

『現代日本語書き言葉均衡コーパス』に対する 述語項構造・共参照関係アノテーション

植田 禎子 (日本システムアプリケーション)

飯田 龍 (情報通信研究機構)

浅原 正幸 (国立国語研究所) *

松本 裕治 (奈良先端科学技術大学院大学)

徳永 健伸 (東京工業大学)

Predicate-Argument Structure and Coreference Relation Annotation on ‘Balanced Corpus of Contemporary Written Japanese’

Yoshiko Ueda (Japan System Applications Co., Ltd.)

Ryu Iida (National Institute of Information and Communications Technology)

Masayuki Asahara (National Institute for Japanese Language and Linguistics)

Yuji Matsumoto (Nara Institute of Science and Technology)

Takenobu Tokunaga (Tokyo Institute of Technology)

要旨

述語項構造・共参照関係は、形態論情報・係り受け構造などの基本的な情報とともに重要であり、京都大学テキストコーパスや NAIST テキストコーパスなど、さまざまな基準に基づいたアノテーションが共有され、それに基づく解析器が整備されてきた。しかしながら、アノテーション先のテキストのほとんどが新聞記事であり、ジャンル横断的な述語項構造・共参照関係アノテーションは整備されていなかった。本発表ではアノテーション基準として NAIST テキストコーパス互換の基準を採用し、『現代日本語書き言葉均衡コーパス』コアデータ全体に対して行った、述語項構造・共参照関係アノテーションについて報告する。2015年7月末時点のアノテーションの統計量を示すとともに、ポスター発表においてアノテーション単位・基準や公開データ形式について紹介する。本アノテーションデータは2015年9月末に公開予定である。

1. はじめに

言語処理の分野において日本語の述語項構造や共参照関係の研究が盛んに進められ、アノテーション基準やそれに基づくアノテーションが整備されている。京都大学テキストコーパス (KTC) は、新聞記事に対して、益岡・田窪品詞体系 (益岡・田窪 (1992)) に基づく形態論情報や

* masayu-a@ninja.ac.jp

文節係り受けとともに、補助動詞や助動詞を含めた述部出現形に対する名詞句の格関係を、格助詞をラベルとして付与している(河原ほか(2002))。さらに名詞句間の属格の情報や、共参照関係・上位/下位関係・総称/非総称関係が付与されている。同様の情報が解析済みブログコーパス(Kyoto University and NTT Blog Corpus: KNBC)に付与されている(橋本ほか(2011))。NAIST テキストコーパス(NTC)は、KTCと同じ新聞記事に述語基本形に対する格関係を格助詞をラベルとして付与している(飯田ほか(2010))。京都大学ウェブ文書リードコーパス(萩行ほか(2014))は、ウェブ上にある多様な文書の書き始め先頭3文に対して、形態論情報・係り受け関係・固有表現のほか、述語項構造・共参照関係を付与している。特に新聞記事などにはあまり出現しない外界ゼロ照応の細分類について規定している。さらに、松林ほか(2014)はNTCの基準を中心に、述語項構造アノテーションの2014年現在の論点について整理して考察を行っている。

本稿では、2008年より進められてきた『現代日本語書き言葉均衡コーパス』(Maekawa et al. (2014)) コアデータ全体に対する述語項構造・共参照関係アノテーション(BCCWJ-PAS)(小町・飯田(2011))について報告する。BCCWJ-PASはアノテーション基準としてNTCの基準を継承している。松林ほか(2014)で示されている論点について全てを反映しているわけではないが、雑誌・書籍などを含む均衡コーパスに対するジャンル横断的なアノテーションとして、2015年9月に公開予定である。

以下、2節ではBCCWJ-PASのアノテーション基準の概略を示す。3節では2015年7月現在のアノテーションの統計を示す。4節では公開の形態について示し、5節にまとめと今後の課題について示す。

2. アノテーション基準

2.1 NAIST テキストコーパス基準

表1 ラベル一覧

	ラベル
述部	「述語」「事態」「助動詞」(「機能語相当」)
格要素	「ガ」「ヲ」「ニ」「ハ」「追加ガ/ニ」
ゼロ照応	「外界一人称」「外界二人称」「外界一般」「節照応」

BCCWJ-PASのアノテーションコーパスのアノテーション基準は基本的にNAISTテキストコーパスの基準に準拠する。しかしながら、新聞記事以外のジャンルのテキストを扱うほか、前提となっている形態論情報が異なるために細かい点で異なっている。

まず、アノテーション対象の述部について述べる。BCCWJに付与されている短単位形態素1単位の基本形を基本単位として、UniDic品詞体系が動詞・形容詞・名詞がアノテーション対象である。動詞・形容詞に対しては「述語」タグを付与する。名詞は述語をなす事態性名詞で「名詞句+助動詞(“だ”)」や「節末の名詞句」を対象とし、「事態」タグを付与する。さらに、後で述べるように、格が増えるような助動詞・補助動詞には、「助動詞」タグを付与する。なお、動詞・形容詞などが機能語・機能表現をなす場合には述語と認定せず、別途「機能語相当」と

いうタグを付与する。

格関係は、述語の基本形に対する「ガ格」、「ヲ格」、「ニ格」の格助詞をラベルとして付与する。「好き」などの形状詞や「できる」などの可能動詞の場合、対象となっている要素を表す「ガ格」とは別に、提題されている要素に対して「ハ格」を付与する。

PM41_00219

私_ハの好き_{述語}な街_ガ 1

格交代によって、主動詞のガ格、ヲ格、ニ格以外に格が増える場合には、対象となる助動詞や補助動詞に「助動詞」タグを付与し、追加で導入される格に「追加ガ/ニ」を付与する。

OY01_00137

自分_{追加ガ/ニ}が40歳になった頃から、税理士会・TKC活動に参加、以後幾多の会議に出席させて頂く機会_ヲに恵ま_{述語}れ_{助動詞}ました。

^a 「恵ま」のガ格は「外界一般」。

共参照関係は文節の主辞をなす名詞句を対象とする。談話内に出現した名詞句（言及）が指示的で、現実世界の实体に写像可能な表現である際に、直接照応可能な先行詞が存在する場合に付与する。総称名詞は照応詞・先行詞として考えない。

PN3b_00001 (共参照をインデックス番号で表示)

★プレゼントはビートルズ₁

米CNNテレビのプロデューサー、ウエンディ・ウィットワースさん₂の50歳の誕生日、夫のラルフさん₃が“プレゼント”したのは何と、元ビートルズ₁のポール・マッカートニー₄のプライベートコンサート₅だった。

誕生パーティーで夫₃がポール₄の登場を告げると、彼₄の大ファンのウエンディさん₂は思わず涙。約1時間半のコンサート₅の終幕近くには、ポール₄はウエンディさん₂をステージに上げ、ビートルズ₁の「バースデイ」を演奏したという。

ポール₄は、報酬の100万ドルを反地雷の慈善団体に寄付するとしている。

ゼロ代名詞の参照の实体が談話における一人称（著者）や二人称（読者）を指す場合、それぞれ「外界一人称」「外界二人称」のタグを付与する。一人称（著者）や二人称（読者）はともに単数であることを前提とする。Yahoo! 知恵袋の場合、質問者-回答者の関係についても「外界二人称」とみなす。その他の談話中に出現する外界照応については「外界一般」のタグを付与する。ゼロ代名詞が節・文・文章などを指す場合、節末の文節の主辞に「節照応」としてタグ付けする。

OY08_00189 (外界のみ表示)

今日は「後ろのヤツ」とはしゃべら_{ガ: 外界一人称}んかった。

まあ授業に集中し_{ガ: 外界一人称}てましたからね。

っていうよりは英語とプログラミングが移動なんでね。

プロは考え_{ガ: 外界一人称}てる暇すら与え_{ニ: 外界一人称}てくれないよ。

でも今日は指の調子が良かったので、早く打ち終わる_{ガ: 外界一人称, ニ: 外界一般} ことが出来_{ガ: 外界一人称}ましてん。

今日は部活があったけど1時半で帰れた_{ガ: 外界一人称, ニ: 外界一般} のよ〜ん。

そいじゃ、まったあしたあ。

表1にラベル一覧を示す。詳細については <https://sites.google.com/site/ryuuida/ntc-annotation-scheme> を参照すること。

2.2 実アノテーション作業

BCCWJ-PAS のアノテーション作業は図 Tagrin (高橋・乾 (2006))⁽¹⁾ を用いて行った。

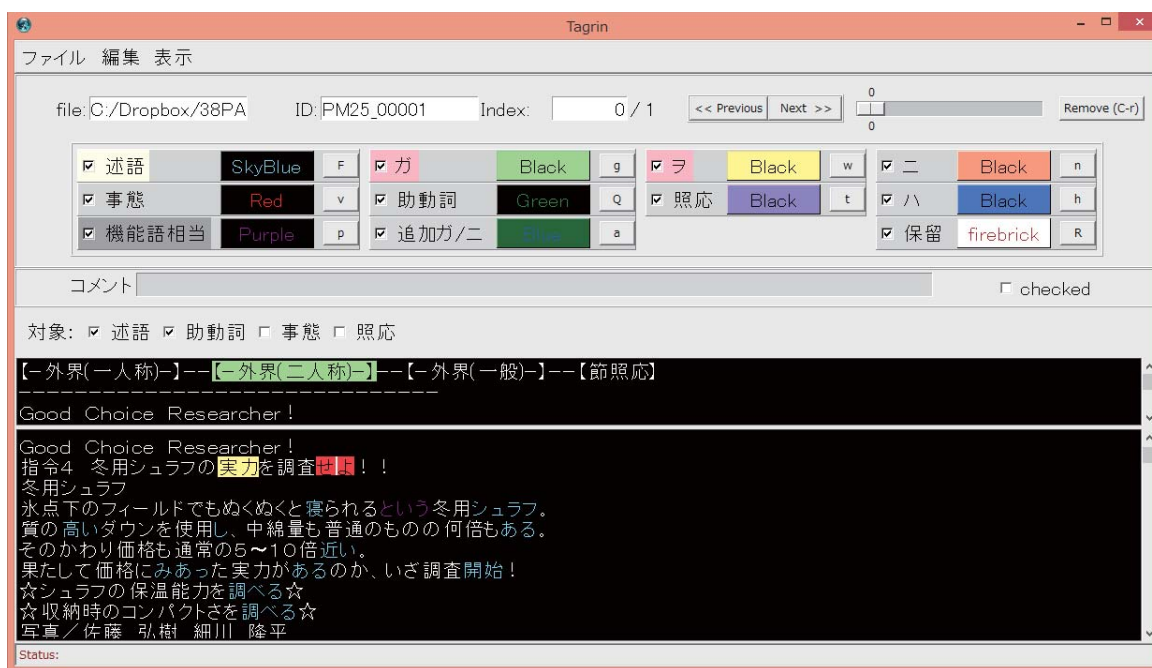


図1 アノテーションツール Tagrin

作業においては以下の点に気をつけながら作業を進めた。

アノテーション誤りよりもアノテーション漏れが多くなるタイプのタスクであり、アノテーション漏れがないように作業者に教示した。

タグづけの範囲の例外として、複合動詞の扱いがある。複合動詞の格は基本的に前の動詞につけるが、後ろの動詞によって格が変化する場合には複合動詞全体を述語とみなし述語タグを付与する。

格要素が直接の係り受け関係にない場合は、作業者間の揺れが大きくなる傾向にある。この場合、格要素のうち直近のものを選択するように教示した。

3. 統計

本節では、2015年7月時点でのアノテーションの統計を示す。残念ながら現時点で、BCCWJ DVD 1.1版や BCCWJ-DepPara との重ね合わせが実現しておらず、形態論情報や係り受け構造との組み合わせによる集計ができていないため、限定的な情報になる。

⁽¹⁾ <http://kagonma.org/tagrin/>

3.1 述語項構造アノテーションの統計

表2に述部と格要素の頻度を示す。

表2 述語と格要素の頻度

ジャンル	述部	述語数	ガ	ヲ	ニ	ハ	追加ガ/ニ
PB	述語	27027	26904	9652	5264	796	9
		100.00 %	99.54 %	35.71 %	19.48 %	2.95 %	0.03 %
PB	事態	5813	5802	2722	656	5	3
		100.00 %	99.81 %	46.83 %	11.29 %	0.09 %	0.05 %
PB	助動詞	478	8	1	3	0	468
		100.00 %	1.67 %	0.21 %	0.63 %	0.00 %	97.91 %
PN	述語	32434	32423	13035	5498	200	1
		100.00 %	99.97 %	40.19 %	16.95 %	0.62 %	0.00 %
PN	事態	14949	14947	6084	1320	1	1
		100.00 %	99.99 %	40.70 %	8.83 %	0.01 %	0.01 %
PN	助動詞	426	2	0	1	0	423
		100.00 %	0.47 %	0.00 %	0.23 %	0.00 %	99.30 %
PM	述語	25275	25272	9206	4140	56	0
		100.00 %	99.99 %	36.42 %	16.38 %	0.22 %	0.00 %
PM	事態	5039	5039	2131	410	1	0
		100.00 %	100.00 %	42.29 %	8.14 %	0.02 %	0.00 %
PM	助動詞	347	0	0	4	0	343
		100.00 %	0.00 %	0.00 %	1.15 %	0.00 %	98.85 %
OW	述語	15436	15384	6941	2388	471	0
		100.00 %	99.66 %	44.97 %	15.47 %	3.05 %	0.00 %
OW	事態	15439	15432	9048	1314	1	1
		100.00 %	99.95 %	58.60 %	8.51 %	0.01 %	0.01 %
OW	助動詞	158	0	0	0	0	158
		100.00 %	0.00 %	0.00 %	0.00 %	0.00 %	100.00 %
OC	述語	13466	13343	4293	2501	606	1
		100.00 %	99.09 %	31.88 %	18.57 %	4.50 %	0.01 %
OC	事態	3472	3472	1756	664	4	0
		100.00 %	100.00 %	50.58 %	19.12 %	0.12 %	0.00 %
OC	助動詞	269	2	0	0	0	268
		100.00 %	0.74 %	0.00 %	0.00 %	0.00 %	99.63 %
OY	述語	12706	12703	3740	1761	138	0
		100.00 %	99.98 %	29.43 %	13.86 %	1.09 %	0.00 %
OY	事態	2997	2997	1043	222	0	0
		100.00 %	100.00 %	34.80 %	7.41 %	0.00 %	0.00 %
OY	助動詞	146	0	0	0	0	146
		100.00 %	0.00 %	0.00 %	0.00 %	0.00 %	100.00 %

ほとんどの述語・事態に対してガ格が規定されていることがわかる。述語については、PN（新聞）・OW（白書）についてはヲ格が多く、OC（Yahoo! 知恵袋）・OY（Yahoo! ブログ）はヲ格が少ない。事態は述語に比してヲ格が多い傾向にある。また、OWは事態が多い傾向にある。ほとんどの助動詞に対して追加ガ/ニが規定されている。

表3に述語のジャンルごとの項の組み合わせの分布について示す。3項（ガヲニ）はOCが若干多い傾向が見られた。2項（ガヲ）は、PN・OWが多く、OC・OYが少ない。1項（ガの

表3 項の組み合わせの分布 (述語)

ジャンル	ガヲニ	ガヲ	ガニ	ヲニ	ガ	ヲ	ニ
PB	1516 5.61 %	8075 29.88 %	3643 13.48 %	6 0.02 %	12904 47.74 %	39 0.14 %	47 0.17 %
PN	1641 5.06 %	11373 35.07 %	3840 11.84 %	3 0.01 %	15369 47.39 %	6 0.02 %	2 0.01 %
PM	1324 5.24 %	7878 31.17 %	2812 11.13 %	1 0.00 %	13202 52.23 %	2 0.01 %	0.00 %
OW	649 4.20 %	6271 40.63 %	1697 10.99 %	0.00 %	6310 40.88 %	17 0.11 %	21 0.14 %
OC	890 6.61 %	3382 25.12 %	1553 11.53 %	0.00 %	6986 51.88 %	19 0.14 %	29 0.22 %
OY	404 3.16 %	3328 26.01 %	1351 10.56 %	1 0.01 %	7482 58.47 %	1 0.01 %	1 0.01 %
ジャンル	ガヲニハ	ガヲハ	ガニハ	ガハ	ニハ	ハ	*
PB	8 0.03 %	2 0.01 %	35 0.13 %	713 2.64 %	1 0.00 %	29 0.11 %	9 0.03 %
PN	4 0.01 %	7 0.02 %	7 0.02 %	181 0.56 %	0.00 %	0.00 %	1 0.00 %
PM	0.00 %	1 0.00 %	3 0.01 %	52 0.21 %	0.00 %	0.00 %	0.00 %
OW	1 0.01 %	3 0.02 %	19 0.12 %	434 2.81 %	1 0.01 %	13 0.08 %	0.00 %
OC	0.00 %	2 0.01 %	27 0.20 %	502 3.73 %	2 0.01 %	73 0.54 %	1 0.01 %
OY	0.00 %	6 0.05 %	4 0.03 %	128 1.00 %	0.00 %	0.00 %	0.00 %

* は「追加ガ/ニを含むその他の組合せ」。

み)は OY が多い。また二重主語である 2 項 (ガハ) は PB・OW・OC が多く、PN・PM が少ない。

表4 項の組み合わせの分布 (事態)

ジャンル	ガヲニ	ガヲ	ガニ	ガ	ヲ	ニ	ガハ	*
PB	234 4.03 %	2479 42.65 %	414 7.12 %	2670 45.93 %	6 0.10 %	5 0.09 %	2 0.03 %	3 0.05 %
PN	203 1.36 %	5879 39.33 %	1117 7.47 %	7746 51.82 %	2 0.01 %		1 0.01 %	1 0.01 %
PM	90 1.79 %	2041 40.50 %	320 6.35 %	2587 51.34 %			1 0.02 %	
OW	511 3.31 %	8531 55.26 %	802 5.19 %	5586 36.18 %	6 0.04 %	1 0.01 %	1 0.01 %	1 0.01 %
OC	243 7.00 %	1512 43.55 %	420 12.10 %	1293 37.24 %			3 0.09 %	
OY	49 1.63 %	994 33.17 %	173 5.77 %	1781 59.43 %				

* は「追加ガ/ニを含むその他の組合せ」。

表4に事態のジャンルごとの項の組み合わせの分布について示す。3項(ガヲニ)はOC・PBが多い傾向にある。OCは他に2項(ガニ)が多く、1項(ガ)が少ない。OWは2項(ガ

ヲ)が多く、1項(ガ)が少ない。一方、PN・PM・OYは2項(ガヲ)が多く、1項(ガ)が多い。

表5 項の組み合わせの分布(助動詞)

ジャンル	ガヲニ	ガ	ニ	追加ガ/ニ	*
PB	1 0.21 %	7 1.46 %	2 0.42 %	468 97.91 %	3 0.63 %
PN		2	1 0.23 %	423 99.30 %	
PM			4 1.15 %	343 98.85 %	
OW				158 100.00 %	
OC		1 0.37 %		267 99.26 %	1 0.37 %
OY				146 100.00 %	

*は「追加ガ/ニを含むその他の組合せ」。

表5に助動詞のジャンルごとの項の組み合わせの分布について示す。ほとんどが「追加ガ/ニ」である。

3.2 外界ゼロ照応・節照応の分布

表6にガ格の外界ゼロ照応・節照応の分布について示す。

まず、紙媒体(PB・PN・PM・OW)には「外界一人称」・「外界二人称」があまり出現しない。紙媒体の中では、OWには「外界一人称」・「外界二人称」がともにほとんど出現しない一方、PMは読者を意識した「外界二人称」が出現することがわかった。Web媒体(OC・OY)は「外界一人称」・「外界二人称」が多く出現することがわかった。質問者-回答者間のインタラクションが発生するOCは「外界二人称」が多く、個人の経験を記述するOYは特に「外界一人称」が多かった。

OWには「外界一般」が多く出現した。これは、OWにおいて、組織・集団を意識した記述が多く、個人を意識した記述が少ないためだと考えられる。報道を目的とするPNは外界ゼロ照応が少ないことがわかった。基本的に文書内に閉じた空間になっていることがわかる。

3.3 共参照関係

本節では共参照関係の統計について示す。

同一の実体を参照する言及間の共参照関係を言及の数⁽²⁾と実体の数⁽³⁾で集計する。

表7に集計結果を示す。「言及/実体」は実体の言及数の平均値である。

サンプル中の形態素数が大きいPB・OWは1つの実体の言及数が多くなる傾向になる一方、サンプル中の形態素数が小さいOCは1つの実体の言及数が少なくなる。PN・PMは1つのサンプルに複数の記事が含まれていることから、実体の言及数が中程度になっていると考える。

(2) 共参照関係を同値関係とみなした場合の同値類の要素ののべ。

(3) 共参照関係を同値関係とみなした場合の同値類の異なり。

表6 外界ゼロの分布 (ガ格)

述語数			ガ	外界一人称	外界二人称	外界一般	節照応
PB	述語	27270	26904	212	88	3384	4
			100.0 %	0.8 %	0.3 %	12.6 %	0.0 %
PB	事態	5835	5802	16	17	1955	2
			100.0 %	0.3 %	0.3 %	33.7 %	0.0 %
PB	助動詞	1865	8	0	0	1	0
PN	述語	32712	32423	298	178	4013	4
			100.0 %	0.9 %	0.5 %	12.4 %	0.0 %
PN	事態	15204	14947	31	95	3921	15
			100.0 %	0.2 %	0.6 %	26.2 %	0.1 %
PN	助動詞	426	2	1	0	0	0
PM	述語	25421	25272	452	289	3686	17
			100.0 %	1.8 %	1.1 %	14.6 %	0.1 %
PM	事態	5141	5039	38	118	1410	6
			100.0 %	0.8 %	2.3 %	28.0 %	0.1 %
PM	助動詞	347	0	0	0	0	0
OW	述語	15654	15384	75	9	4438	0
			100.0 %	0.5 %	0.1 %	28.8 %	0.0 %
OW	事態	15475	15432	12	0	7383	2
			100.0 %	0.1 %	0.0 %	47.8 %	0.0 %
OW	助動詞	260	0	0	0	0	0
OC	述語	13754	13343	1724	1137	1274	3
			100.0 %	12.9 %	8.5 %	9.5 %	0.0 %
OC	事態	3482	3472	406	289	1070	0
			100.0 %	11.7 %	8.3 %	30.8 %	0.0 %
OC	助動詞	1131	2	1	0	0	0
OY	述語	12807	12703	2528	314	1462	5
			100.0 %	19.9 %	2.5 %	11.5 %	0.0 %
OY	事態	3035	2997	534	97	932	0
			100.0 %	17.8 %	3.2 %	31.1 %	0.0 %
OY	助動詞	146	0	0	0	0	0

表7 共参照関係の集計 (言及と実体)

ジャンル	言及	実体	言及/実体
PB	9459	2237	4.33
PN	16476	4445	3.76
PM	9866	2658	3.82
OW	7169	1661	4.47
OC	2468	994	2.49
OY	3506	1106	3.23

表8に実体の言及数⁽⁴⁾の分布を示す。OCにおいて、実体の言及数2が全体の72%を占めている。一方、どのサンプルでも11回以上言及される実体が数%あることが確認された。

⁽⁴⁾ 実体がある同値類を構成する要素数。

表 8 共参照関係の集計 (実体の言及数)

ジャンル	実体の言及数分布							
	2	3	4	5	6..10	11..15	16..20	21..
PB	1278 57.13 %	270 12.07 %	135 6.03 %	92 4.11 %	170 7.60 %	55 2.46 %	50 2.24 %	73 3.26 %
PN	2232 50.21 %	579 13.03 %	617 13.88 %	257 5.78 %	337 7.58 %	70 1.57 %	36 0.81 %	52 1.17 %
PM	1362 51.24 %	426 16.03 %	251 9.44 %	153 5.76 %	173 6.51 %	57 2.14 %	36 1.35 %	46 1.73 %
OW	822 49.49 %	228 13.73 %	163 9.81 %	81 4.88 %	124 7.47 %	66 3.97 %	35 2.11 %	54 3.25 %
OC	717 72.13 %	166 16.70 %	51 5.13 %	30 3.02 %	16 1.61 %	11 1.11 %	16 1.61 %	21 2.11 %
OY	638 57.69 %	139 12.57 %	133 12.03 %	69 6.24 %	53 4.79 %	23 2.08 %	16 1.45 %	25 2.26 %

3.4 機能語相当の統計

表 9 に「機能語相当」タグの統計を示す。機能語相当の表現は OW・PB が多く、PM・OY が少ないことがわかる。

表 9 機能語相当表現の統計

ジャンル	PB	PN	PM	OW	OC	OY
機能語相当	3031	1889	933	5179	1000	428

4. 公開形態

本節では BCCWJ-PAS の公開形態について示す。

他データとの重ね合わせとして、まず BCCWJ-DVD 版 (Version 1.1)(国立国語研究所 (2015)) の形態論情報との統合を行う。さらに係り受け・並列構造アノテーションである BCCWJ-DepPara (浅原・松本 (2013)) との統合を行い、同データに付与されている文境界情報 (小西ほか (2013)) を統合する。さらに、受身に関する格交代の情報を統合するために、れる・られるの用法 (小山田ほか (2012)) との統合を行う。

公開するファイル形式は、2015 年 9 月の時点では NTC に相当するファイル形式での配布を検討する。形態論情報、係り受け、並列構造、れる・られるの用法は、拡張 CaboCha フォーマット (松吉ほか (2014)) に準拠するが、述語項構造、共参照関係は NTC 互換の形式にとどめる。

5. おわりに

本稿では 2015 年 9 月に公開を予定している『現代日本語書き言葉均衡コーパス』に対する述語項構造・共参照関係アノテーション BCCWJ-PAS について紹介した。

本稿執筆時点 (2015 年 7 月) では、形態論情報や係り受けなどのアノテーションと統合がさ

れていないために、アノテーションデータの分析に制限がある。今後の課題として、4節に示した公開形態への整形作業とともに、述語の品詞ごとの分析や係り受け・文節境界・文境界との位置関係による分析がある。また、並行して進めている節境界との統合も行いたい。

謝辞

8年間にわたる述語項構造アノテーションに携わったすべてのみなさまに感謝の意を表します。本研究の一部は科研費特定研究「書き言葉コーパスの自動アノテーションの研究」(18061005)、科研費基盤(B)「言語コーパスに対する読文時間付与とその利用」(25284083)、科研費萌芽「近代語コーパスに対する統語情報アノテーション基準策定」(15K12888)、国語研基幹型共同研究プロジェクト「コーパスアノテーションの基礎研究」および国語研「超大規模コーパス構築プロジェクト」によるものです。

参考文献

- 浅原正幸・松本裕治(2013).『現代日本語書き言葉均衡コーパス』に対する係り受け・並列構造アノテーション」言語処理学会第19回年次大会発表論文集.
- 萩行正嗣・河原大輔・黒橋禎夫(2014).「多様な文書の書き始めに対する意味関係タグ付きコーパスの構築とその分析」自然言語処理, 21:2, pp. 213-247.
- 橋本力・黒橋禎夫・河原大輔・新里圭司・永田昌明(2011).「構文・照応・評価情報つきブログコーパスの構築」自然言語処理, 18:2, pp. 175-201.
- 飯田龍・小町守・井之上直也・乾健太郎・松本裕治(2010).「述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から」自然言語処理, 17:2, pp. 25-50.
- 河原大輔・黒橋禎夫・橋田浩一(2002).「関係」タグ付きコーパスの作成」言語処理学会第8回年次大会発表論文集, pp. 495-498.
- 国立国語研究所コーパス開発センター(2015).『現代日本語書き言葉均衡コーパス』利用の手引き 第1.1版』, http://pj.ninjal.ac.jp/corpus_center/bccwj/doc/manual/BCCWJ_Manual_01.pdf.
- 小町守・飯田龍(2011).「BCCWJに対する述語項構造と照応関係のアノテーション」『現代日本語書き言葉均衡コーパス』完成記念講演会.
- 小西光・小山田由紀・浅原正幸・柏野和佳子・前川喜久雄(2013).「BCCWJ係り受け関係アノテーション付与のための文境界再認定」第4回コーパス日本語学ワークショップ予稿集, pp. 135-142.
- Maekawa, Kikuo, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den (2014). “Balanced Corpus of Contemporary Written Japanese.” *Language Resources and Evaluation*, 48, pp. 345-371.
- 益岡隆志・田窪行則(1992).『基礎日本語文法-改訂版-』くろしお出版.
- 松林優一郎・飯田龍・笹野遼平・横野光・松吉俊・藤田篤・宮尾祐介・乾健太郎(2014).「日本語文章に対する述語項構造アノテーション仕様の考察」自然言語処理, 21:2, pp. 333-377.
- 松吉俊・浅原正幸・飯田龍・森田敏生(2014).「拡張 CaboCha フォーマットの仕様拡張」第5回コーパス日本語学ワークショップ, pp. 223-232.
- 小山田由紀・柏野和佳子・前川喜久雄(2012).「助動詞レル・ラレルへの意味アノテーション作業経過報告」第2回コーパス日本語学ワークショップ予稿集.
- 高橋哲郎・乾健太郎(2006).「アノテーションツール”Tagrin”の紹介」言語処理学会第12回年次大会発表論文集, pp. 228-231.