

BCCWJ-TimeBank に対する時間幅情報付与

浅原 正幸* 森山 奈々美 荻原 亜彩美
 国立国語研究所 国立国語研究所 国立国語研究所

1. はじめに

BCCWJ-TimeBank は、時間表現と事象表現間もしくは 2 事象表現間の時間的順序関係を付与されたデータである。時間的順序関係は、時間幅を考慮した Allen の範囲代数に基づく 13 種類の関係が付与されている。この 13 種類の関係のうち 9 種類は、2 つの時間幅に重なりがあり、残り 4 種類は重ならない。また、5 種類の関係については、時間幅のどちらかの端が接している。時間幅の重なりや接点の有無は、その表現が指し示す時間幅の長短に大きく影響を受ける。この時間幅の分布を得るために、本研究では、BCCWJ-TimeBank (Asahara et al. 2014) に含まれる事象表現 に対してヒトがどのように時間幅の長短を認識するかについて、クラウドソーシングを用いた質問紙調査を実施したので報告する。

2. 調査方法

調査は Yahoo! クラウドソーシングを用いた質問紙調査による。1 文を呈示し、対象となる表現を <<< >>> により示したうえで、各表現の時間幅を「瞬時～1 秒未満」「1 秒以上 1 分未満」「1 分以上 1 時間未満」「1 時間以上 1 日未満」「1 日以上 1 年未満」「1 年以上（常に成り立つを除く）」「常に成り立つ」のいずれに当てはまるかについて選択させた。図 1 にクラウドソーシング画面の例を示す。

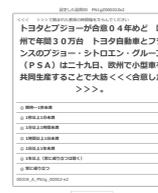


図 1 クラウドソーシング画面

対象は BCCWJ-TimeBank に含まれる事象表現 (<EVENT> タグ) 4952 例と、時間表現 (<TIMEEX3> タグ) 1784 例チェック設問として、時間幅が明瞭な時間表現 (1784 例中 1306 例) に対して正しく割り当てられているか判断する課題を設定した。実験協力者 2101 人から各表現に対し 10 件の回答を得た。調査は 2019 年 11 月 21 日 8:00 に開始し、12 月 5 日 13:04 に終了した。

3. 得られたデータの分析

BCCWJ-TimeBank (保田ほか 2013) の事象表現の class 属性との対照分析を行う。EVENT タグには、事象表現が事象表現を項にとるかなどの情報が class 属性に付与されている。

- OCCURRENCE: 項に事象をとらない静態動詞ではない一般的な事象表現。意志的か否かは問題とせず、無意志的（状態・位置）変化動詞・非意志的（現象一般）動詞・過程を表す動詞なども OCCURRENCE とみなす。
- REPORTING: 項に事象をとる表現活動動詞に相当する事象表現。
- PERCEPTION: 項に事象をとる認識・知覚動詞に相当する事象表現。
- ASPECTUAL: 項に事象をとるアスペクトを表出する事象表現。

* masayu-a@ninja.ac.jp

- I.ACTION (intension): 項に事象をとる遂行動詞に相当する事象表現。
- I.STATE (intension): 項に事象をとる思考・感情動詞に相当する事象表現。
- STATE: 静態動詞、形容詞。
- NULL, NONE: 時間軸上インスタンスが認められない事象表現。

本調査では前後の文脈なし 1 文のみ見せているために、実験協力者間で揺れがある事例がある。また本質的に時間幅についてあいまい性がある事例も存在する。そこで、以下の分析においては、調査した 4952 例の事象表現のうち、クラウドソーシングによるラベルごとの最大回答数が半数以上の 2101 例を対象とした。

表 1 時間幅と BCCWJ-TimeBank の class 属性

		瞬時～ 1 秒未満	1 秒以上 1 分未満	1 分以上 1 時間未満	1 時間以上 1 日未満	1 日以上 1 年未満	1 年以上	常に 成り立つ	総計
OCCURRENCE	項に事象を取らない事象表現一般	22	33	79	86	194	179	360	953
REPORTING	項に事象をとる表現活動動詞	1	5	3	1	1	2	24	37
PERCEPTION	項に事象をとる認識・知覚動詞	3	2	1		1		3	10
ASPECTUAL	項に事象をとるアスペクトを表出する表現			2	3	10	9	8	32
I.ACTION	項に事象をとる遂行動詞	9	5	15	41	90	58	117	335
I.STATE	項に事象をとる思考・感情動詞	4	2			19	16	44	85
STATE	静態動詞、形容詞など		2		3	14	10	63	92
NONE/NULL	時間軸上インスタンスが認められない事象表現	9	11	8	6	41	78	404	557
総計		48	60	108	140	370	352	1023	2101

時間幅と class 属性との対照比較を表 1 に示す。一般的な事象表現 (OCCURRENCE) は (1)–(7) のように全ての時間幅を表出する。

- (1) 左足つま先が氷を<<<削った>>>。 【PN2f.00003】 OCCURRENCE, 瞬時～1 秒未満
- (2) ジャガイモ 1 個、ニンジン小 1 本を<<<洗い>>>、大きければ皮ごと半分。 【PN4a.00001】 OCCURRENCE, 1 秒以上 1 分未満
- (3) 米 1 合をといで普通に水加減をし、だし昆布 5 センチと塩少々を加え、野菜と肉をのせて<<<炊く>>>。 【PN4a.00001】 OCCURRENCE, 1 分以上 1 時間未満
- (4) 相手を<<<圧倒する>>>力強さはないが、「向かっていった」と小田桐が振り返るように、自分たちから先に仕掛ける積極的な試合展開で、ラリーでは主導権を握った。 【PN1e.00003】 OCCURRENCE, 1 時間以上 1 日未満
- (5) 「いまは 8 割の人が残りご飯を<<<冷蔵・冷凍してしまい>>>、炊飯器の出番は少ない」 【PN4a.00001】 OCCURRENCE, 1 日以上 1 年未満
- (6) 長い目で制度を<<<育て上げる>>>ことがぜひとも必要。 【PN3g.00001】 OCCURRENCE, 1 年以上 (常に成り立つは除く)
- (7) ご飯物や肉じゃがといった定番料理、おやつなど 108 品を<<<収めた>>>本は、 【PN4a.00001】 OCCURRENCE, 常に成り立つ

表現活動表現 (REPORTING) は結果 (残存) に意識がおかれ「常に成り立つ」(8) が選択されやすい傾向にある。

(8) 「計画全般に<<<言える>>>のは、部分的に見ると削減効果はあるものの、その裏に相殺する要素が出る施策が多いこと」 【PN5a_00003】 REPORTING, 常に成り立つ

認識・知覚表現 (PERCEPTION) は、瞬間的な知覚 (9)、継続した知覚 (10)、一定期間反復した知覚 (11) などが確認された。

(9) 不気味な枝ぶりを誇る樹木が、たけり狂う竜か化け物に<<<見えた>>>からだ。 【PN2b_00002】 PERCEPTION, 瞬時~1秒未満

(10) エンターテインメントを<<<熟知した>>>作・演出の上で、俳優たちが的確に配置されているからだ。 【PN5c_00002】 PERCEPTION, 常に成り立つ

(11) 銀河系の星がたくさん集まる「天の川」が、鮮明に<<<見える>>>季節でもある。 【PN2g_00004】 PERCEPTION, 1日以上1年未満

アスペクトを表出する表現 (ASPECTUAL) は「続く」という表現が多く、その継続時間がそのまま時間幅に反映される傾向がみられた。

(12) 試合時間に合わせるため、急きょ会場を変更しての練習はリラックスした雰囲気<<<続いた>>>。 【PN2f_00002】 ASPECTUAL, 1時間以上1日未満

(13) サッカーのワールドカップ (W杯) へ調整を<<<続ける>>>日本代表は2日、W杯会場となる神戸市の神戸ウイングスタジアムで 【PN2f_00002】 ASPECTUAL, 1日以上1年未満

(14) 「係数」は前年の病院収入を<<<維持できる>>>ような補正の意味合いも持ち、「当面の激変緩和措置」(厚労省)となっている。 【PN3a_00003】 ASPECTUAL, 1年以上(常に成り立つは除く)

(15) 青い空間に点滅する数字、それは永遠に<<<続く>>>変化。 【PN2d_00004】 ASPECTUAL, 常に成り立つ

項に事象をとる遂行動詞表現 (I_ACTION) や項に事象をとる静態動詞表現 (I_STATE) は、項に割り当てられる事象表現の長さにより時間幅が決まる傾向にある。

(16) 素早く木田にパスを<<<回し>>>試合を決めた。 【PN1e_00003】 I_ACTION, 1秒以上1分未満

(17) 小夜子が電車の中で化粧を<<<し>>>たり着替えをし、自分が注意をしたら 【PN1b_00003】 I_ACTION, 1分以上1時間未満

(18) 「K O B E ビース i ネット」のメンバー9人が、ギターを生演奏などで平和を<<<訴えた>>>。 【PN3a_00002】 I_ACTION, 1時間以上1日未満

(19) 柊乃洋をまったく寄せ付けなかった。「止まらないで攻めようと<<<思った>>>。 【PN1e_00004】 I_STATE, 瞬時~1秒未満

(20) 教師は、個々の奉仕活動を点数できめ細かく評価することは<<<考えるべきではない>>>。 【PN2g_00004】 I_STATE, 常に成り立つ

静態動詞、状態を表す表現 (STATE) は、(21) のように時間幅が長い表現が多いと考えていたが、(22) のように一時的な状態を表す表現も確認された。

- (21) 多くの人(とりわけ母親たち)には、名著「育児の百科」を書いた<<<頼れる>>>小児科の先生として親しみがあっただろう

【PN2b.00002】 STATE, 常に成り立つ

- (22) 男子ハンマー投げで自己ベストを16センチ上回る58メートル48を記録し、5位に入った野口=写真=だが「全然ダメ。前半に58メートル台が出て<<<力みすぎた>>>」

【PN1e.00003】 STATE, 1秒以上1分未満

最後に NONE, NULL タグは、時間軸上に特定の事象が認められない事象表現である。実験協力者は、(23)のような特定の事象ではない場合、不特定の事象として一般にどの程度時間を要するかを判断して回答する傾向がみられた。また、(24)のような事象表現ではない機能表現の場合には、隣接する表現の時間幅を答える傾向がみられた。

- (23) 家庭で<<<食べる>>>サラダの種類や味付けが、マヨネーズ一辺倒ではなく多様化してきたことがうかがわれた。

【PN1a.00002】 NULL, 1分以上1時間未満

- (24) この屋上は01年春に突然、オーストラリアや東南アジアからやって<<<くる>>>カモメ科のコアジサシの子育ての場

【PN5b.00002】 NULL, 1日以上1年未満

4. おわりに

本研究では、BCCWJ-TimeBankの事象表現に対して時間幅の情報を付与し、BCCWJ-TimeBankのEVENTタグのclass属性と対照比較を行った。項に事象をとらない事象表現(OCCURRENCE)は事象自体の時間幅を直接表現され、さまざまな時間幅が確認された。項に事象をとるもののうち、表現活動表現(REPORTING)は結果に意識が置かれ、「常に成り立つ」が選択されやすい。認識・知覚表現(PERCEPTION)は、瞬間的な知覚・継続した知覚・反復した知覚に応じて、時間幅が認識された。アスペクトを表出する表現(ASPECTUAL)は、新聞記事には「続く」の表現が多く、その継続時間がそのまま時間幅に反映された。項に事象をとる遂行動詞表現(I_ACTION)や静態動詞表現(I_STATE)は、項に割り当てられる事象表現の時間幅が認識される傾向がみられた。最後に状態を表す表現(STATE)は一般に時間幅が長い表現が多いと考えていたが、文脈に基づき一時的な状態を表す表現も確認された。

今後、BCCWJ-TimeBankの時間的順序関係と時間幅の関係を調査する。簡単な調査では、文書作成日時と事象表現の関係と時間幅とを対照した際に、時間幅が短いほどcontainsが多く、時間幅が長いほどduringが多いという自然な結果が確認された。さらに2事象表現間関係の関係も含めて検討を進める。

謝辞

本研究は国立国語研究所コーパス開発センター共同研究プロジェクトおよび科研費JP17H00917, JP18H05521によるものです。

文献

- Masayuki Asahara, Kikuo Maekawa, Mizuho Imada, Sachi Kato, and Hikari Konishi (2014). “BCCWJ-TimeBank: Temporal and Event Information Annotation on Japanese Text.” *Computational Linguistics and Chinese Language Processing*, 19:3, pp. 1–24.
- 保田祥・小西光・浅原正幸・今田水穂・前川喜久雄 (2013). 「『現代日本語書き言葉均衡コーパス』に対する時間表現・事象表現間の時間的順序関係アノテーション」 *自然言語処理*, 20:5, pp. 657–682.