

# 国立国語研究所学術情報リポジトリ

## 〈論文紹介〉

保田祥,小西光,浅原正幸,今田水穂,前川喜久雄

「『現代日本語書き言葉均衡コーパス』に対する時間情報表現・事象表現間の時間的順序関係アノテーション」言語処理学会誌『自然言語処理』20(5): 657-681. (2013)

メタデータ	言語: jpn 出版者: 公開日: 2015-10-30 キーワード (Ja): キーワード (En): 作成者: 加藤, 祥 メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00000784">https://doi.org/10.15084/00000784</a>

保田祥, 小西光, 浅原正幸, 今田水穂, 前川喜久雄  
『現代日本語書き言葉均衡コーパス』に対する  
時間情報表現・事象表現間の時間的順序関係アノテーション」  
言語処理学会誌『自然言語処理』20(5) : 657-681. (2013)

## 加藤 祥

本稿は、テキスト中に記述される事象の生起順序解析を目的とし、アノテーション（タグづけ）基準の再定義を行うとともに、『現代日本語書き言葉均衡コーパス』（BCCWJ）コアデータの一部に付与した時間情報表現・事象表現間の時間的順序関係アノテーションの問題点と課題を考察した。英語においては時間情報と事象とを関連づけるアノテーション基準 TimeML や新聞記事にアノテーションを行ったコーパス TimeBank（Pustejovsky et al. 2003）の整備が行われてきた。しかし、日本語では事象の生起順序関係解析のための基礎データは現在まで存在していない。そこで本稿では、TimeML の事象表現を表す〈EVENT〉・事象のインスタンスを表す〈MAKEINSTANCE〉・時間情報表現・事象表現間の時間的順序関係を表す〈TLINK〉タグに基づいたアノテーション基準を日本語向けに再定義し、BCCWJ の一部データにアノテーションを実施した。

日本語特有の性質として、語形によって表出するテンス・アスペクトと時間軸上の過去・現在・未来が必ずしも対応しない一方、豊かなモダリティを表出しているという性質がある。モダリティのアノテーション研究は国立国語研究所の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」にて松吉ら（2011）が行っているほか、田中ら（2013）によっても進められている。このようなモダリティのアノテーション研究により、事象表現が実世界で起きたか否かという事実性解析が可能になった。しかし、それではいつ・どの順序で事象が生じたのかという問題が残る。

このような背景のもと、本稿は TimeML の事象分類を工藤の研究（工藤 1995, 工藤編 2004）に基づいて再定義し、陽に表出されない日本語の時間的順序関係について、3 人の作業員による被験者実験としてアノテーションを行った。本稿の貢献として、日本語学・言語学の事象表現の時間構造の知見を言語資源として国際的に共有可能にしたほか、揺れを許容したアノテーションを行い、揺れを定量的・定性的に評価するという新しい分析手法を提案した点がある。作成されたデータを学習データとして用いた機械学習に基づくテキスト中の事象表現の時間的順序関係推定方法が提案される（吉川ら 2014）など、工学的な利用も進められている。

今後は、TimeML で行われている主節・従属節間の関係について研究を進めたい。具体的には、有田（2007）の完全・不完全時制節性の概念を TimeML の〈SLINK〉タグとして形式化する。また引き続き、日本語学・言語学で進められてきた言語分析をアノテーションの形

で言語資源として表現し、工学応用に発展させるよう努めて行く。

●参考文献●

- 有田節子(2007)『日本語条件文と時制節性』東京：くろしお出版。
- 工藤真由美(1995)『アスペクト・テンス体系とテキスト—現代日本語の時間の表現—』東京：ひつじ書房。
- 工藤真由美(編)(2004)『日本語のアスペクト・テンス・ムード体系—標準語研究を超えて—』東京：ひつじ書房。
- 松吉俊・佐尾ちとせ・乾健太郎・松本裕治(2011)「拡張モダリティタグ付与コーパスの設計と構築」『言語処理学会第17回年次大会発表論文集』147-150。
- Pustejovski, James, Patric Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo (2003) The TimeBank Corpus. *Proceedings of Corpus Linguistics 2003*, 647-656.
- 田中りべか・戸次大介・川添愛(2013)「MCN コーパス：ガイドライン設計とその運用」『言語処理学会第19回年次大会発表論文集』77-80。
- 吉川克正・浅原正幸・飯田龍(2014)「BCCWJ-TimeBank を対象とした時間的順序関係の推定」『言語処理学会第20回年次大会発表論文集』1103-1106。

**加藤 祥** (かとう・さち)

国立国語研究所コーパス開発センター プロジェクト研究員。博士(文学)(神戸大学)。2012年9月より現職。