

文節間限定関係コーパス 1.0 仕様

2013/03/13 初版(1.0)公開

大西真輝, 水野淳太, 渡邊陽太郎, 福原裕一, 乾健太郎 (東北大学)

概要

与えられた 2 文が弱対立関係 (一方が他方に条件付きで同意/対立する関係) にあるかを判定する課題を, 弱対立関係認識課題と呼ぶ. この課題において, 一文内における, 実体や事象に対して, その性質や成立の範囲を限定する文節の存在を認識することが重要である. 実体や事象を含む任意の文節に対して, 他の文節がそれを限定するかを判定する部分課題を, 文節間限定関係認識課題と呼ぶ. 本仕様書は, この課題において重要な正解事例 (限定関係にある文節ペアと, その属性) の付与基準について述べる.

文間弱対立関係の定義

2 文が与えられたときに, 一方が他方に対して, 部分的に賛成・反対しているとき, その間の関係を弱対立関係と呼ぶ [大木環美, 2010].

H1 キシリトールは虫歯予防に効果がある

T1 キシリトールの虫歯予防効果は, 悪くはない程度です

T2 キシリトールは 1日3回毎食後摂取しないと虫歯を抑制できない

上記の例について, *H1* (Hypothesis) に対する *T1*, *T2* (Text) の関係を考える. *T1* は *H1* 中の「予防効果」の程度を限定していることから, 完全に賛成はしていない. しかし, 否定しているわけでもなく, このような事例を *T1* が *H1* に「部分的に賛成している」と呼ぶ. 次に *H1* に対する *T2* の関係だが, *T2* は *H1* 中の「予防効果」を得るための条件を付加している. 下線部の条件を満たさない場合, *H1* 中の「予防効果」を得ることはできない, ということから否定的な意見ではあるが, 条件を満たせば「効果」を得られることが示唆されることから, 部分的に反対している, とする. このように, 一方が他方に対して, 「部分的に賛成/対立」している文対を「弱対立」関係にあるという.

文節間限定関係の定義

弱対立の関係にある文対では、文中の実体や事象に対して、その性質や成立に対する条件や程度の表現が存在する。これらの表現を「限定節」と呼び、それが限定している対象の実体や事象を「被限定節」と呼ぶ。限定節の種類は、条件、程度、範囲の3種類を定義する。それぞれについて順に述べる。

- 条件 (CONDITION) : 実体の性質や、事象の成立に対して、条件を示している表現。以下では、いくつかの事例に基づいて、条件を示す表現について説明する。

H キシリトールは虫歯予防に効果がある

T キシリトールは、毎食後摂ることで虫歯を予防できる

上記の例において、*T*は、「毎食後摂る」という条件を満たした場合、「予防」という事象が成立すると読み取れる。「毎食後摂る」と「予防」が限定関係にある。

H トルマリンはマイナスイオンを発生する

T トルマリンは外から力を加えなければ、マイナスイオンを発生しません

上記の例では、「外から力を加える」という条件を満たさない場合、「発生」という事象が成立しない、逆に言うと、この条件を満たした場合には「発生」という事象が成立するという限定関係になっている。下線部の条件を満たさない場合、「マイナスイオンの発生」という事象が成立しないということから否定的な意見ではあるが、条件を満たせば成立するという限定関係を示している。

H ステロイドは副作用がある

T ひとくちにステロイドと言っても副作用は起きにくく改善されていて、成人では比較的安全だとされています

「成人では」という条件を満たす場合、「比較的安全だ」という事象が成立する。しかし、「成人では」という条件を満たさない場合、程度は不明だが、「比較的安全だ」という事象が成立しない。

H フッ素は虫歯を予防することができる

T フッ素では虫歯を予防することはできない

H キシリトールは虫歯予防に効果がある

T もちろんキシリトールの摂取は虫歯予防にとって非常に大切なことですが
菌数を減少させたとしても母子感染の防止にはなりません

上記 *T* は 「菌数を減少させたとしても」という条件を満たしても、「母子感染の防止」という事象が成立しないという限定関係であり、*H* の「虫歯予防」という事象の成立に対しての条件ではない。このような場合にも **CONDITION** タグを付与する。

- 程度 (DEGREE) : 事象の成立具合 (どのくらい成立するか) の程度を限定する。主に程度副詞により限定されるが、程度を表す副詞節 (句) によっても限定される。

H キシリトールは虫歯予防になる

T1 キシリトールはかなり虫歯予防効果がある

T2 キシリトールはかなり摂取すれば虫歯を予防できる

程度副詞の中には、程度のみを限定する「純粹程度の副詞」と呼ばれるものと、程度と数量それぞれを限定する「量程度の副詞」と呼ばれるものがある。例えば、「かなり」は程度副詞に分類されているが、係り先によっては、*T1* のように「虫歯予防効果」の程度を限定する場合と、*T2* のように「摂取」の量を限定する場合がある。DEGREE ラベルは、事象の成立する具合 (程度) を限定しているものだけに付与し、量を限定しているものには SCOPE ラベルを付与する。また、*T2* 「かなり」は、条件節中にある程度副詞であり、帰結「予防できる」を係り先としないため、事象の成立具合を限定するものではない。このように、条件節中にあり、帰結を係り先としない程度表現については DEGREE ラベルを付与せず、SCOPE ラベルを付与する。

H ステロイド剤は副作用がある

T1 ステロイドは必ずしも副作用があるわけではない

T2 ステロイドの副作用の心配はまず必要ありません

T3 外用薬ステロイドなら、副作用はめったに起きません

上記 T1「必ずしも」と T2「まず」は陳述の副詞と呼ばれ、「必ずしも」は、否定と呼応し、事象の成立具合が低いことを表しており、「まず」は、「格言のムード（述語の基本形，タ形）」や「概言のムード（ようだ，らしい）」と呼応し、事象の成立具合が低いことを表している。T3「めったに」は、アスペクトの副詞と呼ばれ、一般には事態が起こる時間や事態の発生・展開のありかたを表す副詞であるが、事象が成立する頻度を表すこともある。このように程度を表している表現であれば、DEGREE（程度）ラベルを付与する。

H イソフラボンで健康維持ができる

T1 イソフラボンは広告で言われているほど健康維持に効果がない

T2 イソフラボンは驚くくらい健康を維持できる

程度副詞の他にも程度を表す副詞節（句）がある。上記のように「～くらい」（「ぐらい」）「～ほど」などの形で使用される。このような程度を表す副詞節（句）も DEGREE（程度）ラベルを付与する。

程度を限定する表現、主に程度副詞は、限定する程度性が広範であることもあり、複雑である。ひとくちに程度性を持つといっても、量、頻度、蓋然性など、程度性の種類は様々である。さらに、程度副詞間には「とても」「けっこう」「ちょっと」など、程度限定における度合いの違いもある。「キシリトールはとても虫歯を予防する」と「キシリトールはちょっとは虫歯を予防する」では、「予防」という事象の成立する蓋然性に違いが出てくる。このように、ある事象がどの程度成立するのか、その蓋然性を表す表現は、モダリティと関係してくる。現段階では、モダリティとの関係を考慮したラベルは付与せず、単に DEGREE ラベルを付与するに留める。事象成立の蓋然性を表す程度副詞とモダリティの関係については今後の課題である。

- 範囲（SCOPE）：事象の成立する対象を限定、または、実体の性質などを限定する。Hにある事象や実体に対してT側で、対象や性質について限定する事項がある場合、これにSCOPEタグを付与する。H側にある要素がT側の別の要素の対

象や性質を限定する事項になる場合、これは限定関係として認めない。を主に形容詞、数量表現、助詞「の」、とりたて助詞などが SCOPE になりやすい。

キシリトールは子供の虫歯は予防できる

「予防」という事象が成立するのは「子供」の場合のみであるという限定関係。上記は「名詞 1 + 「の」 + 名詞 2」の形で「子供の」が「歯」の限定節になっている。

H キシリトールは虫歯予防に効果がある

T1 フッ素が配合されたキシリトールは虫歯予防に効果がある

T2 高濃度のキシリトールは虫歯予防に効果がある

T3 キシリトールのガムは虫歯予防に効果がある

T1 と *T2* はそれぞれ「フッ素が配合された」や「高濃度の」などによって *H* の「キシリトール」の性質を限定している。一方 *T3* 「キシリトールの」は *H* 側にある要素（「キシリトール」）であり、このような場合「名詞 1 + 「の」 + 名詞 2」の形であっても、*T3* の「ガム」の性質を限定する事項とはせず、*T3* の「ガム」の方を *H* の「キシリトール」の性質を限定する事項として考える。

3 種類の限定関係は、いずれも事象（コト）の成立、または実体（モノ）の性質に対する何らかの限定である。限定は、その情報を含むより広い概念が想定でき、その一部概念をとりたてている。例えば、「純度の高い」という条件は、「純度の高低」という広い概念が想定でき、その一部である「高い純度」に限定している。また、「それなりに」というのは、「効果の強弱」という広い概念に対する、「それなり」という程度に限定している。

- 程度 (DEGREE) や範囲 (SCOPE) が条件 (CONDITION) を兼ねる場合

程度および範囲は、条件を兼ねている場合もある。例えば「子供に対してはキシリトールの虫歯予防は期待できる」という文において「子供に対しては」は「期待できる」に対する範囲であり、かつ条件でもある。また「製品に含まれるキシリトールが多いほど虫歯予防効果が高くなる」という文においては「多いほど」は「高くなる」に対する程

T イソフラボンの サプリメント で健康維持ができる

被限定節

限定節

Tの「イソフラボンの」は「サプリメント」の性質を限定していることから、限定節である。しかし、H側に同じ要素である「イソフラボン」が含まれている場合、これを限定節とは考えず「サプリメント」を限定節と考えてラベルを付与する。ここでは「サプリメント」がイソフラボンの性質を限定していると考えるため、「サプリメント」にSCOPEラベルを付与する。

- T側にHの事象を成立させるための条件になっている事項がある場合、条件(CONDITION)タグを付与する

H キシリトールは虫歯を予防することができる T キシリトールでだ液を

中和させておくと、 少しは虫歯予防になります

上記では、緑の点線の枠内「中和させておくと」が帰結である波線部「虫歯予防になります」の条件(CONDITION)になっている。このような場合、緑の点線の枠内「中和させておくと」にCONDITIONラベルを付与する。

- 特に条件は、どこまでの文節に付与するかを悩みやすいが、必要最低限（部分文のヘッダ）とする（緑の点線の枠内が条件となる）。

H トルマリンはマイナスイオンを発生する

T1 トルマリンは 静止した ままの 状態では ほとんど マイナスイオンを 発生しません

T2 トルマリンは 静止した ままの 状態では ほとんど マイナスイオンを 発生しません

「静止した」「ままの」「状態では」のような文節に分かれている場合、T1のように「状態では」（部分文のヘッダ）のみ、条件として付与する、あるいはT2のように「静止したままの状態では」までを一つの文節とし、条件として付与してもよい。

- 被限定節（限定される対象）には「CONFINED」と付与。

塗り薬の ステロイド
 毎食後 摂取する
 熱を加えると 発生する
 わりと 予防になる

被限定節（青色の線の枠内），条件（緑の点線の枠）によって限定される帰結部分（波線部），程度（紫の点線の枠）によって限定される帰結部分（波線部）それぞれに対し CONFINED ラベルを付与する。

- CONFINED（被限定節）が，他の CONFINED（被限定節）の条件，程度，範囲となる場合もある。その場合，該当の文節に CONFINED と限定属性の両方を付与する

H キシリトールは虫歯予防に効果がある

T キシリトールは 子供の 歯なら 虫歯予防が 期待できる

青い線の枠内「歯なら」は，オレンジの線の枠内「子供の」から見て，被限定節（CONFINED）であるが，帰結部分である波線部「期待できる」に対しては条件（CONDITION）になっている（緑の点線の枠が条件となる）。

- 唯一性（LIMIT）とモダリティ表現：今回は限定関係としてのラベルを付与せず，今後の研究のためにチェックのみし，事例を多く収集する

条件（CONDITION）・程度（DEGREE）・範囲（SCOPE）以外のタグとして，唯一性（LIMIT），モダリティ表現については，それとわかるタグを用意し，それぞれ付与

する. 以下のような表現もモダリティ表現のひとつとしてモダリティとわかるタグを付与する.

H イソフラボンは副作用などの問題がある

T 大豆イソフラボンは一般的な量を摂取するなら問題はない そうです

黄色の点線枠内「そうです (そうだ) 」がモダリティ

H ステロイドは副作用がある

T ステロイドの副作用は全員に出る わけではありません

黄色の点線枠内「わけではありません (わけではない) 」がモダリティ

モダリティ表現といえば, 「だろう」「かもしれない」「ちがいない」などの表現が一般的だが, 義務を表していると思われる「~する必要がある」の「必要がある」や, 確率, 程度を表していると思われる「~する場合がある」「~する可能性がある」の「場合がある」「可能性がある」なども広義のモダリティとして扱い, これらもモダリティ表現に含める.

限定関係タグ仕様

T の CaboCha0.64 による解析結果に対して, *H* および限定関係タグを付与する. これらのデータは, `cc_cab` ディレクトリ以下に含まれている.

1. *T* の解析結果の冒頭に, ” # confinement_reference “から始まる行に, *H* の文を記述する.
2. 被限定節の文節情報を表す行” * “で始まる行の最後に, タブ文字の後, 1 個以上の限定節の文節情報がスペース区切りで付与される. 各限定節は, 条件 : condition, 程度 : degree, 範囲 : scope に続いて, 限定節の文節 ID を 1 個以上記述される. 複数の文節 ID が記述される場合はカンマで区切られる.

また, アノテーションツール brat (<http://brat.nlplab.org/index.html>) 用の定義ファイルおよびアノテーションファイルは `cc_brat` に含まれている.

confinement_reference トルマリンはマイナスイオンを発生する

* 0 4D 0/1 2.626603

トルマリン 名詞,一般,*,*,*,* O
 は 助詞,係助詞,*,*,*,*,は,ハ,ワ,, O
 、 記号,読点,*,*,*,*,、,、,、,、,, O

* 1 4D 1/3 3.387387

静止 名詞,サ変接続,*,*,*,*,静止,セイシ,セイシ,, O
 状態 名詞,一般,*,*,*,*,状態,ジョウタイ,ジョータイ,, O
 で 助詞,格助詞,一般,*,*,*,*,で,デ,デ,, O
 は 助詞,係助詞,*,*,*,*,*,は,ハ,ワ,, O

* 2 4D 0/0 4.612550

殆ど 名詞,副詞可能,*,*,*,*,殆ど,ホトンド,ホトンド,, O

* 3 4D 1/2 0.000000

マイナス 名詞,一般,*,*,*,*,マイナス,マイナス,マイナス,, O
 イオン 名詞,一般,*,*,*,*,イオン,イオン,イオン,, O
 を 助詞,格助詞,一般,*,*,*,*,を,ヲ,ヲ,, O

* 4 -1D 1/3 0.000000 degree="2" condition="1" scope="1"

発生 名詞,サ変接続,*,*,*,*,発生,ハッセイ,ハッセイ,, O
 し 動詞,自立,*,*,*,*,サ変・スル,連用形,する,シ,シ,, O
 ませ 助動詞,*,*,*,*,特殊・マス,未然形,ます,マセ,マセ,, O
 ん 助動詞,*,*,*,*,不変化型,基本形,ん,ン,ン,, O
 。 記号,句点,*,*,*,*,。,,。,,。,, O

EOS