Vol.2012-IFAT-105 No.6
Vol.2012-NL-205 No.6
2012/1/20

並列構造アノテーションの制約を利用した係り受けアノテーション支援

岩立将和[†1] 浅原正幸[†1] 松本裕治[†1]

機械による並列構造解析が係り受け解析より難しいのとは対照的に，並列構造アノテーション規則の方が係り受けアノテーション規則より理解しやすい．そのため，係り受け構造と並列構造がアノテーションされたコーパスを構築する際，並列構造を先に付与し，これに基づいて係り受け構造をアノテーションする方が効率的である．

本研究では，並列構造が与えられると可能な係り受け構造が制約されることを利用して高精度の係り受け解析器を構築することで，係り受けアノテーションの効率をさらに向上させる．現代日本語書き言葉均衡コーパスを用いた実験において，本手法が自動係り受け正解率を向上させることを示す．また本手法は，並列構造に関連する係り受けアノテーション誤りの検出に用いることができる．

Constraints Derived from Coordinated Structure Annotation Make Dependency Structure Annotation More Efficient

Masakazu Iwatate,[†1] Masayuki Asahara[†1]
and Yuji Matsumoto[†1]

When constructing a corpus with dependency and coordinated structures, it is efficient to annotate coordinated structures first. This is because definition of coordinated structures is more understandable than that of dependency structures and, automatic coordination analysis is less accurate than dependency parsing.

Given the coordinated structure annotation of a sentence, possible dependency structures are constrained by the coordinated structure. We propose to boost the efficiency of dependency structure annotation by construction of a precise dependency parser using the annotation. The dependency parser provides dependency structure annotators an automatic parse with fewer errors. We conducted experiments with the Balanced Corpus of Contemporary Written Japanese (BCCWJ) and show that our method improves automatic parsing accuracy and detects dependency annotation errors.

1. はじめに

コーパスアノテーションにおいては，既存のコーパスによって訓練した自動解析器を用いてアノテーション対象のコーパスに自動解析結果を付与することによって，一から人手でアノテーションするよりも労力を減らすということが一般に行われている．

我々が係り受け・並列構造アノテーションを担当している現代日本語書き言葉均衡コーパス (Balanced Corpus of Contemporary Written Japanese ; BCCWJ) は，並列構造を特殊な係り受けラベルで表現している京大コーパス[1] と異なり，係り受けと並列構造 (および同格構造) を独立したアノテーションとする[2] ことで，形態素単位の並列構造などの様々な並列構造を表現できるようにしている．

アノテーション作業においては，最初に並列構造をアノテーションした後，これと自動係り受け結果をマージした構造をアノテーターに提示して，係り受け構造をアノテーションしている．こうすることで係り受け構造と並列構造の整合性を取ることができるとともに，アノテーションを効率化していると考えられる．長い文の構造を理解するには時間がかかるが，並列構造アノテーションが見えることによって文全体の構造を短時間で理解できるようになるからである．

BCCWJ の係り受けアノテーション作業においては，このようにアノテーションの労力を削減しているものの，並列構造の情報を自動解析に用いてはいない．そこで本稿では，並列構造アノテーションから得られる情報を用いて自動係り受け解析の精度を向上させることで，人手による係り受け修正作業の労力を削減することを試みる．

2. 並列構造アノテーションと係り受けアノテーションの関連

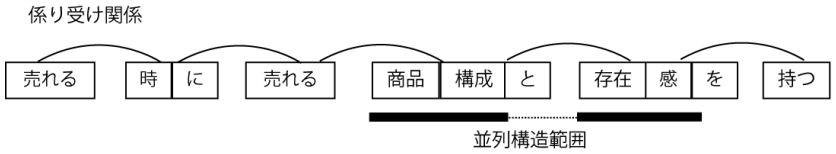
並列構造アノテーションから導かれる係り受け構造の制約については，コーパスアノテーション基準と密接な関係にある．以下作業手順とアノテーション基準について説明する．

2.1 作業手順

並列構造アノテーションと係り受けアノテーションの上流工程として，BCCWJ のコーパスデータのサンプリング，書き起こし，文区切り，形態論情報付与，文節区切りがある．

†1 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

Vol.2012-IFAT-105 No.6
Vol.2012-NL-205 No.6
2012/1/20



[出典：ＰＢ４６＿００００６]

図 1 係り受け関係と並列構造範囲
Fig. 1 Dependency relations and a coordinated structure.

上流工程の作業が終わったデータから順に，作業者１人で全データに対して，並列構造アノテーションを行う．並列構造は形態素単位 (国語研短単位) に認定し，図 1 のように並列構造範囲を表現する．図 1 中の四角１マスは国語研短単位をなし，連結されている単位が文節をなす．文節間の係り受け関係は弧で示し，基本的に最右文節を根とする木をなす．形態素単位に付与される並列構造は，並列句の範囲 (以下「構成句」と呼ぶ) を太下線で示し，その対応関係を点線で示す．アノテーターは構成句を付与し，構成句の対応関係を認定する (作業 A)．並列構造アノテーションについては，BCCWJ のコアデータ約 100 万語すべてについて，作業が完了している．また同じ方法で同格構造の範囲も付与する．

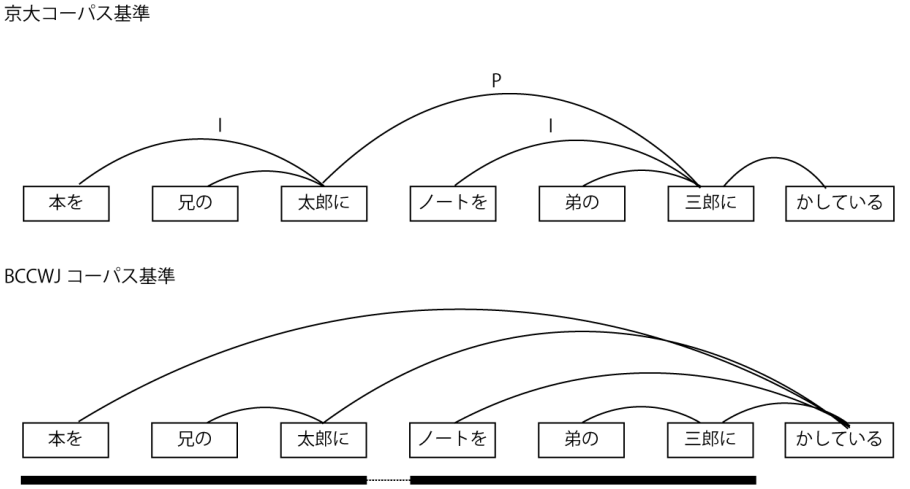
次の工程として，UniDic[3] により国語研短単位形態論情報を自動付与した京大コーパスから，係り受け情報を学習したトーナメントモデルに基づく係り受け解析器を用いて自動解析を行う．自動解析結果を作業 A で付与した並列構造アノテーションと重ね合わせ，係り受けアノテーションの修正作業を行う (作業 B)．BCCWJ で採用する係り受けアノテーション基準は 2.2 節に述べるように京大コーパスと異なる基準を採用している．作業者はこの基準の齟齬を中心に修正を行う．この作業はのべ 11 人の作業者で並行して行う．現在，この作業について BCCWJ の全体の 80% が完了している．

最後の工程として，1〜2 人の熟練した作業者により，並列構造と係り受け構造の両方のアノテーションについて一貫性などの観点から再度修正を行う (作業 C)．

なお，本研究は作業 B に関連する研究である．

2.2 BCCWJ 並列構造・係り受けアノテーション基準の概要

京大コーパスと大きく異なる点として，部分並列の扱いがある．部分並列は係り受けのラベルとして表現するのではなく，並列構造の範囲と真の係り先への係り受け関係の２つにより表現する．



[出典：黒橋 2000]

図 2 京大コーパスにおける関係ラベルと BCCWJ における並列構造の範囲付与
Fig. 2 Annotation of incomplete coordinated structure on the Kyoto Corpus and the BCCWJ.

なお，京大コーパスで並列としているテ形などの述語並列は並列とは捉えず，単に述語間の係り受けとみなす．その他，接続表現や並列構造の構成句すべてに係る左要素の扱いなどについても京大コーパスの基準と異なっている．

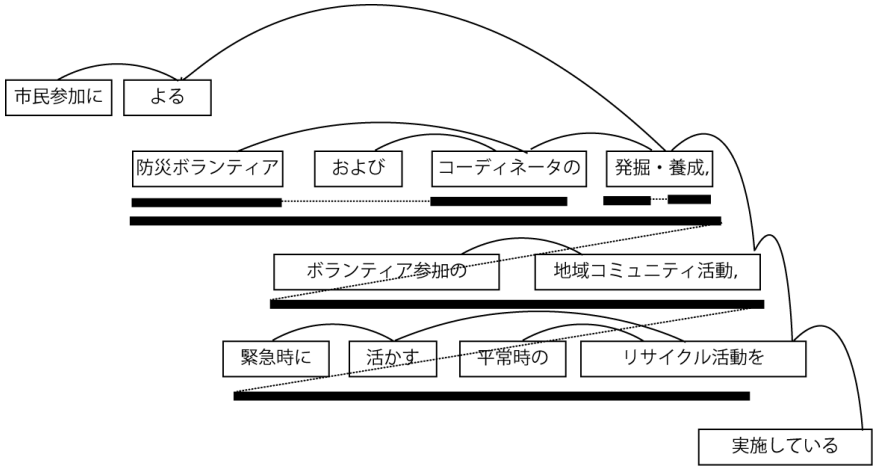
以下，基準の違いについて説明する．

2.2.1 部 分 並 列

京大コーパスでは「並列関係（P ラベル）」・「部分並列内の関係（I ラベル）」・「同格関係（A ラベル）」・「通常の係り受け関係（D ラベル）」の 4 種類を区別して係り受け関係を付与している．この係り受け関係の種別は並列・同格関係を区別するために導入されている．これに対し，BCCWJ では，係り受け関係に種別を設けない．BCCWJ においては，前述のように，別途，並列・同格構造の範囲を付与することによってこの区別を廃止する．

京大コーパスにおいて部分並列内の係り受け関係については，非交差制約を遵守するために，木構造上もっとも近い祖先に係り先を移動する手続きを行い，その移動を意味するために I ラベルが用いられていた．BCCWJ においては，並列・同格構造を範囲で対応関係とともに示し，係り先は真の係り先に係ける．この方法により，部分並列の情報を係り受け関

Vol.2012-IFAT-105 No.6
Vol.2012-NL-205 No.6
2012/1/20



[出典：OW6X_00056]

図 3 並列構造に対する各種アノテーション
Fig. 3 Various annotation rules on a complex coordinated structure.

係のラベルとして保持することを回避する．図 2 に 2 つの基準における係り受け関係基準の対比について示す．なお，以降の図では形態素境界と京大コーパス基準における D ラベルは省略する．

部分並列の廃止により，係り受け関係が交差しうるが交差を許す (非交差制約の廃止)．

2.2.2 3 つ以上の並列

図 3 の例文中「防災ボランティアおよびコーディネータの発掘・養成」と「ボランティア参加の地域コミュニティ活動」と「緊急時に活かす平常時のリサイクル活動」が並列をなし，図のように並列構造を構成する句（以下「構成句」と呼ぶ）の範囲をタグ付けする．この例では，各構成句は，最後の構成句のすぐ右にある「を」を格助詞として共有する．この場合，係り受け関係は隣の構成句に係けることとする．

2.2.3 接 続 表 現

接続表現は並列構造の構成句内の要素とせず，構成句間の要素とする．接続表現が 1 つ以上の文節をなす場合，右隣接構成句の最右文節に係ける．図 3 の例において,「および」は「コーディネータの」に係ける．

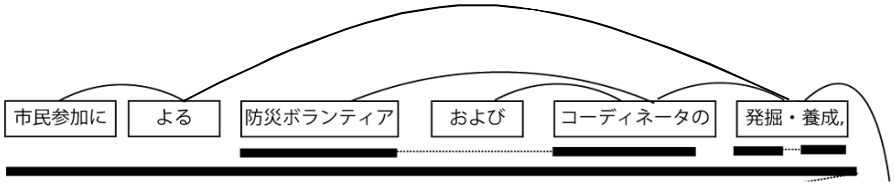


図 4 図 3 の文に対する別の並列範囲アノテーション
Fig. 4 Another coordinated structure annotation on the sentence of Figure 3.

2.2.4 並列構造の構成句すべてに係る左要素

並列構造の左から，並列構造内の構成句すべてに係る要素は，並列構造の最左構成句内の係けるべき要素に係ける．図 3 の例において,「(市民参加に）よる」は「発掘・養成,」に係ける．意味的には,「市民参加による発掘・養成」「市民参加による地域コミュニティ活動,」「市民参加によるリサイクル活動を」ということになる．

もし，図 4 のように「(市民参加に）よる」を最初の構成句に含めると,「(市民参加に）よる」が意味的に係るのは「発掘・養成,」だけとなる。したがってこの場合,「市民参加による地域コミュニティ活動,」「市民参加によるリサイクル活動を」は意味しない．図 3 と図 4 の係り受け構造はまったく同じだが，このように構成句の範囲を変えることで別の意味を表すことができる．

2.2.5 並列構造の入れ子

並列構造の範囲については入れ子（ネスト）を許す．図 3 中の 3 つの並列構造の最左要素「防災ボランティアおよびコーディネータの発掘・養成,」中には,「防災ボランティア」-「コーディネータ」と「発掘」-「養成」の 2 つの別の並列構造を含むが図のように並列構造の範囲を付与する．

京大コーパスにおいても並列構造を入れ子にすることは許されているが，一意に解釈できないという問題があった．図 5 に，2 通りに解釈できる入れ子の並列構造の例を示す．解釈 (a) が正解と考えられるが，(b) の解釈も可能である．BCCWJ においては並列構造を構成句の範囲とその対応関係で表しているため，この問題は起こらない．

2.3 アノテーションの非一貫性

現状，まだアノテーションの見直し (作業 C) が完了していないが，並列構造関連では，以下の問題事例がある．

(1) 同一の並列構造に属する 2 つ以上の構成句が同一文節を共有している事例のうち，一

Vol.2012-IFAT-105 No.6
Vol.2012-NL-205 No.6
2012/1/20

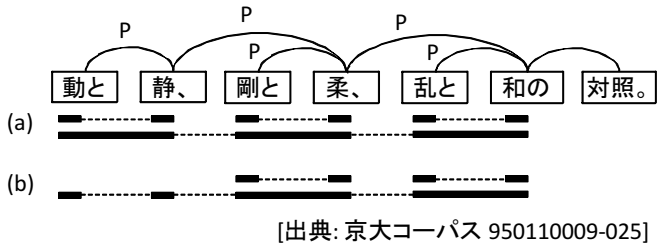


図 5 京大コーパスにおける並列の入れ子の解釈
Fig. 5 Two interpretations for nested coordinated structures on the Kyoto Corpus.

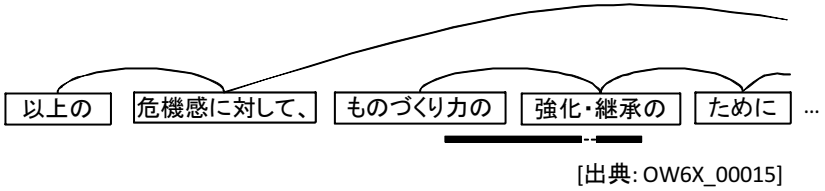


図 6 2 つ（以上）の構成句に関するアノテーション誤りの例
Fig. 6 Annotation error of spans of conjuncts.

文節内で完結している並列構造 (「発掘・養成,」など) でないもの．図 6 の例は本来一文節内で完結する並列構造だが，アノテーション誤りによって 2 文節からなる並列構造となっている．2 つの構成句「力の強化」と「継承」が文節「強化・継承の」を含んでいるため，この事例に該当する．これに該当する事例の多くはアノテーション誤り，あるいはアノテーションをマージする際などのファイル操作誤りと思われる．

(2) 2.2.2 節のアノテーション基準の不徹底．3 つ以上の構成句からなる並列構造において，各構成句の最右文節は右隣接構成句に係けるべきだが，最右構成句に係っている事例．たとえば，図 3 において「発掘・養成,」が「リサイクル活動を」に係っているような事例．

(3) 2.2.4 節のアノテーション基準の不徹底．並列構造の左から構成句内に係る（＝意味的にはすべての構成句の対応する文節に係る）ときは最左構成句に係けるべきだが，最右構成句に係っている事例．たとえば，図 3 において「(市民参加に）よる」が「リサイクル活動を」に係っているような事例．

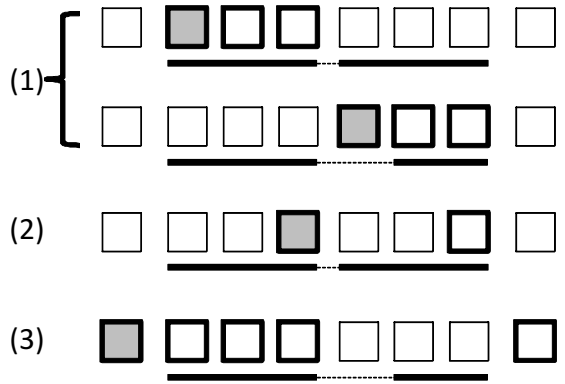


図 7 並列構造アノテーションから導かれる制約．灰色の箱:係り元文節．太線の箱:係り先候補文節．細線の箱:係ることができない文節．
Fig. 7 Constraints derived from coordinated structure annotation. Gray box is the dependent bunsetsu. Thick boxes are its possible heads. Thin boxes are not candidate heads.

3. 手　　法

日本語係り受け解析器は通常，係り元文節より右にあるすべての文節を係り先候補とみなし，この中から最適と思われる候補を選択する．[⋆1] これに対して本稿では，与えられた並列構造アノテーションから導かれる制約に基づき，各々の文節の可能な係り先を制約したうえで係り受け解析を行う．

なお，本研究で取り扱う係り受けアノテーションは文節単位であるため，一文節内で完結している並列構造については取り扱わない．また，2.3 節にあげた問題事例 (1) も取り扱わない．これらの事例は並列構造ではないものとして実験を行う．

3.1 制　　約

図 7 に示した通り，以下のように係り先を制約する．

(1)
- 構成句内の文節（構成句末尾の文節を除く）の係り先は，その構成句内の文節である．たとえば，図 3 の「緊急時に」の係り先は「活かす」「平常時の」「リサイクル活動を」のいずれかである．
- 接続表現 (2.2.3 節) の係り先は，右隣接構成句内のいずれかである．

⋆1 解析アルゴリズムの性質などの理由で，係り受け関係の交差を禁止して解析することもある．

Vol.2012-IFAT-105 No.6
Vol.2012-NL-205 No.6
2012/1/20

表 1 コーパス分割
Table 1 Corpus split.

	訓練データ [文]	テストデータ [文]	合計文節数	合計並列構造数	平均文長 [文節]	平均並列構造数
OW	4500	1326	70126	3901	12.04	0.67
OC	5000	1613	36811	786	5.57	0.12
PN	1500	445	14560	528	7.49	0.27

(2) 最右構成句以外の構成句の末尾の文節の係り先は，右隣接構成句の末尾の文節である．たとえば,「発掘・養成,」の係り先は「地域コミュニティ活動,」に一意に決まる．

(3)
- その並列構造の外部から内部に係るときは，最左構成句に係る．例) たとえば,「よる」の係り先は「防災ボランティア」〜「発掘・養成,」のいずれかか,「実施している」である.「地域コミュニティ活動,」や「リサイクル活動を」などには係ることができない．
- 並列構造の外部から接続表現に係ることは禁止する．

実験の便宜上，制約 (2) に関しては，このアノテーション規則に従っていない個所を自動修正した．[⋆2] 制約 (3) に関しては，修正先が自明でなく，自動修正することができない[⋆3]ため，修正は行っていない．

部分並列については，本研究の対象外とする．後述するとおり事例がきわめて少ないことと，これを考慮すると解析アルゴリズムが複雑になると予想されるためである．

4. 実　　験

BCCWJ の白書 (略称:OW)，Yahoo!知恵袋 (OC)，新聞 (PN) コーパス (2011 年 7 月 22 日時点のもの) を使用した．なお，OC と PN はアノテーション (作業 B) が未完了のため，本実験では完了した部分のみを使用した．文節番号が誤っているなどの，係り受け解析器を動かすために支障のある個所に対し必要な最低限の修正を行った後，コーパスを表 1 のとおり訓練データとテストデータに分割した．2 文節以下の文は，係り受け構造が一意に決まり，係り受け解析器の訓練時にも使用されないが，表 1 ではこれらの文も文数に含んでいる．

係り受け解析器には，トーナメントモデルに基づく解析器[4] を用いた．二値分類器には，Passive-Aggressive アルゴリズムに多項式カーネルを導入した[5] 実装 opal (2010 年 10 月 28 日版)[6] を用いた．2 次と 3 次の多項式カーネルについて実験した．[⋆4] 係り受け解析器の訓練においては一切係り先を制約していない．訓練時，テスト時ともに係り受け関係の交差を許して解析を行った．係り受け正解率の評価においては，各文の最後の一文節（係り先をもたない）を除く．

⋆2 後述するように，この制約に従わない特殊な並列構造が少数あり，それらに対してはこの自動修正は不適切である．

⋆3 係り先が構成句末尾の文節である事例については自動修正が可能と考えられる．

表 2 制約の係り受け正解率への影響．括弧内は制約なし，2 次の多項式カーネルとの比較．
Table 2 Unlabeled attachment scores with and without the constraints.

コーパス	カーネル	制約なし	制約 (1)(2)	全制約 (最左構成句)	全制約 (最右構成句)
OW	2 次	85.08	88.65 (+3.57)	**88.74 (+3.66)**	87.55 (+2.47)
	3 次	84.70	88.19 (+3.11)	88.33 (+3.25)	87.10 (+2.02)
OC	2 次	87.69	89.25 (+1.56)	89.21 (+1.52)	**89.87 (+2.18)**
	3 次	86.99	88.56 (+0.87)	88.57 (+0.88)	89.22 (+1.52)
PN	2 次	84.08	86.96 (+2.88)	87.05 (+2.97)	**87.83 (+3.75)**
	3 次	83.93	86.82 (+2.74)	86.88 (+2.80)	87.63 (+3.55)

4.1 自動係り受け正解率

このコーパスには制約 (3) のアノテーション誤りを含む．そのため，制約を用いない実験，制約 (1)(2) のみを用いた実験，(1)(2)(3) すべての制約を用いた実験を行った．すべての制約を用いた実験では，制約 (3) に関しては「最左構成句に係る」「最右構成句に係る」の両方について実験した．結果を表 2 に示す．

OW は文が長く並列構造が多く含まれるため，制約なしでの係り受け解析は難しく，本手法の効果が大きい．だが，OW に比べると並列構造が少ない PN は制約 (3) の効果が大きく，OW 以上に正解率が向上している．この理由は不明だが，OW は 2〜3 人，OC は 1 人，PN は 5〜7 人で作業 B を行ったために各コーパスで係り受けアノテーションの性質が異なり，並列構造数だけでは説明できないような傾向を示しているのかもしれない．

また，2 次の多項式カーネルを用いた実験の方が，3 次の多項式カーネルを用いるより高い正解率を示している．これは，アノテーションの一貫性がまだあまり高くないため，2 次のカーネルよりも詳細な文脈の学習をする 3 次のカーネルを用いると，かえって性能が落ちるということと考えられる．

表 3 に，制約を入れたことで正解 (gold) の係り先に係ることを禁止された係り受け関係の数を示す．その理由には，部分並列，接続表現およびアノテーション誤りが該当する．部

⋆4 その他のオプション: N=0 a=PA1 ave=1 s=1 iter=40 c=0.00005 M=0

Vol.2012-IFAT-105 No.6
Vol.2012-NL-205 No.6
2012/1/20

表 3 正解不可係り受け数
Table 3 Number of dependency relations which cannot modify their gold heads due to the constraints.

コーパス	制約なし	制約 (1)(2)	全制約 (最左構成句)	全制約 (最右構成句)
OW	(0)	12	23	342
OC	(0)	0	43	1
PN	(0)	3	42	21

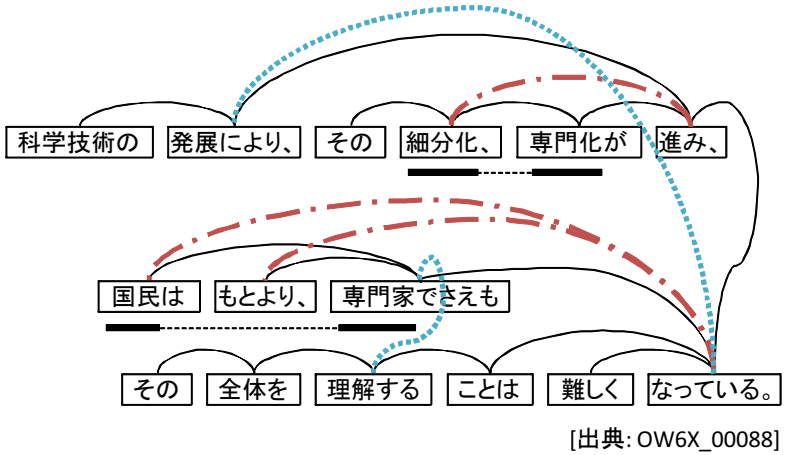


図 8 制約によって正解になった事例．黒実線:正解．赤鎖線:制約によって訂正された係り受け関係であり，訂正前の係り先を指している．青点線:訂正されなかった．
Fig. 8 An example sentence corrected by the constraints. Black solid arcs: gold-standard. Red dashed-dotted arcs: corrected. Blue dotted arcs: not corrected.

分並列は，前述のように，構成句内の文節が構成句外に係る構造となるため，制約 (1) および (2) によって禁止されるからである．これらのパターンを除けば，本手法を導入することによって，正解だった事例が誤ることはない．逆に言えば，表 3 に該当する事例を人手で見て部分並列でも接続表現でもなければアノテーション誤りということであるから，本手法によってアノテーション誤りの個所を効率的に特定することができると考えられる．表 3 からは，制約 (3) に関してアノテーションが統一されていないことがわかる．

4.2 具体的な事例

図 8 に，制約を入れることで正解となった事例を示す．制約なしでは解析を誤っていた 3

(前文脈:
今後、科学者等が社会的責任を果たす上で求められるのは、今までの公開講義のような一方的な情報発信ではなく、双方向的なコミュニケーションを実現するアウトリーチ（outreach）活動である。アウトリーチとは、リーチ・アウト（reach　out）という言葉が名詞化された言葉であり、もともとの意味は「手を伸ばす、差し伸べる」などである。)

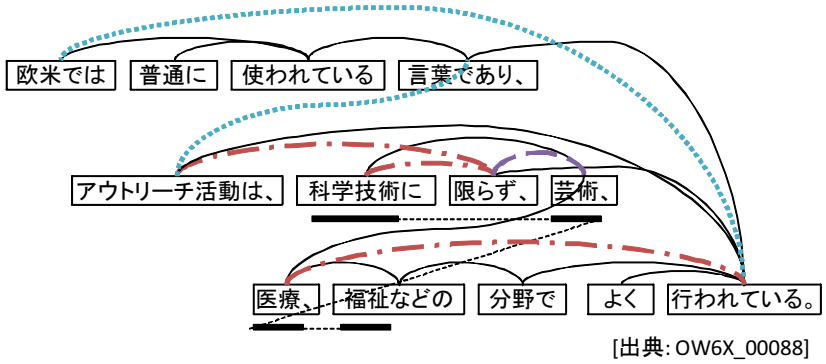


図 9 制約 (3) によって正解になった事例．紫破線: 制約によって誤りになった個所．
Fig. 9 Another example sentence corrected by the constraints. Purple dashed arcs: cannot correct due to the constraints.

個所の係り受け関係が，制約を入れることで正解できている．制約 (2) により，「細分化、」と「国民は」の係り先は一意に決まる．また，「もとより、」の係り先も，右隣の構成句「専門家でさえも」が一文節からなる構成句のため，制約 (1) により，係り先が一意に決まっている．

図 9 にもう一つ例を示す．この例では，並列構造の外から接続表現に係ることを禁止する制約 (3) によって「アウトリーチ活動は、」が「限らず、」に係ることを禁止しているおかげで正解している．一方，制約 (1) によって接続表現「限らず、」の係り先を誤っている．[⋆5]

図 10 に，制約を入れたことで不正解になった事例である，部分並列の事例を示す．この文は，すべての文節が「採択された。」に係るのが正しいが，いくつかの文節は制約 (1) によって正解の文節に係ることを禁止されている．

なお正解自体が，制約 (2) の自動修正処理の影響を受けている．図 9 の文では「科学技術

⋆5 これは特殊な並列構造の一例である．この部分は，英語でいう “not only ... but also ...” に相当すると考え，並列構造とアノテーションしているが，そもそもそうすることに異論があるかもしれない．

Vol.2012-IFAT-105 No.6
Vol.2012-NL-205 No.6
2012/1/20

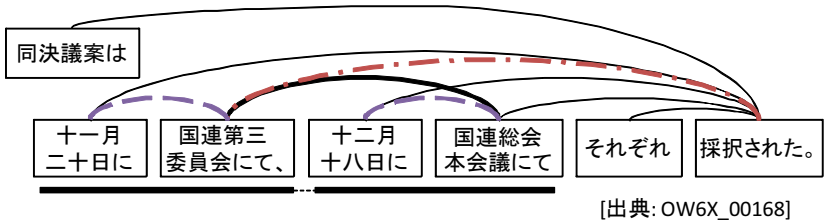


図 10 制約によって不正解になった事例 (部分並列構造)
Fig. 10 An example sentence which made mistakes due to the constraints: incomplete coordinated structure.

に」の正解係り先が「限らず、」から「芸術、」に変更されており，図 10 の文では「国連第三委員会にて,」の正解係り先が「採択された。」から「国連総会本会議にて」に変更されている．⋆6

自動修正処理は,「正解データ」自体の品質がまだあまり高くない現状で自動係り受け正解率を評価するために実験の便宜上行ったものであり，実際のアノテーション作業において行われるものではない．本手法を用いると係り受けアノテーション作業手順は，並列構造アノテーションを用いた制約あり自動係り受け解析，人手による修正作業（2.1 節の作業 B），制約を入れると正解不能になる事例（表 3 に相当）の人手による見直し，最終修正作業（作業 C）という流れになると考えられる．

5. おわりに

本稿では，係り受けアノテーションをする際に並列構造アノテーションが利用可能な場合に，並列構造が，可能な係り受け構造を制約することを利用して自動係り受け解析精度を向上させる手法を提案した．提案手法は，BCCWJ を利用した実験において係り受け正解を 2〜4%向上させた．本手法は，並列構造関連の係り受けアノテーション誤りを効率的に検出することにも有用であり，白書コーパスのような，並列構造が多く含まれるコーパスの構築において特に効果が大きいと考えられる．

今後は，部分並列等の例外的な並列構造の扱いを考える必要がある．もし品詞列などの手掛かりを用いることで例外的な並列構造とそれ以外を確実に区別することができれば，見直し作業の労力を削減できる可能性がある．たとえば，すべての構成句の長さが 2 文節以下の並列構造は，並列構造内のすべての文節（末尾の文節を除く）の係り先が制約 (1)(2) より一意に決まるため，人手で見直す必要がない．

⋆6 評価においては自動修正後の係り受け構造を「正解」とみなしているため，たとえば，制約なしでは「採択された。」に係っていたが，制約によって「国連総会本会議にて」に係るようになったことは，紫線でなく赤線で表している．

参 考 文 献

1) 黒橋禎夫, 居蔵由衣子, 坂口昌子：形態素・構文タグ付きコーパス作成の作業基準 Version 1.8, 京都大学 (2000).
2) 浅原正幸，岩立将和，松本裕治：BCCWJ コアデータへの係り受け・並列構造アノテーション，BCCWJ 公開記念講演会予稿集，pp. 71-76 (2011).
3) 伝康晴, 小木曽智信, 小椋秀樹, 山田篤, 峯松信明, 内元清貴, 小磯花絵：コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用,『日本語科学』22 号, pp.101-122 (2007).
4) Masakazu Iwatate, Masayuki Asahara, Yuji Matsumoto：Japanese Dependency Parsing Using a Tournament Model, In Proceedings of the 22nd International Conference on Computational Linguistics (COLING), pp.361-368 (2008).
5) Naoki Yoshinaga, Masaru Kitsuregawa：Polynomial to Linear: Efficient Classification with Conjunctive Features, In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1542-1551 (2009).
6) Naoki Yoshinaga：opal - C++ header library of online learning with kernel slicing 入手先⟨http://www.tkl.iis.u-tokyo.ac.jp/˜ynaga/opal/⟩