

# 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 2.0

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`

## 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。

科学技術論文の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、医学生物学と自然言語処理では論文の書き方や論旨の展開、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。

本プロジェクトの目的は、医学生物学の論文アブストラクトの談話依存構造解析システムの実現に向けて、CORD19 データセットに収録される論文アブストラクト 1,000 件に対して人手で談話依存構造を付与することである。

## 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 間の談話係り受け関係の同定

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、後続の後処理のために単語分割がされている。テキストファイル (\*.edu.txt) をツール

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

にアップロードすることで、作業を開始することができる。アノテーション結果は、ツールによって出力される JSON データ (\*.edu.txt.dep) として保存する。

### 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。例えば、図 1 では、アブストラクトは 11 個の EDU に分割されており、文書の先頭に便宜的に 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.]

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



図 1: 論文アブストラクトの談話依存構造の例。本プロジェクトで構築したアノテーションツールによるアノテーション画面を表示する。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 間の係り受け関係の有無と方向性を、赤い文字は結合される EDU 間の係り受け関係の種類を表す。“< S >”, “< P >” はそれぞれ文、段落の境界を表す。

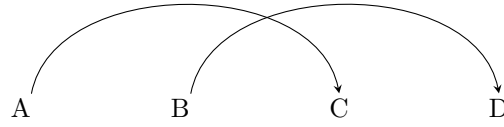


図 2: 交差する談話依存構造の例。A と C のリンクと B と D のリンクが交差している。

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・前提)	Condition, Temporal
7. Contrast (譲歩)	Contrast
8. Definition (定義)	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)	

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。

## 2.2 EDU 間の談話係り受け関係の同定

談話係り受け関係で結ばれる 2 つの EDU のうち、中心的な役割を担う側の EDU を親 (中心部、核)、もう一方の EDU を子 (周辺部、衛星) と呼ぶ。談話係り受け関係は親から子へのリンクとして描く。

1 つのアブストラクト (文書) に対して、1 つのグラフ構造 (木構造) を解析する。これは、(Root EDU を除く) 各 EDU 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。

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。

談話関係としては、SciDTB および RST Discourse Treebank [2], ISO 24617-8 [1] を参考に 16 種類のカテゴリーを定義する。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簡単に説明する。親を太字体、子を斜字体で示す。

### 2.2.1 Root

**定義** Root は、Root EDU (親) と、文書中で意味的に最も中心的な役割を担う EDU (子) との間の関係とする。一般的に、Root 関係における子はアブストラクト中で最も重要な箇所である。Root 関係は、各文書に**必ず 1 回**だけ現れるとする。しばしば、“In this paper, ...” や “This study shows” などの表現によって明示的に示される。

例

- (4) [**ROOT** <sub>head</sub>] ... [*In this paper, we propose to combine the output from a classification-based system and an SMT-based system* <sub>dep</sub>] ...
- (5) [**ROOT** <sub>head</sub>] ... [*This paper presents a negative result on unsupervised domain adaptation for POS tagging.* <sub>dep</sub>] ...
- (6) [**ROOT** <sub>head</sub>] ... [*We describe our initial efforts* <sub>dep</sub>] ...
- (7) [**ROOT** <sub>head</sub>] ... [*To this end, we propose a novel method* <sub>dep</sub>] ...

### 2.2.2 Attribution (帰属)

**定義** Attribution は、報告内容・主張内容・認知内容を表す親と、そのソースを表す子との間の関係とする。基本的には、*attribution verb* 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) [*X shows* <sub>dep</sub>] [**that Y is Z.** <sub>head</sub>]
- (9) [*X argues* <sub>dep</sub>] [**that Y is Z.** <sub>head</sub>]
- (10) [*X found* <sub>dep</sub>] [**that Y is Z.** <sub>head</sub>]
- (11) [*X noticed* <sub>dep</sub>] [**that Y is Z.** <sub>head</sub>]
- (12) [*X investigated* <sub>dep</sub>] [**why Y is Z.** <sub>head</sub>]
- (13) [*X investigated* <sub>dep</sub>] [**whether Y is Z.** <sub>head</sub>]
- (14) [**Y is Z** <sub>head</sub>] [*according to X.* <sub>dep</sub>]

**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。

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

### 2.2.3 Background (背景)

**定義** Background は、親と、その背景情報を表す子との間の関係とする。「背景」には研究の動機となった社会的状況や、既存研究等の学術的動向、または開発した技術が基づいた基盤技術を含む。一般に、アブストラクトの先頭部分に対して現れる。“based on” などのディスコースマーカーを伴うこともあるが、多くの場合は明示的には示されない。また、他の談話関係に比べ、周辺部が中心部よりも前方に位置することが多い。

例

- (16) [*Dependency parsing is a core task in NLP,* <sub>dep</sub>] ... [**We present a new GFL/FUDG-annotated Chinese treebank with more than 18K tokens from Sina Weibo** <sub>head</sub>]
- (17) [*Microblog has become a major platform for information about real-world events.* <sub>dep</sub>] ... [**In this study, we focus on the problem of community-related event detection by community emotions.** <sub>head</sub>]

#### 2.2.4 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) との間の関係とする。原因と結果の場合、どちらが中心部、周辺部になるかは文脈次第であり、文書中でより重要な方を親とする。主張と理由 (根拠) の場合、理由 (根拠) 側を常に子とする。“because” や “since”, “as”, “due to”, “because of”, “as a result” などのディスコースマーカーを伴うことが多い。しかし、しばしば文脈的に判断する必要がある。

例

- (18) [**This is usually problematic** <sub>head</sub>] [*because lexical ambiguity is ubiquitous,* <sub>dep</sub>]
- (19) [**existing models may not do the prediction task well** <sub>head</sub>] [*due to their weakness in sentiment extraction.* <sub>dep</sub>]
- (20) [**Chat language is different from natural language** <sub>head</sub>] [*due to its anomalous and dynamic natures,* <sub>dep</sub>] ...
- (21) [**Language transfer, the characteristic second language usage patterns** <sub>head</sub>] [*caused by native language interference* <sub>dep</sub>] is investigated ...
- (22) [**Parsers typically suffer from the domain mismatch issue,** <sub>head</sub>] [*and thus perform poorly on social media data.* <sub>dep</sub>]
- (23) We find [**that the proposed system is robust to disfluencies,** <sub>head</sub>] [*so that a separate stage to elide disfluencies is not required.* <sub>dep</sub>]
- (24) [*It only requires consistency between training and testing.* <sub>dep</sub>] [**As a result, there is a wide range of possible preprocessing choices for data** <sub>head</sub>]
- (25) We show [**how the expected BLEU objective allows us to train a simple linear discriminative reordering model with millions of sparse features on hundreds of thousands of sentences** <sub>head</sub>] [*resulting in significant improvements.* <sub>dep</sub>]
- (26) [**the existing greedy algorithm often selects poor anchor words,** <sub>head</sub>] [*reducing topic quality and interpretability.* <sub>dep</sub>]
- (27) [**Our work is novel** <sub>head</sub>] [*in that it explicitly addresses the need* <sub>dep</sub>] ...
- (28) [**In general, the recognition problem is undecidable for unification grammars.** <sub>head</sub>] ... [*the problem is computationally hard.* <sub>dep</sub>]

#### 2.2.5 Comparison (比較)

**定義** Comparison は、明確に比較される EDU 間の関係とする。比較対象として明示的に挙げられる方を子とする。後述の Contrast (譲歩) との違いに注意。“when compared with” などのディスコースマーカーを伴うことが多い。

例

- (29) [*Compared to the standard CCG parser,* <sub>dep</sub>] [**our model is more accurate.** <sub>head</sub>]
- (30) [**our ensembles yield significantly better results** <sub>head</sub>] [*when compared with state-of-the-art.* <sub>dep</sub>]
- (31) [**It is able to identify utterances with grammatical errors with an F1-score as high as 0.623,** <sub>head</sub>] [*as compared to a baseline F1 of 0.350 on the same data.* <sub>dep</sub>]

### 2.2.6 Condition (条件・前提)

**定義** Condition は、親とその条件や仮定、前提を表す子との間の関係とする。“if” や “when”, “as far as”, “, given that ...’ などのディスコースマーカーを伴うことが多い。

例

- (32) [**high-accuracy sentiment analysis is only possible** <sub>head</sub>] [*if word senses with different polarity are accurately recognized.* <sub>dep</sub>]
- (33) [**Both methods,** <sub>head</sub>] [*when used with the exponential loss function,* <sub>dep</sub>] bear strong resemblance to the boosting algorithm
- (34) [**The method is adaptable to any language,** <sub>head</sub>] [*as far as resources are available.* <sub>dep</sub>]
- (35) [*Given a parallel corpus,* <sub>dep</sub>] [**semantic projection attempts to transfer semantic role annotations from one language to another** <sub>head</sub>]
- (36) [**The approach is applicable to any type of MWE in any language,** <sub>head</sub>] assuming [*that the MWE is contained in Wikitionary.* <sub>dep</sub>]
- (37) [*When the word is ambiguous,* <sub>dep</sub>] [**a disambiguation procedure must be applied.** <sub>head</sub>]

例 (36) では、少しわかりにくいですが、“assuming” と “that the MWE is contained in Wikitionary” の間には Attribution 関係があり (“assuming” が周辺部)、“that the MWE...” は “The approach is applicable...” が想定する内容なので、Condition という関係が付けられる。

### 2.2.7 Contrast (譲歩)

**定義** Contrast は、逆説にある EDU 間の関係とする。どちらの EDU が親、子になるかは文脈的に重要な方を親とする。“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が、文脈的に判断すべきケースもある。

例

- (38) [**Knowledge graphs are recently used,** <sub>head</sub>] ... [*However, few of the methods pay attention to non-entity words* <sub>dep</sub>]
- (39) [**There is rising interest in vector-space word embeddings and their use in NLP,** <sub>head</sub>] ... [*Nealy all this work, however, assumes a single vector per word type* <sub>dep</sub>]
- (40) [*Although the training objective is no longer concave,* <sub>dep</sub>] [**it can still be used to improve an initial model** <sub>head</sub>]
- (41) [*while labeled data in NLP is heavily biased* <sub>dep</sub>], [**importance weighting has seen only few applications in NLP,** <sub>head</sub>]

- (42) [**We do not trust the best or any specific query segmentation.** <sub>head</sub>] [*Instead, evidence in favor of candidate e2e are aggregated across several segmentations.* <sub>dep</sub>]
- (43) [*Rather than finding an approximate convex hull in a high-dimensional word co-occurrence space,* <sub>dep</sub>] [**we propose to find an exact convex hull in a visualizable 2- or 3-dimensional space.** <sub>head</sub>]

### 2.2.8 Definition (定義)

**定義** Definition は、親と、その定義や言い換えを表す子との間の関係とする。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) もあるため、文脈的に判断する必要がある。

例

- (44) [**When the word is ambiguous** <sub>head</sub>] [*(there are several possible analyses for the word),* <sub>dep</sub>] ...
- (45) [**Context-predicting models** <sub>head</sub>] [*(more commonly known as embeddings or neural language models)* <sub>dep</sub>] are ...
- (46) [**This process models exploratory search:** <sub>head</sub>] [*a user explores a new topic* <sub>dep</sub>] ...
- (47) [**The major NLP challenge for personal assistants is machine understanding:** <sub>head</sub>] [*translating natural language user commands into an executable representation.* <sub>dep</sub>]
- (48) [**We apply this framework to the task of Semantic Textual Similarity (STS)** <sub>head</sub>] [*(i.e., judging the semantic similarity of natural-language sentences).* <sub>dep</sub>]
- (49) [**We present WiBi, an approach to the automatic creation of a bitaxonomy for Wikipedia,** <sub>head</sub>] [*that is, an integrated taxonomy of Wikipedia pages and categories.* <sub>dep</sub>]

### 2.2.9 Elaboration (詳細化)

**定義** Elaboration は、旧情報である親と、それに対してさらに詳細な情報を付加する子との間の関係とする。他の談話関係のいずれも該当しない場合は、Elaboration とする。

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

- (50) [**We introduce a new CCG parsing model** <sub>head</sub>] [*which is factored on lexical category assignments.* <sub>dep</sub>]
- (51) [**Language identification and transliteration for Hindi are two major challenges** <sub>head</sub>] [*that impact POS tagging accuracy.* <sub>dep</sub>]
- (52) [**We investigate the possibility** <sub>head</sub>] [*to automatically generate sports news from live text commentary scripts.* <sub>dep</sub>]
- (53) [**Knowledge takes the form of lexicalized assertions** <sub>head</sub>] [*associated with open-domain classes.* <sub>dep</sub>]

- (54) [The dynamic reranking model achieves an absolute 1.78% accuracy improvement over the deterministic baseline parser on PTB, <sub>head</sub>] [*which is the highest improvement by neural rerankers in the literature.* <sub>dep</sub>]
- (55) [We introduce a new CCG parsing model <sub>head</sub>] which is factored on lexical category assignments. [*Parsing is then simply a deterministic search for the most probable category sequence* <sub>dep</sub>]
- (56) [Different approaches to high-quality grammatical error correction have been proposed recently. <sub>head</sub>] [*Most of these approaches are based on classification or statistical machine translation.* <sub>dep</sub>]
- (57) [Importance weighting is a generalization of various statistical bias correction techniques. <sub>head</sub>] [*Importance weighting has seen only few applications in NLP.* <sub>dep</sub>]
- (58) [The matrices perform better than full tensors, <sub>head</sub>] [*allowing a reduction in the number of parameters* <sub>dep</sub>]
- (59) [In this paper, we propose a novel question difficulty estimation approach <sub>head</sub>] ... [*We further employ a K-Nearest approach* <sub>dep</sub>]
- (60) [We present a new GFL/FUDG-annotated Chinese treebank with more than 18K tokens from Sina Weibo. <sub>head</sub>] [*We formulate the dependency parsing problem as many small and parallelizable arc prediction tasks.* <sub>dep</sub>]
- (61) [We propose a new Chinese abbreviation prediction method <sub>head</sub>] ... [*We introduce the minimum semantic unit* <sub>dep</sub>] ... [*We use an integer linear programming (ILP) formulation with various constraints* <sub>dep</sub>]
- (62) [We present a novel translation model <sub>head</sub>] ... [*A tree-to-string alignment template is capable of generating both terminals and non-terminals* <sub>dep</sub>] ... [*The model is linguistically syntax-based* <sub>dep</sub>] ...
- (63) [In this method, the dependency parsing is executed in two stages: at the clause level and the sentence level. <sub>head</sub>] [*First, the dependencies within a clause are identified* <sub>dep</sub>] ... [*Next, the dependencies over clause boundaries are identified stochastically,* <sub>dep</sub>] ...
- (64) [This paper proposed a method <sub>head</sub>] ... [*First, it learns decision lists from training data* <sub>dep</sub>] ... [*Then, it is augmented by feedback* <sub>dep</sub>] ... [*Finally, it detects errors* <sub>dep</sub>] ...
- (65) [we: (i) leverage content from the local neighborhood of a user; <sub>head</sub>] [(ii) *evaluate batch models as a function of size and the amount of messages in various types of neighborhood;* <sub>dep</sub>] [and (iii) *estimate the amount of time and tweets* <sub>dep</sub>]
- (66) [We describe a new algorithm for PCFG induction <sub>head</sub>] ... [*Moreover, this algorithm can work on large grammars and datasets* <sub>dep</sub>]

### 2.2.10 Enablement (目的)

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

**例**

- (67) [*In order to capture more keywords,* <sub>dep</sub>] [we also incorporate syntactic information into the CBOW model. <sub>head</sub>]



- (68) [*To address this task,* <sub>dep</sub>] [**we exploit extra-textual information** <sub>head</sub>] ...
- (69) [**The system can be used** <sub>head</sub>] [*to aid BIO-NLP directory or as useful material* <sub>dep</sub>]
- (70) [**A number of lexical association measures have been studied** <sub>head</sub>] [*to help extract new scientific terminology or general-language collocations.* <sub>dep</sub>]
- (71) [**In this paper we propose a method** <sub>head</sub>] [*to increase dependency parser performance* <sub>dep</sub>]
- (72) [**This paper proposes to apply the continuous vector representations of words** <sub>head</sub>] [*for discovering keywords from a financial sentiment lexicon.* <sub>dep</sub>]
- (73) [**We present a weakly-supervised algorithm** <sub>head</sub>] [*for harvesting semantic relations.* <sub>dep</sub>]

#### 2.2.11 Evaluation-Conclusion (実験結果と結論)

**定義** Evaluation-Conclusion は、親と、その実験結果、あるいは結論を表す子との間の関係とする。特有のディスコースマーカ―は存在せず、文脈的に判断する必要がある。

**例**

- (74) [**We propose a neural network approach** <sub>head</sub>] ... [*the proposed method marks new state-of-the-art accuracies for English POS tagging tasks.* <sub>dep</sub>]
- (75) [**In this paper, we propose to combine the output from a classification-based system and an SMT-based system** <sub>head</sub>] ... [*We achieve an F0.5 score of 39.39% on the test set of the CoNLL-2014 shared task* <sub>dep</sub>]
- (76) [**We investigate grammatical error detection in spoken language,** <sub>head</sub>] ... [*The proposed system outperforms two baseline systems on two different corpora* <sub>dep</sub>]

#### 2.2.12 Exemplification (例・要素)

**定義** Exemplification は、親と、その具体例や要素 (項目) を最低 1 個以上列挙する子との間の関係とする。“such as” 等のディスコースマーカ―や、Definition 同様に括弧やコロン、“i.e.”, “e.g.” 等のマーカ―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―を伴うこともある。

**例**

- (77) [**Existing methods only employ the internal translation similarity** <sub>head</sub>] [*such as content-based similarity and page structural similarity* <sub>dep</sub>]
- (78) [**It resolves the ambiguities in the main three steps of QALD** <sub>head</sub>] [*(phrase detection, phrase-to-semantic-item mapping, and semantic item grouping).* <sub>dep</sub>]
- (79) [**This paper proposes a generic mathematical formalism for the combination of various structures:** <sub>head</sub>] [*strings, trees, dags, graphs and products of them.* <sub>dep</sub>]
- (80) [**We analyze the importance of seed lexicons for the SBWES induction across different dimensions** <sub>head</sub>] [*(i.e., lexicon source, lexicon size, translation method, translation pair reliability).* <sub>dep</sub>]
- (81) [**Wikification links each concept mention to a concept referent in a knowledge base** <sub>head</sub>] [*(e.g., Wikipedia).* <sub>dep</sub>]
- (82) [**Here we address two limitations of this approach** <sub>head</sub>] ... [*First, Web queries are rarely well-formed questions.* <sub>dep</sub>] ... [*Second, the knowledge graph is always incomplete,* <sub>dep</sub>] ...

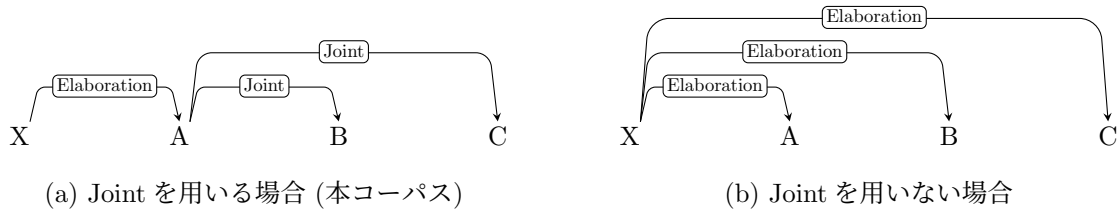


図 3: Joint 関係のアノテーション方針。EDU X に対して EDU A, B, C が同じ談話関係 (e.g., Elaboration) で結合し、かつ A, B, C が並列関係である場合、(b) の代わりに (a) のようにアノテーションする。

### 2.2.13 Joint (並列)

**定義** Joint は、2 つ以上の EDU の並列関係を表す。Joint で結合される EDU は重要性が等しく、どれか 1 つを親とするのが適当ではないが、便宜的に「最初の EDU を親として、それから他の子に Joint 関係で結合する」とする (図 3 (a))。例えば、X (EDU) の詳細を A, B, C の 3 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) の 3 つの談話関係をアノテーションする。図 3 の右図のように、Joint を用いずに、Elaboration(X, A), Elaboration(X, B), Elaboration(X, C) というようにすることもできるが、アノテーションの一貫性を保つためと、並列関係を捉えるために、Joint を用いて前者のようにアノテーションする。

例

- (83) [We propose a probabilistic model for POS guessing of unknown words<sub>head</sub>] [and estimate its parameters.<sub>dep</sub>]
- (84) [We attempt to apply the model to semisupervised learning,<sub>head</sub>] [and conduct experiments on multiple corpora.<sub>dep</sub>]
- (85) As a solution we describe a method [enlarging the vocabulary of a language model to an almost infinite size<sub>head</sub>] [and capturing their context information<sub>dep</sub>]
- (86) [MAGEAD performs an online analysis to or generation from a root+pattern+features representation,<sub>head</sub>] [it has separate phonological and orthographic representations,<sub>dep</sub>] [and it allows for combining morphemes from different dialects.<sub>dep</sub>]

### 2.2.14 Manner-means (手段)

**定義** Manner-means は、親と、そのための手段を表す子との間の関係とする。“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。

例

- (87) [We propagate the enriched features in a graph<sub>head</sub>] [using an unsupervised algorithm.<sub>dep</sub>]
- (88) [We evaluate these models<sub>head</sub>] [using customer agent dialogs from a catalog service domain.<sub>dep</sub>]
- (89) [An experiment<sub>head</sub>] [using a spoken monologue corpus<sub>dep</sub>] shows ...
- (90) [By incorporating textual information,<sub>dep</sub>] [RCM can effectively deal with data sparseness problem.<sub>head</sub>]
- (91) [Existing methods incrementally expand the lexicon<sub>head</sub>] [by greedily adding entities.<sub>dep</sub>]
- (92) [Through a simple bootstrapping procedure,<sub>dep</sub>] [we learn the likelihood of coreference between a pronoun and a candidate noun<sub>head</sub>]

### 2.2.15 Same-unit (同一 EDU)

**定義** 他の談話関係とは異なり、Same-unit はダミーとしての関係カテゴリーである。

名詞句を後ろから修飾する節が EDU として認められる都合上、もともとは 1 つの EDU が、そのような EDU 埋め込みによって 2 つの EDU に分割され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、以下のように A, B, C の 3 つの EDU に分割される。

- (93) [Charles Lutwidge Dodgson,]<sub>A</sub> [better known by his pen name Lewis Carroll,]<sub>B</sub> [was an English writer of children’s fiction.]<sub>C</sub>

これは、B (“better known by ...”) が 1 つの EDU をなすからであり、もし B がなければ A と C は 1 つの EDU (“Charles Lutwidge Dodgson was an English writer of children’s fiction.”) であり、A と C の間に談話関係はない。Same-unit は、これらのもともとは同一の EDU に属するが分割されてしまった EDU を結合するときに使う。

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。常に前方向 (左から右) のリンクとする。

#### 例

- (94) [**the only manual annotations** <sub>head</sub>] needed for training [*are grammatical error labels.* <sub>dep</sub>]
- (95) [**The experimental results** <sub>head</sub>] using open benchmarks [*demonstrate the effectiveness of the proposed method.* <sub>dep</sub>]
- (96) [**A vote prediction system** <sub>head</sub>] that exploits only textual information [*can be improved significantly* <sub>dep</sub>]

### 2.2.16 Textual-organization (テキスト構造)

**Textual-organization** 医学生物学論文では、しばしばアブストラクト中に “Background”, “Method”, “Results” などの境界マーカーが現れる。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り、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)。アブストラクトのテキストと、それらのマーカーを結合するために Textual-organization を活用する。

## 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.