

# 医学生物学論文アブストラクトへの談話依存構造アノテーション ガイドライン ver. 5.0

西田 典起

理化学研究所 革新知能統合研究センター

`noriki.nishida@riken.jp`

2022 年 7 月 11 日

## 1 更新履歴

- Version 4.1 → 5.0
  1. 形容詞節を EDU としないように変更。(3.2 節)
    - 変更前: [I know the woman] [who is sitting at the chair.]
    - 変更後: [I know the woman who is sitting at the chair.]
  2. セクションレベル限定の談話関係の廃止。
    - Background: 汎用的に使用できるように変更。(4.3.9 節)
    - Findings: 廃止。Cause-Result、あるいは Verification で代替。(4.3.4 節、4.3.10 節)
  3. Verification 談話関係の追加。(4.3.10 節)
  4. Addition 談話関係、Same-Unit 談話関係の接続方針の変更(4.3.1 節、4.3.14 節)
  5. Attribution 談話関係の追加。(4.3.12 節)
  6. アノテーション方針の明文化。(5 節)
- Version 4.0 → 4.1 の更新内容
  1. Elaboration と Addition を区別するための例を追加(図 ??)。
  2. アノテーションツールの更新にともなう?? 章の変更。
    - クリップボードへのコピー対象を EDU から全文(“ROOT” や “<S>” 等は除去済み)に変更。
    - アノテーション途中でも保存できるように「途中保存」ボタンを追加(WARNING が出る)。
- 談話関係ごとにランダムサンプルを表示できるように変更。
- EDU 番号を末尾に表示。
- 文(≠ EDU)の範囲がわかりやすいように文ごとに色分け。
- 表示中のデータのファイル名を表示。
- 「ガイドライン」ボタンが新規タブを開くように変更。

### ● Version 3.0 → 4.0 の更新内容

1. Elaboration を分解 → Elaboration (詳細化、例示)、Addition (添加・累加・系列・同列)。
2. Joint の削除(Addition に統合)。
3. Attribution の削除。
4. Same-Unit チェインとの接続ルールの変更: 文内、文間を問わず常にチェインの始点を接合点とする。

## 2 背景と目的

節 (clause) や文 (sentence)、段落は相互に作用し合いながら文章全体の論旨展開を形成している。そのような文章の構造は**談話依存構造** (Discourse Dependency Structure) として表現される。談話依存構造では、文章は Elementary Discourse Units (EDUs) と呼ばれる節単位のノードに分割され、EDU 間の談話関係 (逆説、原因結果、条件など) を表すリンクに基づいて一つのグラフ構造として表現される [3, 2]。談話構造は、文書要約や文書分類、質問応答、情報抽出など様々な自然言語処理技術で有用であることが知られており、談話構造の自動解析技術はこれらの応用の精度を向上させるのに重要である。

談話依存構造の自動解析は、人手によって注釈された談話依存構造の集合を学習データとして、機械学習モデルを訓練することによって行われる。しかし、現存する談話構造データセットはごく限られた分野にしか存在せず、存在していてもそのサイズは非常に小さい。例えば、最もよく使われている RST-DT データセット [1] でもニュース記事に対する 385 件の談話依存構造データしか収録しておらず、SciDTB データセット [4] でも自然言語処理分野の論文アブストラクトに対する 798 件のデータしか収録していない。これらの数やドメインは実用的な談話構造解析器を訓練するのに十分とは言い難い。

本プロジェクトの目的は、医学生物学論文アブストラクトを対象に大規模な談話依存構造ツリーバンクを構築することである。本ガイドラインでは、本プロジェクトにおける談話依存構造の仕様について記述する。本ガイドラインは次の内容で構成される: 3 章では、Elementary Discourse Units (EDUs) への分割基準と例外について記述する。4 章では、談話依存構造における談話依存関係、木構造制約、談話関係を定義する。最後に、5 章ではアノテーションの方針について記述する。

## 3 EDU 分割

本節では、文章を **Elementary Discourse Units (EDUs)** と呼ばれる節 (clause) レベルのスパンに分割するためのルールと例外について記述する。各 EDU は一つの連続したテキスト領域であり、EDU 同士は連続しており、EDU 間にオーバーラップはないとする。

### 3.1 分割基準

本プロジェクトでは、述語または節 (clause) のタイプに基づいて EDU の分割を決定する。基本的には、文などの独立な節や、副詞的な役割を担う従属節・句を EDU とする。

#### 3.1.1 独立節、等位節

独立節 (independent clause)、等位節 (coordinating clause) は独立の EDU とする。

- (1) [John cleaned the kitchen.] [Paul vacuumed the dining room.]

- (2) [John cleaned the kitchen] [and Paul vacuumed the dining room.]

#### 3.1.2 接続詞を伴う従属節

- (3) [Although I have a plan to go back home,] [I took a ticket to Hawaii.]
- (4) [I took a ticket to Hawaii,] [because I have a plan to go back home.]

#### 3.1.3 分詞構文 (participial clause)

- (5) [Having nothing to do,] [I went to bed early.]
- (6) [The rain continued for three days,] [causing a landslide.]

#### 3.1.4 節全体を修飾する関係節

名詞句を修飾する関係代名詞、関係副詞については原則として EDU としないが、文・節全体を修飾する関係節については EDU とする。

- (7) [The rain continued for three days,] [which caused a landslide.]

#### 3.1.5 副詞的役割の to 不定詞節 (adverbial to-infinitive clause)

to 不定詞については、それが別の節の述語を修飾する (i.e., 副詞的役割を担う) 場合は EDU とする。名詞句を修飾する to 不定詞については EDU としない。

- (8) [He hurried back home] [to get his laundry in.]
- (9) [In order to get his laundry in,] [he hurried back home.]

#### 3.1.6 副詞的役割かつ動詞的要素を含む前置詞句

動詞的要素を含む前置詞句については、それが別の節の述語を修飾する場合は EDU とする。

- (10) [He plays an important role] [in building the product.]
- (11) [He figured out the location of the restaurant] [by using Google Map.]

### 3.1.7 動詞を含む相関従属節 (correlative subordinators)

- [...] [than ...]

(12) [It's a lot cheaper and quicker to buy a plan] [than to build one.]

- [... as ...] [as ...]

(13) [It was as easy] [as collecting shells at Malibu.]

- [... so/such ...] [that ...]

(14) [The problem is so vast] [that we need to try innovative solutions.]

- [... enough ...] [to ...]

(15) [A private market like this just isn't big enough] [to absorb all that business.]

- [... too ...] [to ...]

(16) [There were too many phones ringing] [to expect market makers to be as efficient as robots.]

### 3.1.8 attribution verb の補文

動詞の目的格を担う名詞節については原則として EDU としないが、例外的に、動詞が attribution verb であり、目的格が補文である場合は EDU とする。attribution verb とは、say や report、suggest、demonstrate、indicate などの that 節を目的格にとる動詞を指し、補文とはそれらの目的節をいう。

(17) [This paper reports] [that A is B.]

(18) [The experimental results suggests] [that A is B.] [and that C is D.]

### 3.1.9 ディスコースマーカーを伴う句

例外的に、ディスコースマーカーを担う句については EDU とする。表 1 にディスコースマーカーの例を載せる。

表 1: ディスコースマーカーの例

in spite of	in stead of
despite	irrespective of
regardless of	in contrast to
in comparison with	because of
due to	as a result of

(19) [In spite of the rain.] [they went out for a picnic.]

(20) [They couldn't go on a picnic] [due to the typhoon.]

## 3.2 例外

本プロジェクトは以下のタイプの従属節は EDU としない。

### 3.2.1 動詞の主語、目的語、補語 (complement) になっているもの (名詞節)

(21) [Making computers smaller often means sacrificing memory.]

(22) [To deceive him will make him mad.]

(23) [He started digging.]

(24) [We need to find a solution in our project.]

(25) [He tried to get the work done as quickly as possible.]

(26) [He made me what I am.]

(27) [He is interested in climbing Everest.]

(28) [He was cautious about making a fatal mistake.]

### 3.2.2 名詞を修飾しているもの (形容詞節)

(29) [I know the woman sitting at the chair.]

(30) [This is the book stolen by the man.]

(31) [He has a plan to go back home.]

(32) [Sleep deprivation increases the risk of committing cognitive errors.]

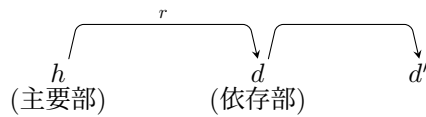


図 1: 談話依存関係 (談話係り受け関係)

- (33) [I have a friend who speaks five different languages.]
- (34) [I was born in Kyoto, which has many historical buildings.]
- (35) [I visited the office where my father works.]
- (36) [I hear the news that she is coming to Tokyo today.]
- (37) [His comment is based on the fact that sleep deprivation increases the risk of health problems.]

### 3.2.3 分裂文 (cleft)、疑似分裂文 (pseudo-cleft)、外置構文 (extraposition)

- (38) [It is sleep deprivation that exacerbates health problems.]
- (39) [What exacerbates health problems is sleep deprivation.]
- (40) [It is obvious that we cannot read the book.]
- (41) [It is difficult to read the book.]
- (42) [This book is difficult to read.]
- (43) [I found it difficult to read the book.]

### 3.2.4 非 EDU である従属節の中の従属節

- (44) [It is reported that A is B although C is Y.]

## 4 談話依存構造

本節では談話依存構造の基本仕様である談話依存関係 (4.1 節)、木構造制約 (4.2 節)、談話関係 (4.3 節) について記述する。

### 4.1 談話依存関係

談話依存構造は EDU 間の談話依存関係または談話係り受け関係 (discourse dependency) の集合とみなすことができる。図 1 は EDU  $h$  と EDU  $d$  の間の談話依存関係を表し、 $d$  が  $h$  に対して談話関係  $r$  の役割を担うことを意味する。 $h$  側を**主要部** (head) または**親**、 $d$  側を**依存部** (dependent) または**子**と呼ぶ。

人間が書いた文章は一般的に首尾一貫しているため、文章中で孤立した EDU は存在しない。したがって、文書中で最も意味的に中心的で重要な EDU を除いて、すべての EDU は依存部としてその主要部 (係り先) を一つ持つと考えてよい。

#### 4.1.1 談話依存関係の非独立性

図 1 において、 $d$  は  $d'$  を子に持つ。したがって、 $d$  だけでなく  $[d, d']$  という部分全体が  $h$  に対して談話関係  $r$  の役割を担っていると解釈する。

#### 4.1.2 従位関係 vs. 等位関係

談話依存関係は、**従位関係** (subordinating relation) と**等位関係** (coordinating relation) に分けることができる。

談話依存関係にある二つの EDU  $h$  と  $d$  の間に主従的な関係を認めることができるとき、言い換えると  $h$  なしで  $d$  が存在すると文章の自然さが損なわれるとき、 $h$  と  $d$  は従位関係であるという。従位関係では、主側を主要部、従側を依存部とする。従位関係において主要部と依存部のどちらが先に現れるかは文脈次第である。

一方、 $h$  と  $d$  が等しく重要であるとき、 $h$  と  $d$  は等位関係であるという。等位関係では、二つの EDU の間に主従の方向性はないため、便宜的に先に現れた方の EDU を主要部、後に現れる方を依存部とする (つまり、図では常に後方への矢印を描く)。三つ以上の EDU が等位的な関係である場合も、先頭の EDU を共通の主要部として、他の EDU をその依存部とすることで等位性を表現する。図 2 では、EDU  $a$  と  $[b, c, d, e]$  が従位関係であり、 $b, c, d, e$  が等位関係である例を示す。

### 4.2 木構造制約

本プロジェクトでは、談話依存構造は木構造であると仮定する。すなわち、“ROOT”と呼ばれる便宜的

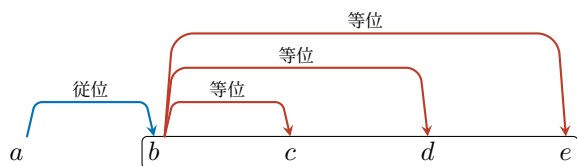


図 2: 三つ以上の EDU の等位関係の例

な EDU (以降、Root EDU と呼ぶ) を文章先頭に追加したとき、以下の制約が満たされるように談話依存構造をアノテーションする。

#### 木構造制約

- Root EDU を除くすべての EDU はその係り先 (主要部、親) を一つ持つ。
- EDU は同時に複数の係り先を持つことはできない。
- Root EDU から矢印の向きに従ってリンクを辿ることで、すべての EDU (ノード) に到達可能である (i.e., すべての EDU 間にパスが存在する)。

また、依存関係は非交差的 (projective) である必要はなく、交差的 (non-projective) であってもよいとする。

## 4.3 談話関係

本プロジェクトでは、談話関係を表 2 に示す 15 種類にカテゴライズする。本節では、各談話関係について説明する。

### 4.3.0 Root

**意味・役割:** Root は、文書中で最も意味的に中心で重要な EDU と、Root EDU との関係とする。Root EDU を主要部、文書中で最重要の EDU を依存部とする。

**制約:** Root 関係は各文書に必ず一度だけ現れる。

### 4.3.1 Addition

**意味・役割:** Addition は、主要部に対して依存部が情報の添加、累加、並列、系列であることを表す。Addition は、等位関係 (coordinating relation) のうち最も基本的で、最頻出の談話関係である。

**(補足) 主要部の決め方:** 4.1.2 節で記述したように、等位関係である Addition では、先に現れた方を主要部、後に現れる方を依存部とする。また、三つ以上の EDU が Addition の関係である場合は、先頭の EDU を共通の主要部とし、それ以外をそれに対する依存部とする。

**(補足) 子を先にまとめてから親と接続 vs. 子と親を独立に接続:** 親 EDU が複数の子 EDU によって同一の談話関係  $r$  で修飾される場合、その修飾内容によって接続方法が変わる。具体的には、その修飾が同一の観点からの修飾である場合は (子側の並列性が強い場合) は、図 3 の左のように、Addition を用いて子側の等位性を先に表現してから、その代表点 (この例では  $b$ ) を親に談話関係  $r$  で接続する。一方、その修飾が異なる観点からのものである場合は (談話関係はあくまでカテゴライズされているものなので、同じカテゴリでも細かく見ると異なる観点からの修飾であることはありうる)、図 3 の右のように、子それぞれを独立に親に談話関係  $r$  で接続する。

例:

- (45) [レモン汁を加えたスライムは液体化した。] <sub>$h$</sub>   
[お酢を加えたスライムも液体化した。] <sub>$d$</sub>
- (46) [スライムにレモン汁を垂らして、] <sub>$h$</sub>  [棒でかき混ぜた。] <sub>$d$</sub>

### 4.3.2 Elaboration

**意味・役割:** Elaboration は、主要部に対して依存部がその詳細化や例示であることを表す。Addition が情報の水平方向への拡張であるのに対し、Elaboration は情報の垂直方向への掘り下げである点に注意が必要である。Elaboration は従位関係 (subordinating relation) のうち最も基本的で、最頻出の談話関係である。

例:

- (47) [レモン汁を加えたスライムは液体化した。] <sub>$h$</sub>   
[液体化にかかる時間は 15 分だった。] <sub>$d$</sub>

### 4.3.3 Comparison

**意味・役割:** Comparison は、主要部に対して依存部が逆接、譲歩、比較などの対比的な関係であることを

表 2: 談話関係とその意味・役割、典型的なディスコースマーカ

談話関係	意味・役割	ディスコースマーカ
0. Root	文章中で最も意味的に中心で重要な EDU	
1. Addition	添加、累加、系列、同列	also, as well as, moreover, furthermore, besides, in addition, next, then
2. Elaboration	詳細化、例示	for example
3. Comparison	逆説、譲歩、対比、比較	but, however, although, yet, despite, whereas, instead of, alternatively, on the other hand
4. Cause-Result	原因、理由、結果	because, so, therefore, thus, due to, consequently, as a result, leading to
5. Condition	条件、仮定	if, as long as, unless, when
6. Temporal	時間、状況	when, before, after, while
7. Enablement	目的、可能化	in order to, for ...ing, so as to, so that, which enables to, which allows to
8. Manner-Means	方法、手段	by, using
9. Background	背景、事前知識	
10. Verification	検証パート	
11. Conclusion	結論、要約、締めくくり	in summary, in conclusion
12. Attribution	補文	attr. verb + that ...
13. Textual-Organization	タイトル、見出し、タグ等	
14. Same-Unit	分離した疑似 EDU の結合	

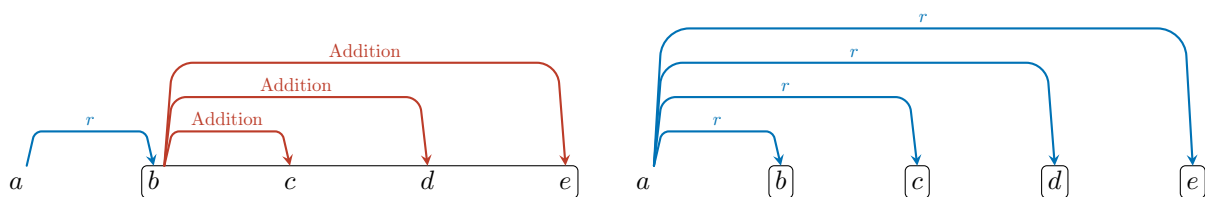


図 3: Addition の接続方法の違い。親  $a$  が子  $[b, c, d, e]$  によって同一の談話関係  $r$  で修飾されるとき、その修飾が同一の観点からなされる場合 (左) と、異なる観点からなされる場合 (右) で、接続方法が異なる。

表す。主要部と依存部が逆の情報を表す negative な対比関係だけでなく、主要部と依存部の類似性を強調するための positive な対比関係についても Comparison に含める。Comparison は従位的でも等位的でもある。

例:

- (48) [レモン汁を加えたスライムは液体化した。]<sub>h</sub>  
[一方、クエン酸を抽出したスライムを加えたスライムは液体化しなかった。]<sub>d</sub>

#### 4.3.4 Cause-Result

**意味・役割:** Cause-Result は、主要部に対して依存部がその原因、理由、または結果であることを表す。原因・理由側、結果側のどちらが主要部、依存部になるかは統語的、または意味的な重要度の差に基づいて決定する。

例:

- (49) [スライムにレモン汁を加えた。]<sub>h</sub> [その結果、スライムは液体化した。]<sub>d</sub>  
(50) [スライムにレモン汁を加えると、]<sub>h</sub> [スライムは液体化した。]<sub>d</sub>

- (51) [レモン汁はクエン酸を含んでいるので、]<sub>d</sub> [スライムは液体化する。]<sub>h</sub>
- (52) [雨が降っていたので、]<sub>d</sub> [太郎は電車出勤した。]<sub>h</sub>

#### 4.3.5 Condition

**意味・役割:** Condition は、主要部に対して依存部がその条件、仮定であることを表す。

例:

- (53) [スライムに加えるものがクエン酸を含むならば、]<sub>d</sub> [スライムは液体化する。]<sub>h</sub>
- (54) [雨が降るならば、]<sub>d</sub> [太郎は電車出勤する。]<sub>h</sub>

#### 4.3.6 Temporal

**意味・役割:** Temporal は、主要部と依存部の時間的、場所的な関係を表す。Condition が仮定的な条件を表すのに対し、Temporal は仮定的ではなく現実的に起こる (起きた) 事象と、主要部との間の時間的、場所的な関係である。時間的な関係はさらに同期的な関係 (“when”, etc.) と非同期的な関係 (“before”, “after”, etc.) に分けることができるが、本プロジェクトではこれらを同一の談話関係カテゴリーとして扱う。

**(補足) Addition との区別:** “First, ... Then, ... Finally, ...” のような系列表現では、時間的な前後関係に焦点が当てられているならば Temporal を、時間ではなく順序的な前後関係に焦点が当てられているば Addition を適用する。

例:

- (55) [スライムにレモン汁を加えたとき、]<sub>d</sub> [スライムは液体化を始めた。]<sub>h</sub>
- (56) [スライムにレモン汁を加えたあと、]<sub>d</sub> [スライムは液体化を始めた。]<sub>h</sub>
- (57) [雨が降ったとき、]<sub>d</sub> [太郎は既に帰宅していた。]<sub>h</sub>

Cause-Result の例 (50) と Temporal の例 (55) の違いは微妙だが、例 (50) では時間的な関係性が意図されてない、あるいは時間的な関係性に重きが置かれ

ていないのに対して、例 (55) では時間的な表現がされているため、Temporal を適用する。

#### 4.3.7 Enablement

**意味・役割:** Enablement は、主要部に対して依存部がその目的や可能になったことであることを表す。

例:

- (58) [スライムを液体化させるために、]<sub>d</sub> [レモン汁を加えた。]<sub>h</sub>

#### 4.3.8 Manner-Means

**意味・役割:** Manner-Means は、主要部に対して依存部がその方法、手段であることを表す。「手段」の “by” や “using ...” などのディスコースマーカを伴うことが多い。

例:

- (59) [レモン汁を加えることで、]<sub>d</sub> [スライムを液体化させた。]<sub>h</sub>

Enablement の例 (58) と Manner-Means の例 (59) を比較することで、Enablement と Manner-Means が反対の関係になっていることがわかる。つまり、Enablement と Manner-Means の違いは二つのスパンのうちどちらを中心とするかによることがわかる。

#### 4.3.9 Background

**意味・役割:** Background は、主要部に対して依存部がその背景、事前知識であることを表す。

例:

- (60) [スライムを液体化できたら素晴らしい。]<sub>d</sub> [我々はスライムを液体化できるか調べた。]<sub>h</sub>

#### 4.3.10 Verification

**意味・役割:** Verification は、仮説・目的パートを表す主要部と、その検証パートを表す依存部との間の関係とする。

例:

- (61) [我々はスライムを液体化できるか調べた。]<sub>h</sub>  
[スライムにレモン汁を加えた。]<sub>d</sub>
- (62) [我々はスライムを液体化できるか調べた。]<sub>h</sub>  
[レモン汁を加えたスライムは液体化した。]<sub>d</sub>

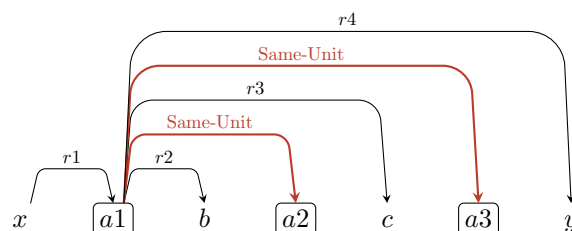


図 4: Same-Unit の接続例

#### 4.3.11 Conclusion

**意味・役割:** Conclusion は、文章全体を表す主要部と、結論や要約を表して末尾で文章を締めくくる依存部との間の関係とする。

例:

- (63) [我々はスライムを液体化できるか調べた。]<sub>h</sub>  
... [スライムはクエン酸によって液体化できることがわかった。]<sub>d</sub>

#### 4.3.12 Attribution

**意味・役割:** Attribution は、attribution verb (know, show, report, find, demonstrate, suggest, indicate, imply, etc.) を含む主要部と、その目的格の名詞節(補文)である依存部との間の関係とする。

例:

- (64) [クエン酸によってスライムが液体化できると]<sub>d</sub>  
[その結果は示唆している。]<sub>h</sub>

#### 4.3.13 Textual-Organization

**意味・役割:** Textual-Organization は、本文(主要部)と、タイトルや見出し、タグ等の本文とは別の要素(依存部)との間の関係とする。特に、医学生物学論文アブストラクトでは“Background:”や“Methods:”、“Results:”などのセクション見出しが頻繁に用いられ、そのような場合では各セクションで談話依存構造を構成し、その代表点を主要部、その見出しを依存部として Textual-Organization で接続する。

例:

- (65) [スライムの液体化についての調査 (タイトル)]<sub>d</sub> ... [本稿では、スライムを液体化できるか調べた。]<sub>h</sub> ...
- (66) [実験結果:]<sub>d</sub> [レモン汁を加えたスライムは液体化した。]<sub>h</sub> ...
- (67) [本稿では、スライムを液体化できるか調べた。]<sub>h</sub> ... [Abstract truncated.]<sub>d</sub>

#### 4.3.14 Same-Unit

**意味・役割:** 自然言語ではしばしば節の前後が逆転したり、一方の節の途中で他の節が埋め込まれるということが行われる。そのような埋め込みは、本来は一つである EDU を複数の非連続なスパンに分割してしまう。Same-Unit は、そのような非連続のスパンを結合することで本来は一つの EDU であることを表すために用いる。

**(補足) 接続方法:** Addition 等の等位関係と同様に、常に、最初に現れた要素を主要部、それ以外を依存部として、主要部と依存部の各要素を Same-Unit で独立に接続する。図 4 の例では、a1, a2, a3 は本来は一つの EDU であるが、b, c が挿入されることで三つ (a1, a2, a3) の非連続スパンに分割されている。a1, a2, a3 を接続するために Same-Unit が使われ、その主要部は a1, 依存部は a2, a3 であり、主要部から依存部の各要素へ独立にリンクを張る。

- (68) [しかし、]<sub>h</sub> [クエン酸を抽出したら、] [レモン汁を加えてもスライムは液体化しなかった。]<sub>d</sub>

## 5 アノテーション方針

現実の医学生物学論文アブストラクトは多様で曖昧性に溢れており、アノテーションの品質を一定の水準



に保つためには、現実の論文アブストラクトに対してどのように談話依存構造をアノテーションするのかわかる程度の方針を明文化しておくことは有益であろうと考える。

## 5.1 基本項目

研究とは仮説検証のことだとするならば、論文アブストラクトは大まかに以下の5項目から構成されることができる。

### 基本項目

#### 1. Background (B)

- なぜその研究 (=仮説検証) をするのか?

#### 2. Hypothesis-Objective (O)

- 何が仮説か? 何を調べたのか?

#### 3. Verification-Method (M)

- どのように検証したのか?

#### 4. Verification-Result (R)

- 検証の結果、どのようなデータが得られたか?

#### 5. Conclusion (C)

- 論文全体の結論として、どのような知見が得られたのか?

談話依存構造のアノテーションでは、各 EDU の項目にも基づいて大局的な構造を判断する。

### 5.1.1 Verification-Result vs. Conclusion

Verification-Result と Conclusion は、研究結果に基づいた知見について記述しているという点では共通・類似している。Conclusion は、研究背景や研究目的も考慮して、研究全体の結論としての知見についての記述とし、それはしばしば残された研究課題や将来の展望、政策への示唆についての記述も対象とする。一方、Verification-Result は、あくまで検証結果のデータから読み取れることについての記述とし、Conclusion に比べて知見の一般化の程度は局所的であると考えられる。

## 5.2 基本型

上述の基本項目の分類に基づき、文章の意味的中心に何を置くのかに着目すると、医学生物学論文アブストラクトの構造は大まかに以下の3つの基本型に分類できる。

### 基本型

1. 仮説中心型
2. 検証中心型
3. 結論中心型

### 5.2.1 仮説中心型

仮説中心型のアブストラクトは、その意味的中心が仮説 (Hypothesis-Objective) であり、アブストラクトの内容は、「どういうモチベーションで (Background)、どのような仮説を (Hypothesis-Objective)、どのように検証して (Verification-Method)、どのような検証結果が得られて (Verification-Result)、そこから何がわかるか (Conclusion)」という順序で記述されることが一般的である。仮説中心型は、分野を問わず最も一般的なアブストラクトの形式である。仮説中心型の例を図5に示す。

### 5.2.2 検証中心型

検証中心型のアブストラクトは、その意味的中心が検証であり、背景 (Background) の記述後、すぐに検証 (Verification-Method、Verification-Result) に入り (メインパート)、最後に結論 (Conclusion) で締め括られる。検証中心型は仮説中心型から Hypothesis-Objective を省略したものと見なせる (背景から仮説・目的が自明な場合など)。検証中心型は、仮説中心型とならんで医学生物学分野の論文アブストラクトでよく見られる形式である (情報学、少なくとも自然言語処理分野では非常に稀な形式である)。検証中心型の例を図6に示す。

### 5.2.3 結論中心型

結論中心型のアブストラクトは、その意味的中心として結論 (Conclusion) が置かれ、次いでその仮説検証の詳細 (Hypothesis-Objective、

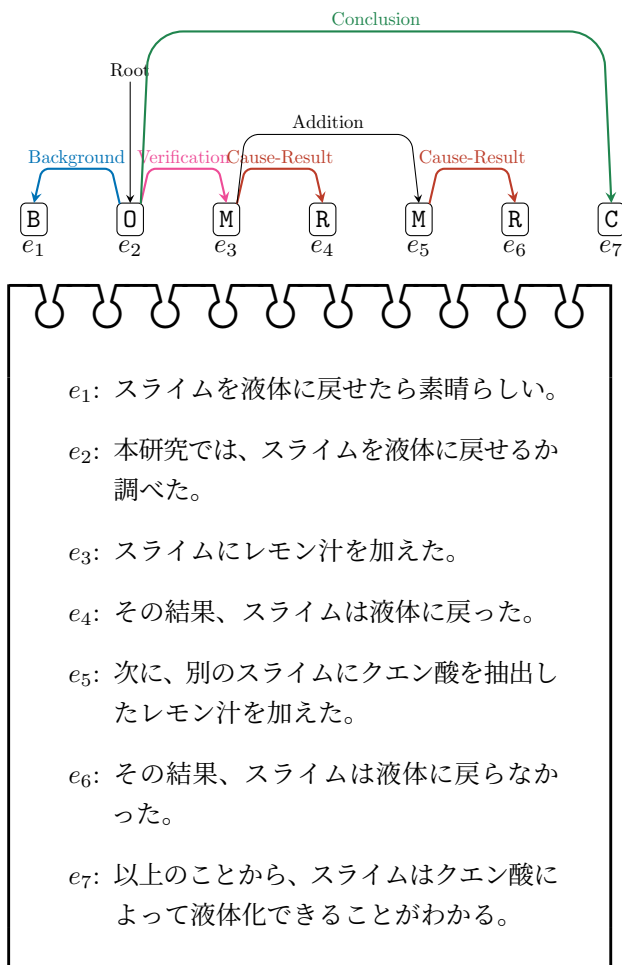


図 5: 仮説中心型

Verification-Method、Verification-Result) が記述され、最後に再び結論 (Conclusion) に戻ってくる形式をとる。結論中心型は仮説中心型の仮説 (Hypothesis-Objective) の記述を結論 (Conclusion) によって代替したものと見なせる。結論中心型の例を図 7 に示す。

### 5.3 階層優先

本プロジェクトでは、まずは文ごとに談話依存構造を付与する。文ごとのアノテーションが完了して、“Background:” などの見出しがついてるような構造化アブストラクトの場合は、各見出しの範囲 (=セクション) ごとに談話依存構造をアノテーションする。最後に、文章全体で談話依存構造を構築する。

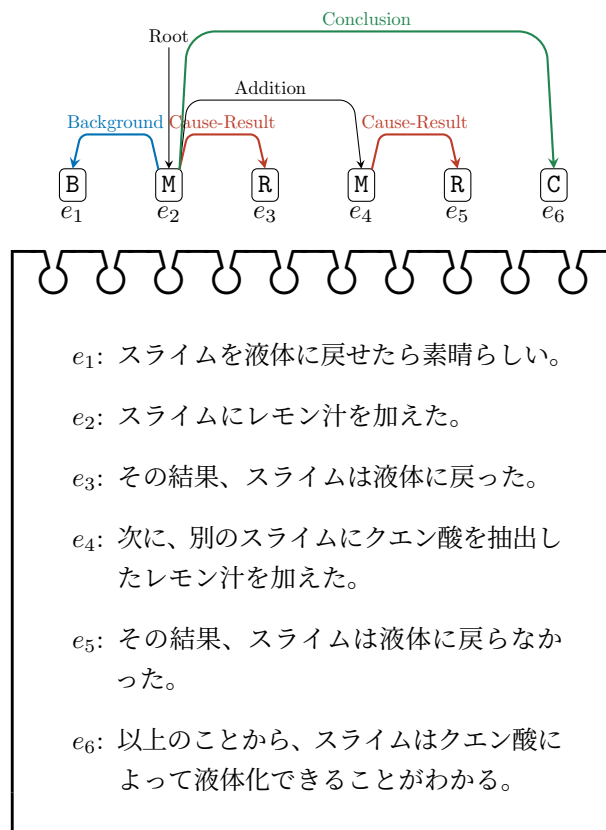
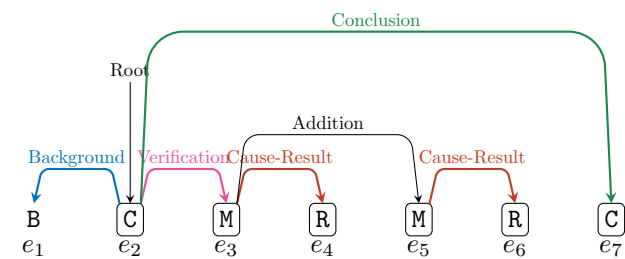


図 6: 検証中心型

## 参考文献

- [1] Lynn Carlson and Daniel Marcu. Discourse tagging reference manual. In *Technical Report ISI-TR-545*. University California Information Sciences Institute, 2001.
- [2] William C. Mann and Sandra A. Thompson. Rhetorical Structure Theory: Towards a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, Vol. 8, No. 3, pp. 243–281, 1988.
- [3] Mathieu Morey, Philippe Muller, and Nicholas Asher. A dependency perspective on RST discourse parsing and evaluation. *Computational Linguistics*, Vol. 44, No. 2, pp. 197–235, 2018.
- [4] An Yang and Sujian Li. SciDTB: Discourse dependency treebank for scientific abstracts. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 444–449, 2018.



- $e_1$ : スライムを液体に戻せたら素晴らしい。
- $e_2$ : 本研究では、スライムはクエン酸によって液体に戻せることを示す。
- $e_3$ : スライムにレモン汁を加えた。
- $e_4$ : その結果、スライムは液体に戻った。
- $e_5$ : 次に、別のスライムにクエン酸を抽出したレモン汁加えた。
- $e_6$ : その結果、スライムは液体に戻らなかった。
- $e_7$ : 以上のことから、スライムはクエン酸によって液体化できることがわかる。

図 7: 結論中心型