きコーパスの欠如のため、英語のようなメジャーな言語以外には開発されていない。

英語に対しては、Penn Discourse Treebank (PDTB) [1]、RST Discourse Treebank [2]、Discourse Graphbank [3]などの談話関係タグ付きコーパスが長い年月をかけて構築されており、それらを用いて機械学習した談話関係解析システムがいくつか公開され実際に使われている。他の言語について同じ枠組みで開発するにはやはり長い時間、莫大なコストがかかり、大きな問題となる。短時間、低コストでこれを実現する一つの手はクラウドソーシングを使うことである。しかし、これまでの典型的な談話関係アノテーションは、テキスト中で関係をもつスパンのペアをまず見つけ、次にスパン間の関係の種類を明らかにするという難

¹ 京都大学 大学院情報学研究所
Graduate School of Informatics, Kyoto University

² 独立行政法人 科学技術振興機構, CREST
CREST, Japan Science and Technology Agency

³ ヤフー株式会社
Yahoo Japan Corporation

a) dk@i.kyoto-u.ac.jp

b) machida@nlp.ist.i.kyoto-u.ac.jp

c) shibata@i.kyoto-u.ac.jp

d) kuro@i.kyoto-u.ac.jp

e) hakobaya@yahoo-corp.jp

f) msassano@yahoo-corp.jp

しい判断となり、これをそのままクラウドソーシングに載せることは困難である。

本論文では、談話関係のアノテーション手続きを単純化し、クラウドソーシングによって談話関係タグ付きコーパスを構築する。まず、関係をもつスパンを同定するのは高コストであるので、自動分割した節を単位とし、この区切りは高精度であるので修正しないこととする。タグ付け対象の各文書は3文からなるものとし、そこに含まれる節は5節までとして単純化する。談話関係タイプの付与については、一回に任意の節間に成り立つ関係を判定するのはクラウドソーシング上では困難であるので、談話関係有無の判定と談話関係タイプの判定の2段階クラウドソーシングで行う。談話関係のタグセットとしては、PDTBをベースに単純化し、2階層からなるタグセットを設計した。上位タイプは、“根拠・条件”、“転換”、“その他”の3タイプからなる。このように、単純化した問題設定ではあるが、リーズナブルな設定であると考えており、ここで得られた知見やモデルは、一般的な談話関係解析に発展させることが十分に可能であると考えている。

実験では、対象言語を日本語とし、1万文書(3万文)からなる談話関係タグ付きコーパスを2段階のクラウドソーシングによって作成した。クラウドソーシングに要した時間は合計8時間弱であり、従来の大規模な談話関係アノテーションと比べて非常に高速に構築することができた。

また、得られた談話関係アノテーションを用いて、談話関係解析システムを構築した。談話関係解析の機械学習モデルを作成し、交差検定によってその精度を評価した。このコーパスを用いて、談話関係解析モデルを機械学習し、その精度を計った。その結果、“根拠・条件”タイプについて37.9%のF値で解析できることがわかった。これは、英語の談話関係解析器の精度と近い値であり、得られたコーパスの有用性が示されたと考えている。構築した談話関係解析システムは、感情分析[4]や対立言明認識[5],[6]のような言語処理アプリケーションにおいて今後利用する予定である。

2. 関連研究

Snowらは5つのアノテーションタスクをクラウドソーシングで行っているが、いずれのタスクも非常に簡単な問題設定にしている[7]。複雑なアノテーションをクラウドソーシングで行った研究がいくつかある。Negriらは、cross-lingual textual entailment コーパスをクラウドソーシングで構築する手法を提案している[8]。彼らは、この複雑なタスクを、文修正、タイプ付与、文翻訳というサブタスクに分割することによって、クラウドソーシングに載せている。このような自由記述を伴う生成的なタスクは、我々の対象としている談話関係アノテーションとは異なっている。Fossatiらは、クラウドソーシングを利用し

たFrameNetアノテーションを提案している[9]。この手法は、項(frame element)だけを同定する1段階の手法であり、小規模な実験によってその有効性を示している。彼らは大規模な実験を実行中としているが、その結果はこれまでのところ明らかになっていない。Hongらは、FrameNetのフレーム選択をクラウドソーシングで行う手法を提案している[10]。このタスクは語彙曖昧性解消と等価であり、FrameNetアノテーション全体と比べると複雑なものではない。これらのFrameNetアノテーションは、我々の対象としている談話関係のアノテーションとはかなり異なっている。談話関係アノテーションをクラウドソーシングで行った先行研究はこれまでのところ存在しない。

人手で構築した談話関係タグ付きコーパスとして、Penn Discourse Treebank [1]、RST Discourse Treebank [2]、Discourse Graphbank [3]が有名である。これらはいずれも英語の新聞記事をタグ付け対象としている。英語以外の言語に対しても、談話関係コーパスを構築する研究がいくつか行われている。それらは、ドイツ語を対象としたPotsdam Commentary Corpus [11] (新聞記事2,900文)、ポルトガル語を対象としたRhetalho [12] (科学技術論文100文書; 1,350文)、スペイン語を対象としたRST Spanish Treebank [13] (複数ジャンルの267文書; 2,256文)である。これらのコーパスは、英語のコーパス(数万文)と比較すると規模が小さい。

また、上記のコーパスを利用した談話関係解析の研究が近年活発に行われている[14],[15],[16],[17],[18],[19],[20],[21],[22],[23],[24]。この談話関係解析研究の盛り上がりは、談話関係が付与された大規模なコーパスが利用可能であることに依るところが大きいと思われる。しかし、そのような大規模コーパスが英語においてのみ利用可能であるため、対象の言語はほとんどの場合、英語となっている。英語以外の言語に対する談話関係解析器を開発し改良するためには、大規模な談話関係タグ付きコーパスを構築することが必要不可欠である。それも、可能なかぎり短期間で構築することが望ましい。

3. クラウドソーシングによる談話関係タグ付きコーパスの構築

3.1 コーパスの仕様

本研究では、談話単位のペアごとに談話関係がタグ付けされたコーパスを構築する。そのためには、対象の文書、談話単位、談話関係タグセットを決める必要がある。それぞれについて、以下で詳細に述べる。

3.1.1 対象文書と談話単位

これまで作られた談話構造タグ付きコーパスはWall Street Journalのような新聞記事を対象にしている。その上で学習したシステムは、他のドメインのテキストに適用するときに分野適応の問題が生じる。つまり、新聞記事

表 1 談話関係タグセット
Table 1 Discourse relation tagset.

上位タイプ	下位タイプ	例
根拠・条件	原因・理由	【ボタンを押したので】【お湯が出た。】
	目的	【試験に受かるために】【必死に勉強した。】
	条件	【ボタンを押せば】【お湯が出る。】
	その他根拠	【ここにカバンがあるから】【まだ社内にいるだろう。】
転換	対比	【あのレストランは寿司はおいしいが】【ラーメンは普通だ。】
	譲歩	【あのレストランは確かにおいしいが】【値段は高い。】
その他	(その他)	【家に着いてから】【雨が降ってきた。】

コーパスで訓練した談話関係解析器は、新聞記事については高精度に解析することができるが、新聞記事以外の分野のテキストについてはあまり精度がよくないという問題が生じる。

この分野適応の問題を起こさないために、我々は新聞記事だけではなく、さまざまなドメインの文書をカバーしたコーパスに対してタグ付けを行う。Web上のテキストはさまざまな分野のものを含んでいるため、本研究では、Webから抽出されたWeb文書書き始め3文コーパス[25]を対象にする。また、このコーパスにおける1文書は、各Webページ先頭の3文からなっており、クラウドソーシングでタグ付けを行うにはこれくらいの短さが簡単で適していると思われる。

談話関係のタグ付けを行う単位である談話単位は節とする。この理由としては、PDTBで採用されているようなスパンは、クラウドソーシングで行うには高精細すぎること、また文を単位とすると談話関係と捉えるには粗すぎるということが挙げられる。節への分割は自動で行う。節分割の精度は高いため、分割結果の手修正は行わないこととした。節分割は、柴田らの談話関係解析器[26]で使われているルールを利用して行った。このルールはたとえば次のようなものである。

- 「～が」「～ので」のような比較的強い区切りとして働く節を談話単位として認定する
- 連体修飾節は談話単位としない

一部の連体修飾節は談話単位と認定した方がよい場合がある[27]が、これについては今後の課題とした。

クラウドソーシングのワーカーは任意の節間の関係を考える必要があるため、3文に含まれる節数が多くなるとワーカーの負荷が組み合わさ的に大きくなる。これを回避するために、3文中に含まれる節数は5節までとした。この制限によって対象外となる文書はもとのコーパスに含まれる文書の約5%であった。

談話関係をタグ付けする文書の本数は1万文書とし、計3万文とした。本コーパスに含まれている節の本数は39,032個であり、1文書あたり平均3.9個の節が含まれていた。コーパスに含まれている節ペアの本数は59,426個であった。

3.1.2 談話関係タグセット

談話関係解析のひとつのアプリケーションは、賛否両論があるトピックに対する俯瞰的なマップを生成する言論マップ[5]やDispute Finder[6]である。これらのシステムにおいて、対立関係を含む、言論間のさまざまな関係の自動同定技術が使われている。この目的には、詳細化や換言のような付加的な談話関係や時間関係は重要ではないと考えた。この設定は、因果関係とは独立に時間関係を付与している[28]と似ている。この研究のように、時間関係を必要とするアプリケーションのために、時間関係を別途付与することも考えられる。

このような考えのもとに、談話関係タグセットはPDTBを参考にして決めた。タグセットは、表1に示すように、2階層、下位7つの分類からなる。なお、談話関係の方向性は、クラウドソーシングのためにタスクを単純化するために扱わないこととした。

我々のタグ付けタスクは、文書中の任意の節ペアに上記のいずれかの談話関係が成り立つかどうかをアノテーションすることである。以下にタグ付け例を示す。ここでは、節区切りは“::”で示している。また、明示的に談話関係を示していない節ペアには“その他”関係が与えられている。

原因・理由 気がつけば::梅雨も明けてました。::毎日暑い日が続きますね。::【父の手術も無事に終わり、】::【少しだけほっとしてます。】

対比 今日とある企業のトップの話聞くことが出来た。::経営者として何事も全てビジネスチャンスに変えるマインドが大切だと感じた。::【生きていく上で追い風もあれば、】::【逆風もある。】

3.2 2段階クラウドソーシングの手続き

本研究では、2段階のクラウドソーシングによって、談話関係タグ付きコーパスを構築する。この2つのステージは、節ペアに対する談話関係の有無の判定と、談話関係があると判定された節ペアに対する談話関係タイプの判定からなる。以下ではそれぞれについて詳細に述べる。

表 2 談話関係有無判定の結果

Table 2 Number of clause pairs resulting from the judgments of discourse relation existence.

確率	頻度
= 1.0	64
> 0.99	554
> 0.9	1,065
> 0.8	1,379
> 0.5	2,655
> 0.2	4,827
> 0.1	5,895
> 0.01	9,068
> 0.001	12,277
> 0.0001	15,554

3.2.1 ステージ 1: 談話関係有無判定

まず、各文書 3 文中に含まれる節ペアごとに、“原因結果”、“目的”、“条件”、“その他根拠”、“対比”、“譲歩”の談話関係(すなわち、“その他”以外)があるかどうかをクラウドソーシングで判定する。

ワーカーに提示される 1 設問は、1 文書に含まれるすべての節ペアに対する判定からなる。こうすることによって、文書全体を考慮して、談話関係の有無を判断することができる。

3.2.2 ステージ 2: 談話関係タイプ判定

次に、第 1 段階において談話関係があると判断された節ペアに対して、談話関係のタイプを判定する。タイプ判定は、談話関係タグセットの下位の 7 つの関係から一つを選択することによって行う。もし、ワーカーが“その他”を選択すれば、これは関係ありと判定された第 1 段階の結果をキャンセルすることに相当する。

ワーカーに提示される 1 設問は、1 つの節ペアの判定である。すなわち、1 文書中に判断すべき複数の節ペアがある場合には、複数の設問に分割されるが、このような場合は稀である。

3.3 結果

Yahoo!クラウドソーシング^{*1}を用いて、3.2 節の 2 段階クラウドソーシングを実行した。構築するコーパスの信頼性を高めるために、それぞれの判定における各設問を判定するワーカーの数を 10 人とした。Snow ら [7] は、専門家レベルの質のアノテーションをクラウドソーシングで実現するには、ワーカーの数を 4 人以上の設定すればよいと結論付けているが、ワーカーの数を 10 人まで増やすに連れて、いくつかのタスクにおける質が向上しているため、本タスクにおいては 10 人と設定した。また、ワーカーの質を担保するために、10 設問ごとに 1 つの隠れゴールド問題を設定した。これは正解をあらかじめ与えてある簡単な問

^{*1} <http://crowdsourcing.yahoo.co.jp/>

表 3 談話関係下位クラスに対する談話関係タイプ判定の結果

Table 3 Results of the judgments of lower discourse relation types.

下位クラス	すべて	確率 > 0.8
原因・理由	2,104	1,839 (87.4%)
目的	755	584 (77.4%)
条件	1,109	925 (83.4%)
その他根拠	442	273 (61.8%)
対比	437	354 (81.0%)
譲歩	80	49 (61.3%)
上記の合計	4,927	4,024 (81.7%)
その他	4,141	3,753 (90.6%)
合計	9,068	7,777 (85.8%)

表 4 談話関係上位クラスに対する談話関係タイプ判定の結果

Table 4 Results of the judgments of upper discourse relation types.

上位クラス	すべて	確率 > 0.8
根拠・条件 転換	4,439	3,993 (90.0%)
転換	516	417 (80.8%)
上記の合計	4,955	4,410 (89.0%)
その他	4,113	3,753 (91.2%)
合計	9,068	8,163 (90.0%)

題であり、ワーカーがこの問題を連続して間違えればテスト問題が課され、タスクを継続するためにはこれに正解する必要がある。

第 1 段階のクラウドソーシング、つまり談話関係有無判定タスクを実行し、10,000 文書中の 59,426 節ペアそれぞれについて 10 個の判定を得た。この結果から、Whitehill らによって提案された GLAD[29] を用いて、それぞれの節ペアに対するラベル(談話関係の有無)の確率を計算した。この手法は、単純な多数決を用いるよりも高い精度でラベルを推定することができる報告されている。表 2 に結果を示す。第 2 段階のクラウドソーシングで関係タイプを判定する対象となる節ペアを選択するために、この確率に閾値を設定した。本論文では、閾値を 0.01 と設定し、この閾値を上回る 9,068 個の節ペア(全体の 15.3%)を選択した。この閾値はかなり低いだが、これは、確率の低い節ペアについて第 2 段階のクラウドソーシングでもう一度判定するためである。

第 2 段階のクラウドソーシングにおいては、9,068 個の節ペアについて談話関係タイプの判定を行い、それぞれの節ペアについて 10 個の判定を得た。第 1 段階と同様に、GLAD を多値に拡張して、節ペアごとに各ラベル(談話関係タイプ)の確率を求めた。最終的に、節ペアごとに、もっとも高い確率をもつ談話関係タイプを付与した。表 3 に談話関係タグセット下位クラスに対する分布を示す。表において、2 つ目の列は、それぞれのタイプの頻度を示し、3 つ目の列は確率が 0.8 以上の頻度を示している。表 4 には、

表 5 アノテーション結果の例。第 1 列は確率、第 2 列は第 3 列のタイプを付与したワーカーの数を示す。第 4 列において、第 3 列のタイプをもつ節ペアを【】で示す。

Table 5 Examples of Annotations. The first column denotes the estimated label probability and the second column denotes the number of workers that assigned the designated type. In the fourth column, the clause pair annotated with the type is marked with 【】.

確率	#ワーカー	タイプ	文書
1.00	6/10	原因・理由	ツツジ科・ツツジ属。【花が陰暦五月に咲くため】【「臯月」と呼ばれている。】市制 20 年を記念して、1979 年 11 月 3 日に制定された。
0.99	4/10	条件	【↓マップ上の吹き出しをクリックすると】【おすすめルートがご覧になれます。】市町村名をクリックすると「見どころ・体験・食」の情報がご覧になれます。緑色の表記は各スポットの写真正覧になれます。
0.81	3/10	目的	ダイランティアはマナによって支えられた世界。しかし、人類の繁栄と共に世界樹が 3 年に一度結実させる「大いなる実り」だけでは人類の繁栄を支えることができなくなってしまった。【そして「大いなる実り」を求めて】【各国が戦争を繰り返して行く。】
0.61	2/10	原因・理由	スケールは（一部を除き）1/32 とされている。これは単 3 形乾電池 2 本が入りやすいようにしたサイズである。動力は単 3 形乾電池 2 本と FA-130 サイズのモーター 1 個で、【ギヤーとシャフトの組み合わせにより動力を前後の車軸に伝達し、】【4 輪を駆動する。】
0.54	3/10	対比	来年春には、阪急百貨店が新博多駅に東急ハンズと共に目見えする。そうすると【百貨店による顧客の奪い合いが厳しくなる。】【そこに浮上するのが、三越福岡の閉鎖の可能性である。】

下位クラスを上位クラスにマージした結果の分布を示す。表 5 に結果のアノテーションの例を示す。

クラウドソーシングによる上記 2 つのタスクの実行には、それぞれ約 3 時間、5 時間かかり、1,458 人、1,100 人のワーカーが作業を行った。もしこのタスクを 1 つのタスクとして実行したならば、約 33 時間 (5 時間 / 0.153) かかり、これは、我々の提案する 2 段階の手法と比べて 4 倍長いことになる。また、このような 1 段階の手法は、2 段階の手法がもっているようなダブルチェックの機構がないため、頑健さに欠けると思われる。金額的なコストとしては、それぞれのタスクに 11.1 万円、11.3 万円かかり、従来の大規模な談話関係アノテーションプロジェクトと比べて、非常に安価にコーパスを構築できたと考えられる。

表 5 のアノテーション例において、上から 4 つ目までの例の談話関係タイプは正しいが、一番下の例の“対比”は正しくないと思われる。この例では、2 つ目の括弧で示された節は、1 つ目の例示であると考えられるので、正しいタイプは“その他”である。このような間違いは、特に確率 0.8 以下の節ペアについて見つかった。

4. 談話関係解析器の構築

構築した談話関係付きコーパスの有用性を検証するために、このコーパスを教師データに用いた機械学習に基づく談話関係解析器を作成し、評価した。解析器としては、談話関係タグセットの上位タイプおよび下位タイプのそれぞれを対象とした 2 種類を作成した。

第 1 段階の談話関係有無判定の結果において、関係ありの確率が 0.01 より低い節ペアに“その他”タイプを与えた。それ以外の節ペアについては、第 2 段階の談話関係タイプ判定の結果において、確率 0.8 を越える談話関係タイ

表 6 談話関係解析器の素性

Table 6 Features for our discourse parsers.

名前	説明
節距離	節間の距離 (節単位)
文距離	節間の距離 (文単位)
単語	節に含まれる単語 (原形)
述語	節の述語 (原形)
述語活用型	節の述語の活用型
接続詞	節の最初の単語が接続詞ならばその単語 (原形)
単語重複率	2 節間における単語重複率
節タイプ	KNP によって出力される節のタイプ (約 100 種類)
「は」有無	節中に係助詞「は」があるかどうか
「は」共起	2 節ともに係助詞「は」があるかどうか

プがあればそれを採用し、なければその節ペアを実験に用いないこととした。この処理の結果、下位タイプについて 58,135 個 (50,358 + 7,777) の節ペア、上位タイプについて 58,521 個 (50,358 + 8,163) の節ペアが得られた。それらのうち、4,024 個 (6.9%) と 4,410 個 (7.5%) の節ペアに“その他”以外のタイプが付与されていた。この 2 種類のデータを用いて、5 分割交差検定を行った。

機械学習の素性を抽出するために、日本語形態素解析器 JUMAN*2 と日本語構文・格解析器 KNP*3 を上記コーパスに適用した。用いた素性を表 6 に示す。これらは、他言語の談話関係解析器で標準的に使われているものである。

機械学習ツールとしては、opal[30]*4 を用いた。このツールでは、多項式カーネルを用いたオンライン学習が可能である。学習には、2 次の多項式カーネルを用いた

*2 <http://nlp.ist.i.kyoto-u.ac.jp/?JUMAN>

*3 <http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

*4 <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/opal/>

表 7 下位タイプに対する談話関係解析器の精度
Table 7 Performance of our lower-type discourse parser.

タイプ	適合率	再現率	F1
原因・理由	0.623 (441/708)	0.240 (441/1,839)	0.346
目的	0.489 (44/90)	0.075 (44/584)	0.131
条件	0.581 (256/441)	0.277 (256/925)	0.375
その他根拠	0.000 (0/12)	0.000 (0/273)	0.000
対比	0.857 (6/7)	0.017 (6/354)	0.033
譲歩	0.000 (0/0)	0.000 (0/49)	0.000
その他	0.944 (53,702/56,877)	0.992 (53,702/54,111)	0.968

表 8 上位タイプに対する談話関係解析器の精度
Table 8 Performance of our upper-type discourse parser.

タイプ	適合率	再現率	F1
根拠・条件	0.625 (1,084/1,735)	0.272 (1,084/3,993)	0.379
転換	0.412 (7/17)	0.017 (7/417)	0.032
その他	0.942 (53,454/56,769)	0.988 (53,454/54,111)	0.964

passive-aggressive (PA-I) アルゴリズムを多クラス分類に拡張 [31] したものを利用した。クラス数は、下位クラス解析器の場合は7つで、上位クラス解析器の場合は3つである。また、aggressiveness パラメータについては、他の分類タスクで精度がよかった 0.001 に設定し、他のパラメータは opal のデフォルトとした。

談話関係解析器を評価するために、評価尺度として、それぞれの談話関係タイプごとの適合率、再現率および F 値 (F1) を採用した。表 7 と表 8 に、下位クラスおよび上位クラス談話関係解析器の精度を示す。

表 8 を見ると、上位クラス談話関係解析器は、“根拠・条件” タイプに対して 37.9% の F 値を出している。言語やデータが異なるため、先行研究と比較するのは困難であるが、英語の談話関係解析器と比べて同等の精度を達成していると考えられる。たとえば、Lin ら [24] の談話関係解析器は、PDTB に対して 20.6%~46.8% の F 値を達成している。

また、我々の談話関係解析器は、“転換”(下位クラスでは対比もしくは譲歩) に対する F 値が低い。これは PDTB に対する解析器と同じ傾向である。この最大の原因は、この関係を明示するような表現(たとえば「X は～したが、一方で Y は～した」) があまり使われないことである。Pitler ら [14] などの研究に示されているように、明示的な表現が使われない場合の談話関係認識は非常に難しい。これらの精度向上のためには、訓練データ以外に外部の知識源を利用する必要がある。大規模な生コーパスから獲得した知識を利用することによって、この問題を解決したいと考えている。

5. おわりに

本稿では、2段階のクラウドソーシングを用いて、談話関係タグ付きコーパスを構築する手法を述べた。談話構造のタグ付けは従来は非常に長い時間とコストがかかるものであったが、提案手法ではこれを2段階のクラウドソーシングのタスクに分けることによって、8時間弱で構築することができた。完成したコーパスは、一般に公開する予定であり、これによって、談話関係解析の研究が進展することが期待される。

参考文献

- [1] Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. and Webber, B.: The Penn Discourse Tree-Bank 2.0, *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pp. 2961–2968 (2008).
- [2] Carlson, L., Marcu, D. and Okurowski, M. E.: Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory, *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue* (2001).
- [3] Wolf, F. and Gibson, E.: Representing Discourse Coherence: A Corpus-Based Study, *Computational Linguistics*, Vol. 31, No. 2, pp. 249–287 (2005).
- [4] Zirn, C., Niepert, M., Stuckenschmidt, H. and Strube, M.: Fine-Grained Sentiment Analysis with Structural Features, *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp. 336–344 (2011).
- [5] Murakami, K., Nichols, E., Matsuyoshi, S., Sumida, A., Masuda, S., Inui, K. and Matsumoto, Y.: Statement Map: Assisting Information Credibility Analysis by Visualizing Arguments, *Proceedings of the 3rd Workshop on Information Credibility on the Web*, pp. 43–50 (2009).
- [6] Ennals, R., Trushkowsky, B. and Agosta, J. M.: High-

- lighting Disputed Claims on the Web, *Proceedings of the 19th international conference on World Wide Web*, pp. 341–350 (2010).
- [7] Snow, R., O'Connor, B., Jurafsky, D. and Ng, A.: Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks, *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 254–263 (2008).
- [8] Negri, M., Bentivogli, L., Mehdad, Y., Giampiccolo, D. and Marchetti, A.: Divide and Conquer: Crowdsourcing the Creation of Cross-Lingual Textual Entailment Corpora, *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 670–679 (2011).
- [9] Fossati, M., Giuliano, C. and Tonelli, S.: Outsourcing FrameNet to the Crowd, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 742–747 (2013).
- [10] Hong, J. and Baker, C. F.: How Good is the Crowd at “real” WSD?, *Proceedings of the 5th Linguistic Annotation Workshop*, pp. 30–37 (2011).
- [11] Stede, M.: The Potsdam commentary corpus, *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, pp. 96–102 (2004).
- [12] Pardo, T. A. S., Nunes, M. d. G. V. and Rino, L. H. M.: Dizer: An automatic discourse analyzer for Brazilian Portuguese, *Advances in Artificial Intelligence—SBIA 2004*, Springer, pp. 224–234 (2004).
- [13] da Cunha, I., Torres-Moreno, J.-M. and Sierra, G.: On the development of the RST Spanish treebank, *Proceedings of the 5th Linguistic Annotation Workshop (LAW V)*, pp. 1–10 (2011).
- [14] Pitler, E., Louis, A. and Nenkova, A.: Automatic sense prediction for implicit discourse relations in text, *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 683–691 (2009).
- [15] Pitler, E. and Nenkova, A.: Using Syntax to Disambiguate Explicit Discourse Connectives in Text, *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 13–16 (2009).
- [16] Subba, R. and Di Eugenio, B.: An effective Discourse Parser that uses Rich Linguistic Information, *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 566–574 (2009).
- [17] Hernault, H., Prendinger, H., duVerle, D. and Ishizuka, M.: HILDA: A discourse parser using support vector machine classification, *Dialogue & Discourse*, Vol. 1, No. 3, pp. 1–33 (2010).
- [18] Ghosh, S., Tonelli, S., Riccardi, G. and Johansson, R.: End-to-End Discourse Parser Evaluation, *Fifth IEEE International Conference on Semantic Computing (ICSC)*, pp. 169–172 (2011).
- [19] Feng, V. W. and Hirst, G.: Text-level Discourse Parsing with Rich Linguistic Features, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, pp. 60–68 (2012).
- [20] Joty, S., Carenini, G. and Ng, R.: A Novel Discriminative Framework for Sentence-Level Discourse Analysis, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 904–915 (2012).
- [21] Joty, S., Carenini, G., Ng, R. and Mehdad, Y.: Combining Intra- and Multi-sentential Rhetorical Parsing for Document-level Discourse Analysis, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 486–496 (2013).
- [22] Biran, O. and McKeown, K.: Aggregated Word Pair Features for Implicit Discourse Relation Disambiguation, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 69–73 (2013).
- [23] Lan, M., Xu, Y. and Niu, Z.: Leveraging Synthetic Discourse Data via Multi-task Learning for Implicit Discourse Relation Recognition, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 476–485 (2013).
- [24] Lin, Z., Ng, H. T. and Kan, M.-Y.: A PDTB-styled end-to-end discourse parser, *Natural Language Engineering*, Vol. 20, No. 2, pp. 151–184 (2014).
- [25] Hangyo, M., Kawahara, D. and Kurohashi, S.: Building a Diverse Document Leads Corpus Annotated with Semantic Relations, *Proceedings of 26th Pacific Asia Conference on Language Information and Computing*, pp. 535–544 (2012).
- [26] Shibata, T. and Kurohashi, S.: Automatic Slide Generation Based on Discourse Structure Analysis, *Proceedings of Second International Joint Conference on Natural Language Processing*, pp. 754–766 (2005).
- [27] 飯田龍, 徳永健伸: 文内に出現する談話関係を認定するための接続表現の調査, 言語処理学会第20回年次大会, pp. 173–176 (2014).
- [28] Bethard, S., Corvey, W., Klingenstein, S. and Martin, J. H.: Building a Corpus of Temporal-Causal Structure, *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pp. 908–915 (2008).
- [29] Whitehill, J., Ruvolo, P., fan Wu, T., Bergsma, J. and Movellan, J.: Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise, *Advances in Neural Information Processing Systems 22* (Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C. K. I. and Culotta, A., eds.), pp. 2035–2043 (2009).
- [30] Yoshinaga, N. and Kitsuregawa, M.: Kernel slicing: Scalable online training with conjunctive features, *Proceedings of the 23rd International Conference on Computational Linguistics (COLING2010)*, pp. 1245–1253 (2010).
- [31] Matsushima, S., Shimizu, N., Yoshida, K., Ninomiya, T. and Nakagawa, H.: Exact Passive-Aggressive Algorithm for Multiclass Classification Using Support Class, *Proceedings of 2010 SIAM International Conference on Data Mining (SDM2010)*, pp. 303–314 (2010).