

医学生物学論文アブストラクトに対する談話依存構造の アノテーション・ガイドライン ver 2.1

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`

1 背景

談話依存構造 (Discourse Dependency Structure) [4, 3] は、文書中の文や節 (clause) の間の関係性 (背景、手段、目的、例示など) をもとに文書全体をグラフ構造 (木構造) として表現する。談話依存構造の例を図 1 に示す。談話依存構造の仕様については、次節で説明する。談話構造は、文書要約や極性分類、情報抽出などの自然言語処理タスクで有用であることが示されている。

論文アブストラクトの談話依存構造を収録するコーパスとして SciDTB [5] が存在する。SciDTB は自然言語処理分野の論文アブストラクト 798 件に対して人手で談話依存構造をアノテーションしている。しかし、分野によって論文の書き方や論旨の展開方法、用いられる語彙の傾向は大きく異なる。例えば、自然言語処理では手法そのものがアブストラクトの中心トピックに置かれる傾向がある一方で、医学生物学では実験・解析の結果得られた知見に重きが置かれる傾向にある。実際に、SciDTB を用いて学習した談話構造解析システムは、医学生物学の論文アブストラクトでは解析精度が著しく低下する。

本プロジェクトでは、CORD19 データセットに収録される医学生物学論文のアブストラクト 1,000 件に対して人手で談話依存構造をアノテーションし、談話依存構造解析法およびそのドメイン適応について研究する。

2 アノテーション仕様

談話依存構造のアノテーションは、一般に

1. 談話単位 (Elementary Discourse Unit; EDU) への分割
2. EDU 間の係り受けの等位・従位 (subordination, coordination) と談話関係の同定

の 2 段階で行われる。尚、今回は **EDU 分割** については既に完了しているとする。

EDU 分割済みの論文データは、1 行 1EDU のテキストファイル (*.edu.txt) として保存されている。尚、テキストは既に単語単位に分割されており (スペース区切り)、文境界、段落境界を表すマーカー “〈 S 〉”, “〈 P 〉” が対応する EDU の末尾に挿入されている。また、テキストは pdf データから自動抽出されているため、しばしば抽出エラー (単語の結合や、無意味な記号の挿入) が見られるが、これらは今回は修正しない。

作業はアノテーションツール¹

`https://norikinishida.github.io/tools/discdep/`

で行う。テキストファイル (*.edu.txt) をアップロードすることで作業を始めることができる。ツールによって出力される JSON データ (*.edu.txt.dep) をアノテーション結果とする。

¹Yang ら [5] のツールをもとに拡張して作成した。



図 1: 論文アブストラクトの談話依存構造の例。アノテーションツールによるアノテーション画面を示す。白枠で囲まれたテキストブロックは談話単位 (Elementary Discourse Unit; EDU) に対応し、青い矢印は EDU 間の談話関係の有無と等位・従位を、赤い文字は EDU 間の談話関係を表す。“< S)”, “< P)” はそれぞれ文、段落の境界を表す。

2.1 談話単位 (EDU) への分割

EDU は、節 (述語) を中心とする最小単位のテキストスパンであり、EDU 間にオーバーラップはない。文書の先頭には便宜的に Root EDU が挿入されている。EDU は節を基準に定義されるが、RST-DT の構築に用いられた Carlson ら (2001) のマニュアル [2] に則り、いくつかの例外を設定する。

- 文の main verb の主語、目的語になる節は EDU としない。
 - (1) [Making computers smaller often means sacrificing memory.]
- 述語を含まなくても、明示的なディスコースマーカーが付随している句は EDU とする。
 - (2) [They went on a picnic] [in spite of the typhoon.]

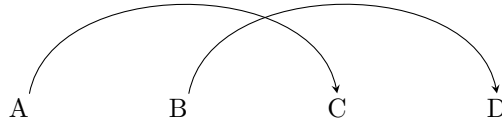


図 2: 交差する談話依存構造の例。A と C のリンクと B と D のリンクが交差している。

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

2.2 係り受けの等位・従位の同定

談話依存構造では、係り受けは親 EDU (中心部、核) からその子 (周辺部、衛星) へのリンクとして表す。本稿では、二つの EDU 間に従属的な関係性があることを**従位 (subordination)** という。一方、二つの EDU 間に等位的 (並列的) な関係性があることを**等位 (coordination)** という。従位の場合、主要な方を親、もう一方を子とする。等位の場合は二つの EDU の重みは等しいため、便宜的に前の EDU を親、後ろの EDU を子とする。

木構造制約 本プロジェクトでは、1つのアブストラクトに対して、1つの談話依存構造をアノテーションする。これは、**各 EDU (Root EDU を除く) は必ず親を 1 つだけもち、かつすべての EDU はグラフ上で連結されている (任意の 2 つの EDU 間にパスが存在する)** と定式化することができる。アブストラクト全体の中心的な EDU は Root EDU をその親としてもち、Root EDU は親をもたない。

非交差制約 また、リンク同士は交差することは少ないと考えられる。交差の例を図 2 に示す。しかし、アブストラクトによっては交差するリンクの方が適当な場合もある。

2.3 係り受けの談話関係の同定

本プロジェクトでは、談話関係として 16 種類のカテゴリーを定義する。これらは、SciDTB および RST Discourse Treebank [2], ISO 24617-8 [1] を参考に、特に将来的に科学技術論文からの情報抽出のために活用されることを念頭に設定されている。表 1 に談話関係カテゴリーの一覧を示す。以降で、各談話関係カテゴリーについて例を使って簡単に説明する。親を太字体、子を斜字体で示す。

2.3.1 Root

定義 Root は、Root EDU (親) と、文章中で最も重要な EDU (子) との間の従位的な関係とする。Root 関係は、各文書に**必ず 1 回だけ**現れるとする。“In this paper, ...” や “This study shows” などの表現を伴うことが多い。

例

- (4) [**ROOT**_h] ... [*In this paper we analyze the Twitter narratives around this decision making*_a] by applying ...
- (5) [**ROOT**_h] ... [*This review examines the recent discovery of novel influenza viruses in bats and cattle,*_a] ...
- (6) [**ROOT**_h] ... In this study we identify [*which areas in the country are most vulnerable for COVID-19, both in terms of ...*_a]

表 1: 本コーパスで採用する談話関係カテゴリー。SciDTB における談話関係との対応付けも示す。

本コーパス	SciDTB
1. Root	Root
2. Attribution (帰属)	Attribution
3. Background (背景)	Related, Goal, General
4. Cause-Result-Reason (原因・結果・理由)	Cause, Result, Evidence, Reason
5. Comparison (比較)	Comparison
6. Condition (条件・仮定・前提)	Condition, Temporal
7. Contrast (譲歩)	Contrast
8. Definition (定義・言い換え)	Definition
9. Elaboration (詳細化)	Elaboration, Aspect, Process-step, Progression, Summary
10. Enablement (目的)	Enablement
11. Evaluation-Conclusion (実験結果と結論)	Evaluation
12. Exemplification (例・要素)	Enumerate, Example
13. Joint (並列)	Joint
14. Manner-means (手段)	Manner-means
15. Same-unit (同一 EDU)	Same-unit
16. Textual-organization (テキスト構造)	

- (7) [**ROOT**_h] ... In order to develop an effective vaccine and diagnostic tools, [*we prepared UV-inactivated SARS coronavirus on a large scale under the strict Biosafety Level 3 (BSL3) regulation.*_a]

2.3.2 Attribution (帰属)

定義 Attribution は、報告内容・主張内容・認知内容を表す親 EDU と、そのソースを表す子 EDU との間の従位的な関係とする。基本的には、*attribution verb* と呼ばれる動詞 (know, say, show, demonstrate, indicate, argue, find, notice, investigate など) と *that* 節によって明示的に示される。*that* 節だけではなく、疑問詞で始まる節も対象にする。また、“according to” によってソースが明記される場合は、“according to” で始まる句 (子) との間に Attribution 関係を認める。

例

- (8) [*Results show*_a] [**that the optimal strategy changes with environment.**_h]
(9) [*Our previous study in chickens has shown*_a] [**that inhalation of a non-adjuvanted dry powder influenza vaccine formulation ...**_h]
(10) [*Overall, our results indicate*_a] [**that passive inhalation is feasible, ...**_h]
(11) Our work offers new perspectives [*by demonstrating*_a] [**that small-worldness and non-Markovianity can stabilize a classical discrete time crystal,**_h] ...

例外 ソースが明記されていない場合は、Attribution 関係は認めず、EDU 分割も行わない。

- (12) [It is said that Y is Z.]

2.3.3 Background (背景)

定義 Background は、親 EDU と、その背景を表す子 EDU との間の従位的な関係とする。背景には親 EDU の内容に関する社会的状況や学術的動向なども含む。一般的に研究背景はアブストラクトの先頭に書かれるため、アブストラクトの先頭部分に子 EDU が現れることが多い。親 EDU は Root の子 EDU であることが多い。しかし、なかには背景が書かれていないアブストラクトも存在するため注意。特徴的なディスコースマーカ―は存在しないと考えられる。

例

- (13) [*Infection with Mycobacterium tuberculosis remains a major cause of morbidity and mortality all over the world.* _d] ... [**We generated several subtypes of attenuated recombinant influenza A viruses** _h] ...
- (14) [*There is limited evidence* _d] as to how COVID-19 infection fatality rates (IFR) may vary by ethnicity. [**We combine demographic and health data for ethnic groupings in ...** _h]
- (15) [*The COVID-19 pandemic has resulted in a proliferation of clinical trials* _d] ... [**Here we have provided a foundation for cardiovascular and cardiooncology physicians** _h] ...
- (16) [*Mucosal vaccination is an effective strategy* _d] for generating antigen-specific immune responses against [**In this study, Lactovacillus plantarum strains NC8 and WCFS1 were used as oral delivery vehicles** _h] ...

2.3.4 Cause-Result-Reason (原因・結果・理由)

定義 Cause-Result-Reason は、原因と結果、または主張と理由 (根拠) を表す従位的・等位的な関係とする。“because” や “since”, “as”, “due to”, “because of”, “as a result” などのディスコースマーカ―を伴うことが多い。

例

- (17) [**These viruses resemble influenza C viruses** _h] [*because they share not only the receptor determinant,* _d] ...
- (18) [*As the outbreak of Coronavirus disease 2019 (COVID-19) is a worldwide pandemic,* _d] and it is rapidly expanding in Iran, [**real-time analyses of epidemiological data are needed to increase situational awareness and inform interventions.** _h]
- (19) [**PKR phosphorylates the α subunit of eukaryotic translation initiation factor 2 (eIF2 α),** _h] [*which in turn results in global protein synthesis inhibition and formation of stress granules (SGs).* _d]
- (20) [**The recent outbreak of severe acute respiratory syndrome (SARS)** _h] [*caused by SARS coronavirus (SARS-CoV)* _d] has necessitated as in-depth molecular understanding of the virus ...
- (21) [**Absence of a-casein protein significantly curtails secretion of other milk proteins and calcium-phosphate,** _h] [*suggesting a role for a-casein in the establishment of casein micelles.* _d]
- (22) [**For most environmental pollutants, experimental data are lacking,** _h] [*resulting in default assumptions* _d] ...

例 (21)、(22) では、親 EDU の事実が子 EDU の根拠や原因となっているので Cause-Result-Reason 関係としている。

2.3.5 Comparison (比較)

定義 Comparison は、比較される EDU 間の等位的・従位的な関係とする。“compared to” や “on the other hand”, “while” などのディスコースマーカーを伴うことが多い。

例

- (23) Where slippage is stimulated, [**the resulting products have one or more additional base(s)** _h] [*compared to the corresponding template motif.* _a]
- (24) Bactericidal assays demonstrated [**that there was a significant increase in the antimicrobial properties of the cell culture supernatants of porcine ear fibroblasts from the TG pigs** _h] [*in comparison to those from the WT pigs.* _a]
- (25) [**In previous studies, the conventional Loopamp kit for Norovirus G2 showed a relatively high detection rate,** _h] [*while that for Norovirus G1 showed a relatively low detection rate.* _a]

2.3.6 Condition (条件・仮定・前提)

定義 Condition は、親 EDU と、その条件や仮定、前提を表す子 EDU との間の従位的な関係とする。また、時間的な関係も含む。“if” や “when”, “as far as”, “, given that ...”, “before”, “after” などのディスコースマーカーを伴うことが多い。

例

- (26) [*Given the immense need for science led innovative solutions,* _a] [**this manuscript critically reviews recent literature** _h] ...
- (27) [*Where slippage is stimulated,* _a] [**the resulting products have one or more additional base(s)** _h] ...
- (28) [*In the absence of disruption to the intestinal microbiota,* _a] [**kittens are resistant to clinical signs of aEPEC carriage** _h] ...
- (29) However, [*after states reopened* _a] we find [**that Red states had higher travel-infection correlations than Blue states in all five state clusters.** _h]

2.3.7 Contrast (譲歩)

定義 Contrast は、逆説を表す従位的・等位的な関係とする。“however” や “although”, “while”, “instead”, “in spite of” などのディスコースマーカーを伴うことが多い。

例

- (30) [**The molecular mechanisms ... are poorly understood,** _h] [*but there is increasing evidence* _a] that ...
- (31) [*Although the idea ... has attracted considerable interest,* _a] [**this technique has never been demonstrated as the necessary scale.** _h]
- (32) [**Absence of a-casein protein significantly curtails secretion of other milk proteins and calcium-phosphate,** _h] ... [*In contrast, secretion of albumin ... into milk is not reduced.* _a]

- (33) ... [in that they do not encode the essential pathogenic factor nonstructural protein 1 (nsp1)_h] [and instead start with nsp2._a]

2.3.8 Definition (定義・言い換え)

定義 Definition は、親 EDU と、その定義や言い換えを表す子 EDU との間の従位的な関係とする。しばしば括弧やコロン、“i.e.”, “that is,” 等のマーカーを伴うが、それらのマーカーが付随しても厳密には定義、言い換えではない場合 (後述の Exemplification など) もあるため注意が必要である。

例

- (34) [When the word is ambiguous_h] [(there are several possible analyses for the word)_a] ...
- (35) [Context-predicting models_h] [(more commonly known as embeddings or neural language models)_a] are ...
- (36) [This process models exploratory search;_h] [a user explores a new topic_a] ...
- (37) [The major NLP challenge for personal assistants is machine understanding;_h] [translating natural language user commands into an executable representation._a]
- (38) [We apply this framework to the task of Semantic Textual Similarity (STS)_h] [(i.e., judging the semantic similarity of natural-language sentences)._a]
- (39) [We present WiBi, an approach to the automatic creation of a bitaxonomy for Wikipedia,_h] [that is, an integrated taxonomy of Wikipedia pages and categories._a]

2.3.9 Elaboration (詳細化)

定義 Elaboration は、親 EDU と、それに関するより詳細な情報を表す子 EDU との間の従位的な関係とする。Elaboration は最もスタンダードな従位的な関係である。文内の Elaboration については、構文情報が有用な指標になる。文間の Elaboration については、“moreover”, “furthermore” 等のディスコースマーカーは存在するが、多くの場合は明示的ではなく、共参照関係や文脈情報を用いて同定する必要がある。

例

- (40) [Emerging and re-emerging infectious diseases, and their determinants, have recently attracted substantial scientific and popular attention._h] [HIV/AIDS, severe acute respiratory syndrome, H5N1 avian influenza, and many other emerging diseases have either proved fatal_a] ...
- (41) [We were able to demonstrate the potency of influenza virus NS vectors_h] [to induce an M. tuberculosis specific Th1 immune response in mice._a]
- (42) [In this demo, we introduce an interactive system,_h] [which effectively applies multiple criteria analysis_a] ...

2.3.10 Enablement (目的)

定義 Enablement は、親 EDU と、その目的を表す子 EDU との間の従位的な関係とする。“in order to” や “so as to” などのディスコースマーカーや、「目的」を表す to 不定詞、for + 現在分詞によって明示的に示されることが多い。

例

- (43) [*In order to develop an effective vaccine and diagnostic tools*,_a] [**we prepared UV-inactivated SARS coronavirus on a large scale under the strict Biosafety Level 3 (BSL3) regulation.**_h]
- (44) [**The present study was carried out**_h] [*to apply the RNAi technology*_a] ...
- (45) [**Computational approach**_h] [*to identify protein-coding genes and their putative functions*_a] will help in designing experimental protocols.
- (46) [**Especially during quarantine and COVID emergence around the world there is a need of programs**_h] [*aimed to prevent acts of domestic violence*_a] ...

2.3.11 Evaluation-Conclusion (実験結果と結論)

定義 Evaluation-Conclusion は、親 EDU と、その実験結果あるいは研究の結論を表す子 EDU との間の従位的な関係とする。一般的に実験結果および結論はアブストラクトの後半に書かれるため、アブストラクトの後半部分に子 EDU が現れることが多い。親 EDU は Root の子 EDU であることが多い。また実験結果と結論の両方がアブストラクト内に記述される場合は、親 EDU からそれらにそれぞれ Evaluation-Conclusion 関係があると考えられる。特有のディスコースマーカースは存在せず、文脈的に判断する必要がある。

例

- (47) [**We study seasonal epidemic spreading in a susceptible-infected-removed-susceptible (SIRS) model on smallworld graphs.**_h] ... [*A numerical scaling analysis exhibits a diverging autocorrelation time in the thermodynamic limit,*_a] ...
- (48) ... [**we elucidate cellular and signaling mechanisms**_h] ... [*S1P1 agonism suppresses cytokines and innate immune cell recruitment in wild-type and lymphocyte-deficient mice,*_a] ...
- (49) [**This paper offers a state-of-the-art overview of the intertwined privacy, confidentiality, and security issues**_h] ... [*The paper concludes*_a] by briefly charting ...

2.3.12 Exemplification (例・要素)

定義 Exemplification は、親 EDU と、その具体例や要素 (項目) を最低 1 個以上列挙する子 EDU との間の従位的な関係とする。“such as” 等のディスコースマーカースや、Definition 同様に括弧やコロン、“i.e.”, “e.g.” 等のマーカースを伴うことが多い。また、“first”, “second” や“(1)”, “(2)” のような明示的なリスティングマーカースを伴うこともある。

例

- (50) [**Existing methods only employ the internal translation similarity**_h] [*such as content-based similarity and page structural similarity*_a]
- (51) [**It resolves the ambiguities in the main three steps of QALD**_h] [*(phrase detection, phrase-to-semantic-item mapping, and semantic item grouping).*_a]
- (52) [**This paper proposes a generic mathematical formalism for the combination of various structures:**_h] [*strings, trees, dags, graphs and products of them.*_a]
- (53) [**Wikification links each concept mention to a concept referent in a knowledge base**_h] [*(e.g., Wikipedia).*_a]

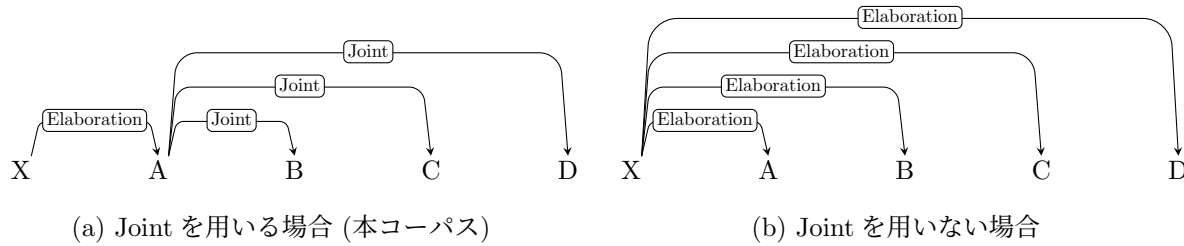


図 3: Joint 関係のアノテーション方針。EDU X に対して EDU A, B, C, D が同じ談話関係 (e.g., Elaboration) で結合し、かつ A, B, C, D が並列である場合、(b) ではなく (a) のようにアノテーションする。

- (54) [We analyze the importance of seed lexicons for the SBWES induction across different dimensions_n] [(i.e., lexicon source, lexicon size, translation method, translation pair reliability)._a]
- (55) [Here we address two limitations of this approach_n] ... [First, Web queries are rarely well-formed questions._a] ...

2.3.13 Joint (並列)

定義 Joint は並列にある EDU を表すための等位関係とする。Joint は最もスタンダードな等位関係である。Joint は常に等位的であるため、常に前の EDU を親、後ろの EDU を子とする。また、三つ以上の EDU が等位である場合、Universal Dependency にならい、**等位にある最初の EDU を親とし、後続するその他の EDU をその子とする** (図 3 (a))。例えば、X (EDU) の詳細を A, B, C, D の四つの EDU によって説明するとき (Elaboration 関係)、Elaboration(X, A), Joint(A, B), Joint(A, C), Joint(A, D) の四つの係り受けをアノテーションする。Joint を用いず、A, B, C, D それぞれが X の詳細化だとして図 3 (b) のように表現することも可能だが、アノテーションの一貫性を保つためと、A, B, C, D が並列関係にあることを表すために、図 3 (a) のように表現する。研究背景、論文の貢献、実験結果などの様々なトピック、様々な粒度で並列関係は出現する。“and” や “also”, “moreover” などの明示的なディスコースマーカを伴うこともあれば、内容で判断しなくてはならない場合もある。

例

- (56) Our method also suggests [that polyprotein 1ab, polyprotein 1a, S, M and N are proteins of viral origin_n] [and others are of prokaryotic._a]
- (57) [Absence of a-casein protein significantly curtails secretion of other milk proteins and calcium-phosphate,_n] ... [The absence of a-casein also significantly inhibits transcription of the other casein genes._a] [a-Casein deficiency severely delays pup growth during lactation_a]
- (58) I use the COVID-19 death rate in South Korea and a method ... [to estimate the total cases of COVID-19 in the U.S._n] [and to estimate the extent of infection and the unidentified share of the infected population in each of ..._a]

例 (57) では、“Absence of” の EDU, “The absence of ...” の EDU, そして “a-Casein deficiency ...” の三つの EDU が並列であり、それぞれ実験の結果について述べている。また最後の EDU については、明示的なディスコースマーカは出現しておらず、内容的に読み解く必要がある。

2.3.14 Manner-means (手段)

定義 Manner-means は、親 EDU と、そのための手段を表す子 EDU との間の従位的な関係とする。“using ...” や「手段」の “by” などのディスコースマーカを伴うことが多いが、アブストラクト中で研究目的 (おもに Root

の子 EDU) に対する具体的な方法 (Method) が述べられている場合にも、研究目的部分と方法部分との間に Manner-means 関係を認める (例 (61) 参照)。

例

- (59) [The competitiveness of transshipment hubs is then assessed_h] [*using an AHP approach.*_a]
- (60) [Relative importance of these factors is determined_h] [*based on collective views of logistics stakeholders.*_a]
- (61) [In this paper, influenza A, B and coronavirus antibody titers were measured in 257 subjects with ..._h] [*Pearson's χ^2 tests and logistic regression models were applied*_a] ...
- (62) [Our work offers new perspectives_h] by demonstrating [*that small-worldness and non-Markovianity can stabilize a classical discrete time crystal,*_a] ...

例 (62) は少し複雑であり、“Our work ...” の EDU は “by demonstraing that small-worldness ... crystal,” の部分スパンと Manner-means 関係によって結合するが、その部分スパンはさらに “by demonstrating” と “that small-worldness ... crystal,” の二つの EDU に分割され、Attribution 関係によって that 節の EDU の方が親になるため、“One work ...” の EDU と “that small-worldness ...” の EDU の間に Manner-means 関係をアノテーションする。

2.3.15 Same-unit (同一 EDU)

定義 他の談話関係とは異なり、Same-unit は便宜的なカテゴリである。常に前方向 (上から下へ) のリンクとする。

今回のアノテーション仕様では、名詞句を後ろから修飾する節 (e.g., 関係節) が独立した EDU として分割されるため、もともとは一つであるはずの EDU が二つのスパンに分離してしまうことがある。例えば、“Charles Lutwidge Dodgson, better known by his pen name Lewis Carroll, was an English writer of children’s fiction.” という文は、“better known by his pen name Lewis Carroll,” が一つの EDU として分割されるため、以下のように A, B, C の三つの EDU に分割される。

- (63) [Charles Lutwidge Dodgson,]_A [better known by his pen name Lewis Carroll,]_B [was an English writer of children’s fiction.]_C

しかし、A と C はそれぞれ EDU の要件を満たしておらず、B が挿入されなければもともとは “Charles Lutwidge Dodgson was an English writer of children’s fiction.” という一つの EDU である。Same-unit は、A と C のように分割されてしまったスパンを同一の EDU として結合するための関係カテゴリである。

Same-unit は EDU 分割時に決定される関係であるため、EDU 分割時にマーカー “(SU – X)” を EDU の先頭に挿入している。X は 2 以上の整数であり、マーカーが挿入されている EDU を子、そこから X 個前の EDU を親として、Same-unit 関係をアノテーションする。文よりも大きい EDU は存在しないため、Same-unit は同一文内の EDU 間でのみ生じる。

例

- (64) [the only manual annotations_h] needed for training [*are grammatical error labels.*_a]
- (65) [The experimental results_h] using open benchmarks [*demonstrate the effectiveness of the proposed method.*_a]
- (66) [A vote prediction system_h] that exploits only textual information [*can be improved significantly*_a]

2.3.16 Textual-organization (テキスト構造)

定義 Textual-organization は、アブストラクト本文中のテキスト (親 EDU) と、アブストラクト中に現れるサブタイトルや記号 (子 EDU) を結合するための従位的な関係である。Textual-organization は SciDTB では定義されていない。しかし、今回対象にする医学生物学論文では、しばしばアブストラクト中に “Background” や “Method”, “Results”, “Conclusion” などのサブタイトルが現れる。また、本コーパスが対象とする論文アブストラクトは pdf から自動抽出されたものであるため、しばしば抽出エラーによるノイズが含まれる (出版番号など)。

例

- (67) [*Background*:_d] [**In this study we evaluated the RespoCheck Mycoplasma triplex real-time PCR for ...** _h]
- (68) [*OBJECTIVE*:_d] [**We quantitatively examined the relationship between PERC toxicokinetics and toxicodynamics at the population level** _h] ...
- (69) [*Methods*:_d] [**Influenza A, B and coronavirus antibody titers were measured in 257 subjects with ...** _h]

参考文献

- [1] Harry Bunt and Rashmi Prasad. ISO DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations. In *Proceedings of 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-12)*, 2016.
- [2] Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, 2001.
- [3] 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.
- [4] 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.
- [5] 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*, 2018.