

文章内談話関係タグ付けの基準

河原大輔 澤田晋之介 岸本裕大 黒橋禎夫
京都大学

2022 年 5 月 31 日

1 はじめに

本稿は、文章内の節・文間に談話関係をタグ付けするための基準を示したものである。談話関係は意味的な関係であり、作業者間のタグ付けの不一致が、構文関係のタグ付けなどと比べてより大きな問題となる。この問題に対処するために、談話関係のタグ付け基準・作業を簡単化、明瞭化した。これは主には、談話関係を付与する単位を、自動分割した節にすること、談話関係のタグセットを 2 階層 7 分類の簡単なものにすることによって実現している。

また、本基準に従って、談話関係をクラウドソーシングで付与することも可能である。クラウドソーシングによる談話関係のタグ付けは 5.2 節で述べる。

2 基本スペック

2.1 談話関係の単位

述語項構造を中心とした連続する句の列を節と定義し、節を談話単位とする。なお、1 つの節からなる文も節と呼ぶこととする。談話単位として節を採用した理由は、文を単位とすると談話関係と捉えるには粗すぎ、句を単位とすると逆に粒度が小さすぎるからである。文の節への自動分割は高精度にできるため、自動分割した節を談話単位とし、分割結果の人手修正は行わないこととする。節分割は、構文解析システム KNP を利用して、次のようなルールを用いて行う。

- (i) 文末、主節、並列節の末尾を節の境界とする

表 1 談話関係タグセット

上位タイプ	下位タイプ	例
順接系	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。】
	根拠	【ここにカバンがあるから】【まだ社内にいるだろう。】
逆接系	対比	【大阪は雨だが、】【東京は晴れだ。】
	逆接・譲歩	【あのレストランはおいしいが】【値段は高い。】
関係なしまたは弱い関係		【家に着いてから】【雨が降ってきた。】

- (ii) 主節に対する従属度が低い従属節の末尾を節の境界とする
- (iii) 主節に対する従属度が中程度で、かつ項を 1 つ以上持つ従属節の末尾を節の境界とする
- (iv) 連体修飾節、補文節の末尾は節の境界としない
- (v) 括弧内の句は節の境界としない

文を節に自動分割する際、係り受け関係を無視して分割している。そのため、談話単位として不十分な場合が生じる。具体例を示す。

- (1) 【テーブルは時間があれば】【こまめに拭いていらっしゃるようですが、】床は掃除していないのですね。

本稿中の例において、2 つの【…】によってタグ付け対象の節ペアを表す。順序として前側に出現した節を前節、後側の節を後節と呼ぶ。

例 (1) の場合、「テーブルは」は後節の「拭いていらっしゃる」に係っている。しかし節は連続する句の集合体と定義しているため、下線部は前節に含まれてしまう。このように、節が入れ子構造になっている場合の取り扱いは今後の課題とする。

2.2 談話関係タグセット

PDTB 2.0 は付録 A に示すとおり、3 階層、下位 30 種類のタグからなる。日本語の談話関係タグセットは PDTB 2.0 よりも簡単かつ明瞭なものを目指し、談話関係解析を応用したアプリケーションを想定した上で談話関係タグの取捨選択を行なった。談話関係解

析のアプリケーションとして、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 [1] や Dispute Finder[2] が挙げられるが、これらのシステムでは、時間経過や詳細化、換言といった関係よりも、因果、根拠、比較などの比較的強い談話関係が重要となる。そこで、因果や根拠を示す PDTB 2.0 の“CONTINGENCY” タグや、比較を示す PDTB 2.0 の“COMPARISON” タグ、『論理トレーニング』 [3] を参考に、2 階層、下位 7 種類からなるタグセットを設計した (表 1)。表中の上位タイプは上位のタグ 3 種類の総称、下位タイプは下位のタグ 7 種類の総称である。“関係なしまたは弱い関係” (以降、“談話関係なし” と呼ぶ) は談話単位間に意味的なつながりがない場合だけでなく、詳細化や換言などの付加的な関係、時間関係も含まれる。

2.3 談話関係の方向

順接系および逆接系の下位タイプ“逆接”については、談話関係の方向が存在する。具体例を挙げる。

- (2) 【ボタンを押したので、原因・理由】【お湯が出た。】
- (3) 【お湯が出た。】【ボタンを押したためだ。原因・理由】

例 (2) では、前節の「ボタンを押した」ことが原因で、後節の「お湯が出た。」という結果になったと解釈できる。このような前節が従、後節が主となる場合を順方向と呼ぶ。逆に、例 (3) では前節の「お湯が出た」理由を後節で述べている。このような前節が主、後節が従となる関係を関係を逆方向と呼ぶ。以降、例 (2)・例 (3) のように、従属的な節の末尾に関係名を記載することによって談話関係と方向の両方を示す。すなわち、順方向の場合には前節の末尾に、逆方向の場合には後節の末尾に関係名を記載する。

2.4 談話関係タグ付け

本コーパスは、文書中に含まれる全ての節ペアに対し、談話標識と表 1 に示した談話関係タグのいずれか 1 つを付与する。なお、談話標識が節ペア中に存在しない場合はアノテーションしない。例 (4) を用いてタグ付け方法を説明する。

- (4) 一人で悩んでいてはいつまでたっても解決できません。まずは気軽に相談してください。臨床心理士などが、いつでも相談にのります。

構文解析システム KNP を用いると、例 (4) は以下の 4 節に分割される。

- (i) 一人で悩んでいては、
- (ii) いつまでたっても解決できません。
- (iii) まずは気軽に相談してください。
- (iv) 臨床心理士などが、いつでも相談にのります。

この 4 節の全組み合わせに対して談話関係のタグ付けを行う。具体的には以下の 6 例にタグ付けをする。

- (5) 【一人で悩んでいては 条件】【いつまでたっても解決できません。】まずは気軽に相談してください。臨床心理士などが、いつでも相談にのります。
- (6) 【一人で悩んでいては】いつまでたっても解決できません。【まずは気軽に相談してください。】臨床心理士などが、いつでも相談にのります。
- (7) 【一人で悩んでいては】いつまでたっても解決できません。まずは気軽に相談してください。【臨床心理士などが、いつでも相談にのります。】
- (8) 一人で悩んでいては【いつまでたっても解決できません。】【まずは気軽に相談してください。】臨床心理士などが、いつでも相談にのります。
- (9) 一人で悩んでいては【いつまでたっても解決できません。】まずは気軽に相談してください。【臨床心理士などが、いつでも相談にのります。】
- (10) 一人で悩んでいてはいつまでたっても解決できません。【まずは気軽に相談してください。 条件】【臨床心理士などが、いつでも相談にのります。】

例 (5) は前節が条件となる節ペアなので順方向の“条件”を付与する。例 (6) は節ペアに談話関係がないため“談話関係なし”を付与する。同様に、例 (7)、例 (8)、例 (9) も談話関係がないため“談話関係なし”を付与する。例 (10) は前節を条件とみなした順方向の“条件”と、後節を前節の理由とみなした逆方向の“原因・理由”の両方の解釈ができる。この場合、順方向の関係を優先するため例 (10) は順方向の“条件”を付与する。

3 談話関係タグごとの基準

3.1 順接系

3.1.1 原因・理由

「～ので」、「だから」のような原因・理由を示す談話標識が節ペアに含まれている、もしくは原因・理由を示す談話標識を節ペアに挿入しても意味的なつながりが変化しない場合、“原因・理由”のタグを付与する^{*1}。この関係をもつ節ペアには方向があり、一般に、一方の節が原因もしくは理由、他方の節が結果となる^{*2}。例を示す。

- (11) 【必死に勉強したので、原因・理由】【試験に受かった。】
- (12) 【お客様一人ひとりにあったまつ毛のスタイルや、ケアの方法をご提案しています。原因・理由】【何でもお気軽にご相談ください。】
- (13) 【明るく、分かりやすく、楽しく教えていきたいです。】【勉強は誰もが嫌いなものです。原因・理由】

なお、例 (12) のような比較的弱い因果関係を持つ場合や、例 (13) のような書き手の主観的な理由も“原因・理由”とする。

3.1.2 目的

「～するため」のような目的を示す談話標識が使われている、もしくは目的を示す談話標識を節ペアに挿入しても意味的なつながりが変化しない場合、“目的”とする^{*3}。この関係をもつ節ペアにも方向があり、一般に、一方の節が目的、他方の節が手段となる。例を以下に示す。

- (14) 【試験に受かるために、目的】【必死に勉強した。】
- (15) 【都会から離れて、】【体をリフレッシュする。目的】

^{*1} PDTB 2.0 および 3.0 の“CONTINGENCY.Cause”に相当する。

^{*2} PDTB 2.0 では順方向の“原因・理由”を“CONTINGENCY.Cause.result”、逆方向の“原因・理由”を“CONTINGENCY.Cause.reason”としている。PDTB3.0 では前述のタグに加え“CONTINGENCY.cause+belief”も“原因・理由”に相当する。

^{*3} PDTB 3.0 の“CONTINGENCY.purpose”に相当する。PDTB 2.0 では“目的”を直接的に示すタグは存在しない。

3.1.3 条件

「～すれば」「～すると」のような条件を示す談話標識が使われている、もしくは条件を示す談話標識を節ペアに挿入しても意味的なつながりが変化しない場合、“条件”とする*4。この関係をもつ節ペアにも方向があり、一般に、一方の節が条件、他方の節が結果となる。例を以下に示す。

(16) 【酒を飲むと 条件】【すぐ顔が赤くなる。】

(17) 【雨が降れば、 条件】【道がぬれる。】

例 (17) のように仮定的な条件の場合も“条件”とする。

3.1.4 根拠

“原因・理由”、“目的”、“条件”には該当しないが、節ペアの関係が根拠に基づく推量や認識と解釈できれば、“その他根拠”とする*5。この関係をもつ節ペアには方向があり、一般に、一方の節が根拠、他方の節が推量や認識となる。例を以下に示す。

(18) 【試験に受かったのだから、 根拠】【必死に勉強したのだろう。】

(19) 【ツバメは、巣材集めの真っ最中のようにです。】【毎日、泥や枯れ草をくちばしにくわえて飛び立っていきます。 根拠】

3.2 逆接系

3.2.1 対比

「～の一方で～」のような対比を示す談話標識が使われている、もしくは対比を示す談話標識を節ペアに挿入しても意味的なつながりが変化しない場合、“対比”とする*6。一方の節が他方の節を含意している場合も含む。この関係をもつ節ペアは方向をもたないため、本稿では前節と後節の間に談話関係タグ“対比”を記す。例を以下に示す。

*4 PDTB 2.0 の“CONTINGENCY.Condition”に相当する。PDTB 3.0 では前述のタグに加え、否定条件を示す“CONTINGENCY.negative-condition”も“条件”に相当する。

*5 PDTB 2.0 および 3.0 では中位・下位ラベルを伴わない“CONTINGENCY”としてアノテーションされている。

*6 PDTB 2.0 および 3.0 の“COMPARISON.Contrast”に相当する。

- (20) 【大阪は雨だが、】対比【東京は晴れた。】
- (21) 【本施設は会員の方は使えますが、】対比【非会員の方は使えません。】

3.2.2 逆接

節ペアの関係を表す際、「～しかし～」のような逆接を示す談話標識や「確かに～だが～」のような譲歩を示す談話標識が使われている、もしくは逆接や譲歩を示す談話標識を節ペアに挿入しても意味的なつながりが変化しない場合、“逆接”とする*7。“対比”は方向をもたないが、逆接はどちらかの節に主眼があるので方向をもつ。例を以下に示す。

- (22) 【昨日は雨が降っていた。逆接】【しかし、野球の試合があった。】
- (23) 【明日は1日休みだ。】【もしかすると急に仕事が入るかもしれないが。逆接】
- (24) 【彼が来たとしても、逆接】【状況は改善されなかっただろう。】

例(24)のような仮定的な逆接表現（以降、“逆条件”と呼ぶ）の場合も“逆接”とする。

3.3 関係なしまたは弱い関係

節ペアに意味的なつながりがない、もしくは節ペアの関係が時間経過*8や、例示、補足、列挙、要約・敷衍・換言・含意、付加など*9の弱い関係、謝罪や感謝、話題導入を示す定型表現である場合、“関係なしまたは弱い関係”とする。例を以下に示す。

- (25) 【毎日暑い日が続きますね。】【父の手術も無事に終わり、】少しだけほっとします。
- (26) 【これ以外の環境ではご利用いただけない可能性がありますので、】【あしからずご了承ください。】
- (27) 【画像を利用する際にコメントなどを頂戴し】【ありがとうございます。】
- (28) 【コンビニのドル箱商品といえば】【やはり「おにぎり」です！】

例(26)は“原因・理由”とも解釈できるが、後節の「ご了承ください」が定型表現なので“

*7 PDTB 2.0 および 3.0 の“COMPARISON.Concession”に相当する。

*8 PDTB 2.0 および 3.0 の“TEMPORAL”に相当。

*9 PDTB 2.0 および 3.0 の“EXPANSION”に相当。

関係なしまたは弱い関係”とする。同様に例 (27) は後節が定型表現、例 (28) は文全体が話題導入を示す定型表現なので“関係なしまたは弱い関係”となる。

4 その他の基準

節間の談話関係を定義する際、談話関係の判断が難しい例がいくつか存在する。そこで、談話関係を付与する際の基準を具体例とともに示す。

4.1 優先規則

複数の談話関係が考えられ、かつ順方向と逆方向の二つの関係が考えられる場合には、順方向の関係を優先する。また、複数の談話関係が考えられ、かつ談話関係の方向が同じとなる場合は、談話標識を手がかりにタグを決定する。

- (29) 【まずは気軽に相談してください。条件】【臨床心理士などが、いつでも相談にのります。】

例 (29) は前節を条件とみなした順方向の“条件”と、後節を前節の理由とみなした逆方向の“原因・理由”の両方の解釈ができる。このような場合は順方向の関係を優先し、順方向の“条件”を付与する。

- (30) 【生乳本来の風味と栄養を大切にするため、目的】【低温殺菌を行っているおいしい牛乳です。】

例 (30) は前節が「低温殺菌を行っている」の目的とみなした順方向の“目的”と、「生乳本来の風味と栄養を大切にする」から「美味しい牛乳です」とみなした順方向の“原因・理由”の両方の解釈ができる。この例では、前節に「～するため」という目的を示す談話標識が含まれるため、順方向の“目的”を付与する。

4.2 三段論法

$A \rightarrow B \rightarrow C$ の三段論法の A, C 間は原則、「関係なしまたは弱い関係」とする。ただし、A, C 間に直接、「根拠・条件」関係が認められる場合にはその関係を付与する。

- (31) 【強風が吹いたので、原因・理由】【傘が壊れた。】また傘を買った。

- (32) 強風が吹いたので、【傘が壊れた。原因・理由】【また傘を買った。】
- (33) 【強風が吹いたので、】傘が壊れた。【また傘を買った。】
※「強風が吹いた」ので「傘を買った」とは考えられないので、「関係なしまたは弱い関係」とする。

次に、A, C 間に直接、「根拠・条件」関係が認められる例を示す。

- (34) 【たった幅107。5mm×奥行き68。0mm×高さ13。5mmというサイズで非常に小型のスイッチングハブです。原因・理由】【場所を取らないので、】設置場所を選びません。
- (35) たった幅107。5mm×奥行き68。0mm×高さ13。5mmというサイズで非常に小型のスイッチングハブです。【場所を取らないので、原因・理由】【設置場所を選びません。】
- (36) 【たった幅107。5mm×奥行き68。0mm×高さ13。5mmというサイズで非常に小型のスイッチングハブです。原因・理由】場所を取らないので、【設置場所を選びません。】
※「非常に小型のスイッチングハブ」なので「設置場所を選びません」と考えられるため、「原因・理由」とする。
- (37) 【前回のコラムでは「管理職は人を管理するな。」仕事を管理せよ」と書きました。人を管理すると、どうしても個人的な好き嫌いの管理になる。【それでは組織はよんどでしまう。原因・理由】
- (38) 前回のコラムでは「管理職は人を管理するな。仕事を管理せよ」と書きました。人を管理すると、【どうしても個人的な好き嫌いの管理になる。原因・理由】【それでは組織はよんどでしまう。】
- (39) 【前回のコラムでは「管理職は人を管理するな。」仕事を管理せよ」と書きました。人を管理すると、【どうしても個人的な好き嫌いの管理になる。原因・理由】それでは組織はよんどでしまう。
※「どうしても個人的な好き嫌いの管理になる」ので「人を管理するな。」と考えられるため、「原因・理由」とする。

4.3 節末表現・括弧の扱い

「～したことが分かる」「～といわれる」のような補文表現・節末表現が同一節内に含まれることがある。このような場合は、補文表現・節末表現を無視して談話関係を付与する。また、節区切の仕様上、例(42)のように括弧の内部と外部が同一節内に含まれる場合もある。この場合は、他の節と関係をもっている括弧の内部のみ、もしくは括弧の外部のみを考慮して談話関係を付与する。

- (40) 【これは義詮が正行の人柄を慕い、原因・理由】【そばに葬るように遺言したためといわれる。】

※節末表現「～といわれる」を無視し、「原因・理由」と考える。

- (41) 【このイルミネーションをプロのカメラマンに頼んで】【写真におさめる家庭もあるとか。目的】

※「写真におさめる」ことが目的であるので、「～もあるとか」を無視し、「目的」と考える。

- (42) 【前回のコラムでは「管理職は人を管理するな。」仕事を管理せよ」と書きました。人を管理すると、【どうしても個人的な好き嫌いの管理になる。原因・理由】

※「どうしても個人的な好き嫌いの管理になる」ので「管理職は人を管理するな」と考えられるので、鉤括弧の外部を無視し、「原因・理由」と考える。

- (43) 【「思ってた数字がポンと出てきた。原因・理由】【厳しいと思ってたので」と、にこやかな表情を浮かべていた。】

※「思ってた数字がポンと出てきた」ので「にこやかな表情を浮かべていた」と考えられるので、鉤括弧の内部を無視し、「原因・理由」と考える。

5 談話関係タグ付きコーパスの構築

本節では、談話関係タグ付きコーパスの構築方法について述べる。本コーパスでは、談話関係タグは文書中に含まれる全ての節ペアに対して、2種類のアノテータがそれぞれ付与している。1つ目は言語学を習得した熟練のアノテータによるものである。このタグ付けでは高品質なアノテーションが期待される。2つ目はクラウドソーシングを用いた談話関係タグ付けである。このタグ付けでは高速かつ大規模なアノテーションが期待される。

No.0001

- 1) 青森県は“青い森”というように自然に恵まれています。
- 2) 森や山が多くあります。
- 3) いわゆる高山はありませんが、
- 4) 高いゆえに、その土地の人たちと密接にかかわってきました。

1と2の間の談話関係: ☐原因・理由 ☐目的 ☐条件 ☐その他根拠 ☐対比 ☐逆接・譲歩 ☐談話関係なし [☐順方向 ☐逆方向]

1と3の間の談話関係: ☐原因・理由 ☐目的 ☐条件 ☐その他根拠 ☐対比 ☐逆接・譲歩 ☐談話関係なし [☐順方向 ☐逆方向]

1と4の間の談話関係: ☐原因・理由 ☐目的 ☐条件 ☐その他根拠 ☐対比 ☐逆接・譲歩 ☐談話関係なし [☐順方向 ☐逆方向]

2と3の間の談話関係: ☐原因・理由 ☐目的 ☐条件 ☐その他根拠 ☐対比 ☐逆接・譲歩 ☐談話関係なし [☐順方向 ☐逆方向]

2と4の間の談話関係: ☐原因・理由 ☐目的 ☐条件 ☐その他根拠 ☐対比 ☐逆接・譲歩 ☐談話関係なし [☐順方向 ☐逆方向]

3と4の間の談話関係: ☐原因・理由 ☐目的 ☐条件 ☐その他根拠 ☐対比 ☐逆接・譲歩 ☐談話関係なし [☐順方向 ☐逆方向]

特記事項（迷った点、節区切りがおかしい等の問題点をご記入ください。）

データを送信する

図 1 専門家による談話関係タグ付けのタスク画面

5.1 専門家による談話関係タグ付け

言語学を習得した熟練のアノテータ（以降、専門家と呼ぶ）3名が、談話関係タグと談話関係の方向のアノテーションを行った。このアノテーションを以降、専門家セットと呼ぶ。専門家セットは京都大学ウェブ文書リードコーパスから文書が無作為に抽出した380文書からなる。専門家セットの談話関係タグは、専門家による多数決によって決定した。全員の意見が一致しなかった場合は、合議により談話関係を決定した。なお、専門家の意見が2:1に別れた場合、少数派であった談話関係タグも付記という形でコーパスに記載している。

5.2 クラウドソーシングによる談話関係タグ付け

専門家がアノテーションを実施する場合、長い時間と莫大なコストがかかるという問題がある。この問題を解決するために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。任意の節間に成り立つ談話関係を一回のクラウドソーシングで判定するのは困難であるので、談話関係の有無の判定と、談話関係タグの判定の2段階クラウドソーシングで行う。

【…】で示されるフレーズの間に、論理的な関係（原因・理由、目的、条件、逆接・対比）があるかどうかを判断し、選択肢から選んでください。

- 【1 多摩霊園の中にある芝生の上で、週末サッカーを楽しんでいます。】
 【2 ボールを蹴って】
 【3 汗かいて…。】
 【4 愛犬と散歩する方など多く、邪魔にならない様にしています。】

- ☐ 1 と 2 の間に論理的な関係がある
☐ 1 と 3 の間に論理的な関係がある
☐ 1 と 4 の間に論理的な関係がある
☐ 2 と 3 の間に論理的な関係がある
☐ 2 と 4 の間に論理的な関係がある
☐ 3 と 4 の間に論理的な関係がある
☐ どのフレーズの間にも論理的な関係はない

図 2 クラウドソーシングのタスク画面（1 段階目）

文章中の【…】で示されることばの間に成り立つ関係について、もっとも自然なものを選択肢から1つ選んでください。

サヌールの海岸を散歩した。
【1 ビデオで撮影する私を見て】
【2 優しく微笑みをくれた。】
浜辺でお祈りをしている若いご婦人がいた。

- ☐ 原因・理由（「1したがって2」と言え、かつ「1さらに2」と言えない関係。逆向きも可。）
☐ 目的（「1するために2」と言える関係。逆向きも可。）
☐ 条件（「1すれば2」と言える関係。逆向きも可。）
☐ その他根拠（「1だから2だろう」と言える関係。逆向きも可。）
☐ 対比（「1しかし2」と言え、かつ「1そして2」と言える関係。逆向きも可。）
☐ 逆接・譲歩（「1しかし2」と言え、かつ「1そして2」と言えない関係。逆向きも可。）
☐ 上記いずれの関係もない（時間経過・定型表現・詳細化など）

図 3 クラウドソーシングのタスク画面（2 段階目）

5.2.1 1 段階目：談話関係の有無の判定

意味的なつながりを明らかに持たない節ペアを除外するために、各文書に含まれる節ペアごとに談話関係があるかないかの 2 値分類を 1 段階目で行う。まずクラウドワーカーに、ある文書に含まれるすべての節ペアを提示する（図 2）。クラウドワーカーは談話関係を持つ節ペアをすべて選択する。10 人に対して同じ設問を実施し、1 人以上が談話関係があると選択した節ペアを談話関係がある可能性があると判定した。なお、明示的である節ペアはクラウドワーカーの投票結果にかかわらず、談話関係がある可能性があると判定した。

5.2.2 ステージ 2: 談話関係タイプ判定

2 段階目では、1 段階目で談話関係がある可能性があると判定された節ペアを対象とする。まず、クラウドワーカーにアノテーション対象である節ペアとその前後の文を提示する（図 3）。クラウドワーカーは提示された節ペアに対して、表 1 の下位タイプの中から最も適切な関係を選択する。この際、談話関係の方向は無視している。10 人に対して同

じ設問を実施し、多数決によって談話関係タグを決定する。最多投票となった談話関係タグが複数ある場合は、最多投票のタグを全て付与する。なお、“談話関係なし”が最多投票となった場合は第1段階の結果を無視し、談話関係がないと判定する。

5.2.3 実験結果

Yahoo!クラウドソーシング*¹⁰を用いて、京都大学ウェブ文書リードコーパス全体(6,333 文書*¹¹、40,067 節ペア)に対してアノテーションをおこなった。この際、Yahoo!クラウドソーシングに実装されているチェック問題という機能を活用した。チェック問題とは、事前に正解を用意しておいた簡単な問題を出題し、その問題が不正解だったクラウドワーカーをタスクから除外する仕組みである。この仕組みを使用することで、クラウドワーカーが注意深く・真面目に取り組んでいるかどうかを識別することができ、その結果アノテーション品質の向上が期待できる。

アノテーションの結果、京都大学ウェブ文書リードコーパス全体の 52.1% にあたる 20,878 ペアが第1段階を通過し、そのうち 6,312 ペアに“談話関係なし”以外のタグが付与された。これは、コーパス全体の 15.7% に相当する。

*¹⁰ <http://crowdsourcing.yahoo.co.jp/>

*¹¹ 京都大学ウェブ文書リードコーパス(約 5,000 文書)に加え、諸事情により公開していない約 1,300 文書に対しても談話関係を付与している。また 6,333 文書中には専門家セットとしてアノテーションした文書も含まれている。

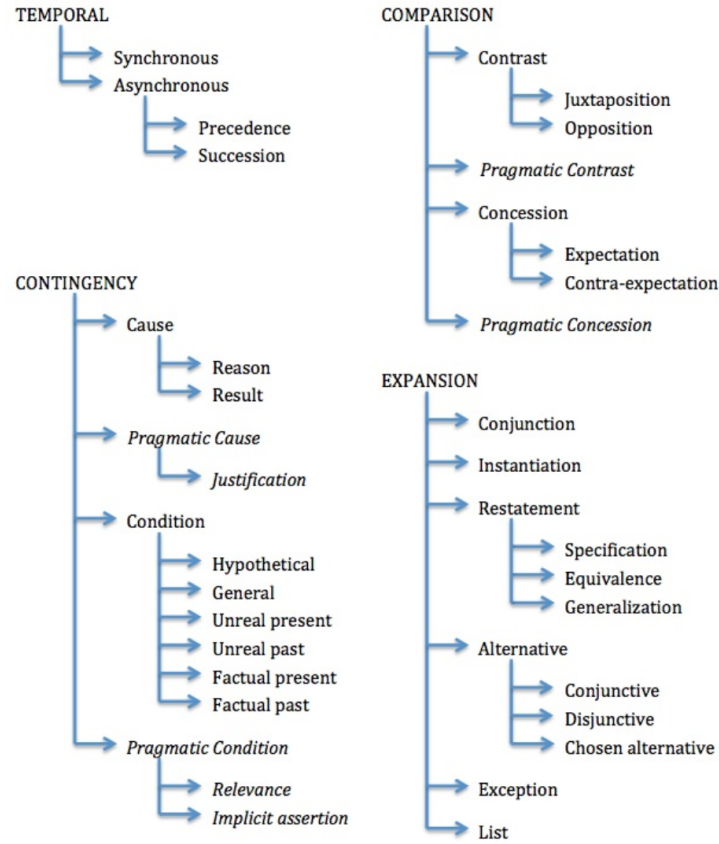


図4 PDTB 2.0 における談話関係タグセット. [4] より引用.

参考文献

- [1] Koji Murakami, Eric Nichols, Suguru Matsuyoshi, Asuka Sumida, Shouko Masuda, Kentaro Inui, and Yuji Matsumoto. Statement map: Assisting information credibility analysis by visualizing arguments. In *Proceedings of the 3rd Workshop on Information Credibility on the Web*, pages 43–50, 2009.
- [2] Rob Ennals, Beth Trushkowsky, and John Mark Agosta. Highlighting disputed claims on the web. In *Proceedings of the 19th international conference on World Wide Web*, pages 341–350, 2010.
- [3] 野矢 茂樹. 新版 論理トレーニング. 産業図書, 2006.
- [4] Rashmi Prasad, Bonnie Webber, and Aravind Joshi. Reflections on the Penn discourse TreeBank, comparable corpora, and complementary annotation. *Com-*

putational Linguistics, 40(4):921–950, December 2014.