

形式意味論に基づく出来事間関係認識に向けて リソース構築の展望とテンス「タ」のアノテーション

宇津木 舞香^{1,a} 稲田 和明^{2,b} 金子 貴美^{3,d} 戸次 大介^{3,4,5,e} 乾 健太郎^{2,4,5,c}
¹ お茶の水女子大学理学部情報科学科 ² 東北大学情報科学研究科
^ag1120508@is.ocha.ac.jp {^bkazuaki.inada,^cinui}@ecei.tohoku.ac.jp
³ お茶の水女子大学大学院人間文化創成科学研究科
{^dkaneko.kimi,^ebekki}@is.ocha.ac.jp
⁴ 国立情報学研究所 ⁵ 独立行政法人科学技術振興機構, CREST

1 はじめに

現在、意味処理の分野では、出来事間の時間関係や含意関係の推定などの「複数の出来事間の関係」の解析が盛んに行われている。これらの解析では、人手で正解を付与したアノテーション付きコーパスによる機械学習手法がよく採用されており、たとえば時間関係の推定で約6割の精度を達成する [1] など、一定の成果をあげている。しかし、このアプローチは以下の二つの本質的な問題に直面している。

第一の問題は、出来事と、それを指すテキスト上の範囲の対応が自明ではないことである。先行研究の多くは、「毎日飲んでた」という表現において「飲んで」のように「述語の語幹+活用語尾」を出来事の表現と見なしているが、これは出来事間の関係の単位として適切とは言えない。たとえば、「三ヶ月間毎日走ってから水を飲んだ」では、「走る」「飲む」という「述語の語幹+活用語尾」が表す出来事間の時間関係が毎日繰り返されている、という読みが自然であるが、「三ヶ月間毎日走ってからマラソン大会に出場した」では、「三ヶ月毎日走る」「出場する」という出来事間の時間関係であり、前者には時間副詞「三ヶ月間」、量化表現「毎日」を含めなければならない。量化表現に限らず、モダリティや談話関係についても同様である。

第二の問題は、出来事間の関係を分類するラベル体系が自明ではないことである。「述語の語幹+活用語尾」の外側には、「テンス、アスペクト、量化、モダリティ、従属節・談話関係」の5つが重なり、複数の出来事間の関係に影響する。さらに、これらが複数段階埋め込まれる場合を考慮すると、任意の順序で任意の回数出現しうるため、二つの出来事間の関係を10数個のラベルに分類することは不自然である。

したがって、現状において出来事間の関係を記述したアノテーション付きコーパスは、多くのノイズが混ざったものとならざるを得ない。単純にリソースを増やすことが性能向上に必ずしもつながらないと考える理由である。「正解」の概念的正当性も疑わしい。

この二つの問題を、どのように捉えるべきであろうか。以下の3つの論点がポイントとなる。

1) 出来事とテキスト上の範囲の対応は現代的な意味論の体系で表すことによって解決される。テキスト上の範囲には、統語構造と、そこからマップされる意味表示が存在し、「出来事」とはその意味表示中の要

素であるが、日本語の統語構造と意味合成の仕組みについては大部分が解明されている [2]。

- 2) 複数の出来事間の関係はいずれも演繹関係に還元される。たとえば「太郎が毎日来たかもしれない」というテキストにおいて「来た」という出来事の実在性とは、「太郎が毎日来たかもしれないト 太郎が来た」という含意関係に他ならない。時間的前後関係や含意関係認識も同様に定式化できる。
- 3) テンス、アスペクト、量化、モダリティ、従属節・談話関係が演繹関係にもたらす影響は、各表現の意味表示が決定する。形式意味論の意味表示は再帰的構造であり、出来事間の関係の再帰的構造をもっとも直裁的に表す体系であり、我々はこれらの各要素について記述・計算が可能な意味論の体系を確立している [3]。

ここで問題となるのは「テンス、アスペクト、量化、モダリティ、従属節・談話関係」などの表現に存在する曖昧性である。たとえばテンスでは、後述するように、過去の助動詞「タ」には、意味表示が異なるという観点から分類すると、少なくとも14個の独立した用法がある。テキストが与えられたとき、これらの表現の曖昧性が解消できない限り、テキストを意味表示にマップすることはできない。

しかしその一方で、筆者らのこれまでの研究によって、モダリティ [4, 5] と談話関係 [6] に関しては、各表現がもつ意味表示が異なる用法は数え上げることができ、「ラベル判定問題」に還元することができる見通しを得ている。よって、信頼度の高いアノテーション付きコーパスを作成し、機械学習によってラベル分類を行うことで、各表現を意味表示にマップし、文の意味を意味表示から合成することで導出でき、演繹関係は証明探索の問題に帰着することになる。

このような見通しから、テキスト内の出来事についての事実性・確実性判断、出来事間の時間関係認識、含意関係認識などの「複数の出来事間の関係」を問うタスクを対象とした研究プロジェクトを始動する (図1参照)。

- i) テンス、アスペクト、量化、モダリティ、従属節・談話関係のうち、主要な表現のもつ異なる用法について、MCN コーパスの方法論を用いてアノテーション体系を設計し、アノテーション付きコーパスを作成する。
- ii) 各表現の異なる用法を判定する機械学習ベースの判

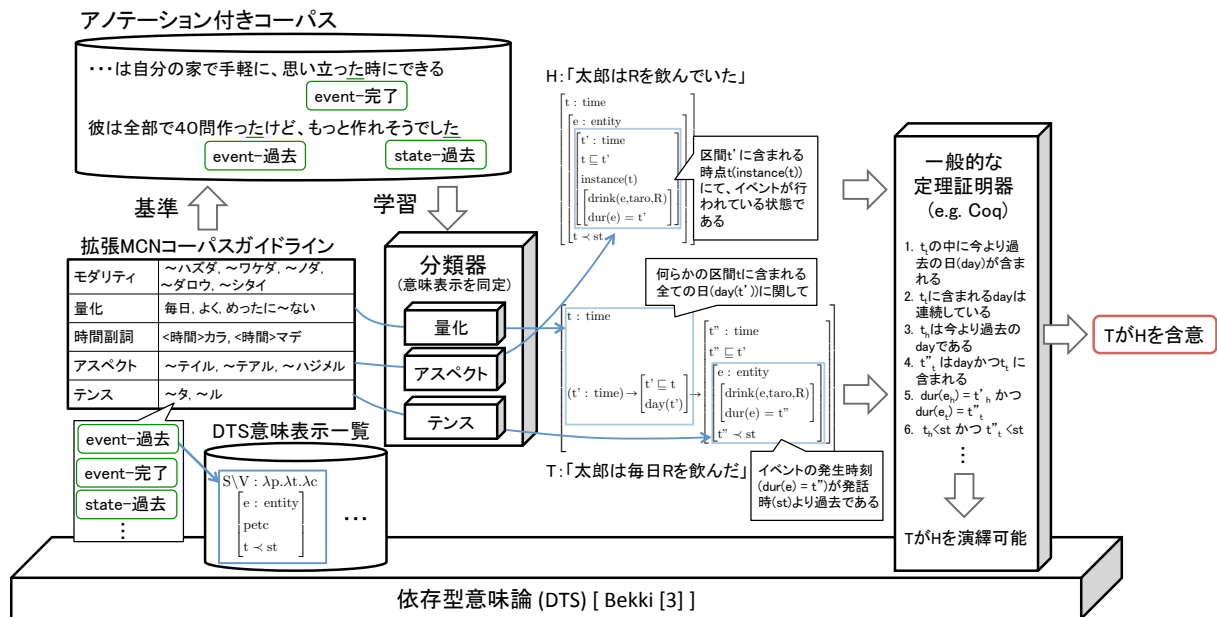


図 1: 形式意味論に基づく出来事間関係認識へ向けたプロジェクトの全体像

定器を開発する。

- iii) 各表現の異なる用法について、依存型意味論による意味表示を与え、出来事間の関係は意味表示間の演繹関係として定式化し、証明探索等によって問題を解く。

本論文は、当研究プロジェクトの最初の試みとして、日本語のテンスのうち過去の助動詞 [～タ] に焦点をあてたものである。はじめに日本語のテンスについて述べたのち、本研究において設計した [～タ] のアノテーション体系と、それに基づいて作成した [～タ] コーパスについて紹介する。

2 日本語のテンス

テンスとは、文書中の出来事の時刻を発話時を基準として位置づける文法形式を指す。テンスはほぼすべての文に現れ、時間関係はもとより含意関係の観点からも非常に重要である。日本語ではテンスに関わる言語表現として助動詞 [～タ]、終止形活用語尾 [～ル] が挙げられる。一般に、[～ル] は非過去、[～タ] は過去を表すとされる。

Kamp ら [7] では eventuality (出来事) は event が state のどちらかであるとしている。中村 [8] によると、非過去の場合に state が発話時と重なることができるのに対し、event は発話時と重なることができないということから、event/state 間にはテンスにおける振る舞いにおいて違いがある。この違いは [～タ] のもたらす意味にも影響し、event に接続する [～タ] と state に接続する [～タ] はそれぞれ異なる意味表示を持つ。

以上のことから、文書中の出来事の時間的な位置づけを把握するには [～タ][～ル] が後接する述語の eventuality が event/state のどちらであるかの判断と、eventuality が [～タ] や [～ル] によってどのように位置づけられるかの判断を行う必要がある。3 節では eventuality の特定方法を提案し、4 節では event に後続する [～タ]、state に後続する [～タ] の分類についてそれぞれ述べる。

3 eventuality

3.1 eventuality を決定する要素

本節では eventuality の特定方法を提案する。実テキスト上には様々な語句が結びつくため、event/state の判断が自明でないケースが存在する。一般に文中の出来事は述語によって表現される。動作を表す動態述語と状態の存在を表す状態述語に分けられ、動態述語は event、状態述語は state にそれぞれ分類される。これらの述語に対して以下に示す 4 種類の要素が接続する度に eventuality が変化する。

- アスペクトに関する表現: eventuality の中の様々な局面を表す文法形式であり、その例として [～ハジメル] [～テイル] [～オワル] などが挙げられる。たとえば、[食べる] という event はアスペクト表現が接続することで、[食べ始める] [現在食べている] [食べ終わる] などの局面を表現する。
- 否定表現: [～ナイ][～ズ][～ヌ] の形をとる語句であり、否定表現と接続したフレーズは直前の eventuality によらず state となる。従って、動態動詞 [食べる] は event であるが、助動詞 [～ナイ] に接続した形の [食べない] は state となる。
- モダリティに関する表現: あるフレーズに対する書き手の認識や態度を表す言語表現であり、[～ヨウダ] [～モノダ] [～ハズダ] などの接尾語によって表現される。これらの語句が接続したフレーズは直前の eventuality によらず state となる。
- テンスに関する表現: 2 節で述べた通りである。テンスに関する表現の接続では eventuality の変化は生じず、直前のフレーズの eventuality をそのまま保持する。

3.2 eventuality の特定方法

本研究では、上記の各要素が述語と接続した際に event/state のどちらを指定するかを判断するために、戸次 [2] の品詞分類に基づき一覧を作成した。表 1 はその抜粋である。以降では、表 1 を用いていくつかの語句の eventuality を特定する過程を示す。

表 1: eventuality の event/state の分類

品詞分類	eventuality	特徴	例	
動詞	event	動態動詞	動きを表す動詞	アルク、タベル、ノム
	state	状態動詞	状態を表す動詞	アル、イル、デキル
連用形複合動詞	event			ナグリタオス
形容詞	state			アカイ、ウツクシイ
状詞	state			キレイダ、シスカダ
動詞性接尾語	event	形式動詞	動きを表す形式動詞	～スル、～ナル
	state		状態を表す形式動詞	～イル、～アル
	event	動詞の連用形に接続	アスペクトに關係	～ハジメル、～ダス
			完遂の意味を表す	～ツクス、～ヌク
			不遂の意味を表す	～ウスレル、～ソコナウ
			その他	～アウ、～ナオス
			アスペクトに關係	～デル、～テアル
	state	動詞のテ形に接続	授受に關係する	～テモラウ、～テアル
	* ¹		その他	～テオク、～テミル
	event	形容詞/名詞の語幹に接続		～ガル
	event	普通名詞に接続		～メク、～バム
	* ¹	應	尊敬	～レル
	event		受身	～レル、～ル
	event		使役	～セル
state		可能	～ル、～レル、～ウル	
状詞性接尾語	state		～ヨウダ、～ハスタ	
形容詞性接尾語	state	否定の意味をもつ	～ナイ	
		動詞の連用形に接続	～タイ	
		タロウ接続形に接続	～ラジイ	
		イ形容詞/状詞語幹に接続	～ツボイ	
		名詞に接続	～クサイ	
		動詞の終止形に接続	～ベキ、～ベシ	
		テ形に接続	～ラジイ	
		その他	～カモシレナイ	
			～タ、～ダ	
			～デス、～ナイ、～ダロウ	

¹ * は直前の eventuality を保持する

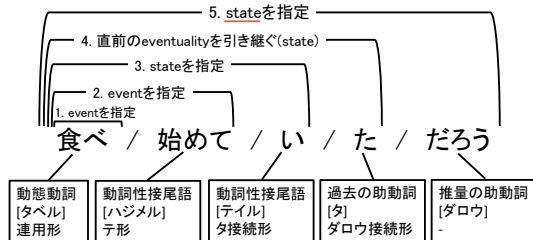


図 2: 「食べ始めていただろう」の eventuality の特定

まず、動詞、連用形複合動詞、形容詞、状詞¹などの単一の要素からなる語句は、他の要素と接続せずに eventuality を指定することができる。たとえば、動態動詞「歩く」や連用形複合動詞「殴り倒す」は event となり、状詞「静かだ」や形容詞「楽しい」は state となる。

次に、複数の要素が接続された複雑なフレーズの eventuality を特定する例として、「食べ始めていただろう」という表現について考える。図 2 に示す通り、「食べ始めていただろう」は、戸次の品詞分類 [2] に従い 5 つの要素に分解された後、1~5 の順で eventuality が指定され、最終的に state が指定される。このように表 1 の event/state の分類表は、複雑な語句に対する eventuality の特定にも対応している。

しかしながら、慣用的な意味をもつ語句に対しては、表 1 のみでは eventuality の判断が困難な事例も存在する。以下に例を示す。

(1) サプリというサプリはすべて試した。

たとえば、(1) に現れる「いう」は、直感的には event とは言い難いが、動態動詞「イウ」の連体形であることから、表 1 に従うと event を挿入してしまう。このような慣用的な用法を持つ語句については、表 1 のような品詞による分類ではなく、MCN コーパスガイドライン [5] による分類を使用し、各分類ごとに event/state の判断を指定する必要があると考えている²。

¹ 戸次 [2] では、状詞の概念は、形容動詞と、名詞+判定詞、副詞、連体詞にまたがる分類であるとしている。

² MCN コーパスは、様相表現・条件表現・否定表現といった出来

4 助動詞 [～タ]

助動詞 [～タ] は eventuality の時刻を過去に位置づけるはたらきをもつ。以下の (2) では、動態動詞「食べる」により指定された event を、(3) では、状詞「サンマだ」により指定された state を過去に位置づけている。

(2) 今日は 7 時に朝ご飯を食べた。

(3) 昨日の晩ご飯はサンマだった。

しかし、助動詞 [～タ] は eventuality を過去に位置づけるだけに留まらず、様々な意味を持つことが言語学において知られている。

(4) さあ行った行った!

(5) あ、そういえば午後からミーティングだった。

これらは eventuality を過去に位置づける意味を持たない [～タ] の例であり、(4) は命令のニュアンスを、(5) は発話者が何かを思い出すというニュアンスをもたらし。よって、日本語で記述されたテキストから過去の eventuality を抽出するためには、eventuality を過去に位置づける [～タ] とそれ以外の [～タ] を区別しなければならない。

4.1 助動詞 [～タ] のガイドラインの概要

本研究では、過去の意味をもたらし [～タ] とそれ以外の [～タ] を区別してアノテーションするために、MCN コーパスのガイドライン設計 [4] と井上による「タ」の分析 [9] に基づき、表 2 に示す [～タ] のアノテーションガイドラインを作成した。

全カテゴリ数は 14 種類であり、event に後接する [～タ] には 5 種類、state に後接する [～タ] には 9 種類のカテゴリを設けた。このうち eventuality を過去に位置づけるはたらきを持つカテゴリは、event に接続する [event-過去][event-完了]、state に接続する [state-過去][state-完了] の 4 種類であり、eventuality を過去に位置づける働きを持たない [～タ] については、[event-動作開始時] やムード (話し手の心的態度) を表すようなカテゴリ 10 種類を設けている。(4)(5) はどちらもムードの例であり、(4) は [event-命令]、(5) は [state-思い出し] のカテゴリに属する。

4.2 助動詞 [～タ] へのアノテーション方法・結果

本研究では現代日本語書き言葉均衡コーパス³の書籍ドメインのテキストの一部に対し [～タ] のアノテーションを行った。書籍ドメインのコアデータからランダムにサンプルした 5 つのドキュメントに対して形態素解析⁴を行い、表層が「タ」もしくは「ダ」であった 533 件をアノテーション対象とした。

表 1 に従い、予め [～タ] の直前の eventuality の属性情報 (event/state) を付与したテキストに対し、言語学専攻でない大学院生 2 名が [～タ] の意味分類を行った。今回用いた [～タ] の分類一覧を表 2 に示す。今回のアノテーション作業における単純一致率は 0.80、Kappa 値は 0.65 であった。単純一致率は、以下の式で算出した。

$$\text{単純一致率} = \frac{\text{カテゴリ一致数}}{\text{全体数}}$$

事の確実性に関わる表現について用法分類をアノテーションした言語データ [4] である。MCN コーパスではガイドラインに言語学的テストを用いることによって、各分類の独立性を保証している [5]。

³ <http://www.ninjal.ac.jp/corpus.center/bccwj>

⁴ MeCab (<http://code.google.com/p/mecab/>) を使用した。

表 2: 助動詞 [～タ] のガイドライン

直前の eventuality	カテゴリ	特徴	例文	言語学的テスト
event	過去	A. event が過去に位置することを表す B. 対応する否定形は [(結局) シナカッタ] であり、当該の event が実現想定区間内に実現されないまま終わったことを表す	「8時に朝ご飯を食べた」「(その日の夕方に) 今日、朝ご飯食べた?」「明日の決勝で勝ったチームがオリンピックへの出場権を得る」	
	完了	A. 現在までに出来事が実現済みという現在完了的な意味を表す B. 実現済みの意味をともなう副詞 [もう、すでに、とくに] と共起される C. 対応する否定形は [(マダ) シナイ] であり、発話時が当該の出来事の実現想定区間にあり、出来事の実現が確定されないうちに使用できる	「私の父は、ガンでもう死んだ」「(レポート提出日は翌日) レポートもう出した?」「ありふれた話」	[シテイル] に置き換え可能
	動作開始時	A. 発話時と動作の開始時点が同時であることを表す	(野球の実況中継) イチロー、走った!」「(赤ちゃんを眺めていたら、たまたま笑い出した) あ、笑った!」	[～ハジメル+タ] に置き換え可能
	ムード 仮定 命令	A. 仮定の事項を述べる [～トスル] の補文に用いられる	「仮に走ったとします」「さあ、行った行った!」「買った、買った!」	[タ+トスル] の形式であり、現在形に置き換え可能 命令形に置き換え可能
state	過去	A. 過去において存在した (観測された) 状態であることを表す B. 対応する否定形は [(結局) シナカッタ] であり、当該の event が実現想定区間内に実現されないまま終わったことを表す	「太郎は自宅にいた」「太郎は子供に甘ずきた」「昔はここから富士山が見えた」「昔は英語が話せた」	
	完了	A. 当該の出来事、ある過去の時点までに実現済みの出来事、あるいはある過去の時点までの経過・経歴として述べる B. 現時点において、それよりも前に実現した運動がひきつづき開き、効力を持っている	「父は既に3時間前に死んでいた」	[テイル+タ] の形式
	ムード 発見 判明	A. 発話時に知られた状態が過去のある時点から成立している事象を述べる B. 発話時において存在することが明らかでない状態に対して、過去形を用いて「発話時以前にその状態が存在した」ということを述べる B. 発話時以前に「観察行為＝状態の判明」のプロセスが存在したことを暗示する	「あ、ここにあった」「おや、こんなところに銭があった」「(今日、CDを買った) ベートーベンの「第九」だった」「X: ちょっとこれ味見してみて、Y: いいよ、X: どう? おいしかった?」	現在形に置き換え可能 観察行為 [～デミル+タラ] 等の語句が挿入可能
		思い出し	A. その場の問題解決に有効な情報が話し手自身の過去の体験の中に見いだされたことを述べる	「そうか、そういえばそういう手があったねえ」「そういえば、そんな本があったねえ」「ええと、あなたは井之上さんでしたか?」「今日は何曜日でしたか?」
	修正補強	A. 話し手が新規情報を受け入れつつあることを表す [...] ノカ [疑問情報受容文] において用いられる B. 発話時以前の状態に関する話し手の認識を修正し、現在の認識との間のギャップを解消させる	「(取り物が予想外の場所で見つかった) なんだ、こんなところがあったのか?」「(誰だろう、と振り向いて) 何だ、井之上さんでしたか?」	[～タ] の直後に疑問情報受容文 [...] ノカ が接続
	仮定文	A. 仮定の事項を述べる [...] トスル] の補文に用いられる	「ここに100万円あったとします、あなたなら何に使いますか?」「Xと2xとの間に1対1対応があったと仮定せよ」	[タ+トスル] の形式であり、現在形に置き換え可能
	正解	A. 発話時まで隠していた正解を述べる際に用いられる B. 聞き手の認識を修正補強する	「正解は3番でした」	
	他		「ありがたいございました」「ご苦労様でした」	

これらの値から、今回のアノテーションは比較的高い一貫率で作業できたと考えられる。

表 3 に各カテゴリにおける作業間の一貫数を示す。この結果から、今回のアノテーションでは [event-過去][event-完了] の判断について揺れが生じていたことが分かる。このような判断の揺れは、[event-過去] と [event-完了] の言語学的テストが明確ではないことに起因しており、少なくとも次に述べる 3 点を改善する必要があると考えられる。1 点目は、[event-完了] がもたらす現在完了の意味が、共起される副詞句 [もう][既に] などの [～タ] 以外の要素からもたらされる可能性が残っていることである。2 点目は、[event-完了] の言語学的テストには [～テイル] が用いられている。[～テイル] には「進行」「経験」「結果残存」「習慣」の 4 つの用法があるといわれているが、今回のテストでは、これらを区別するための判断基準を設けなかった。このことが判断の揺れに繋がったと推察されるため、[～タ] と同様に、[～テイル] の分類も整理する必要があると考えられる。3 点目は、[event-過去] と [event-完了] の区別が明確ではないことである。[event-過去][event-完了] の区別は 1 文のみでは難しく、前後の文脈を参照する必要があるケースが存在する。これらのことを踏まえ、言語学的テストの設計を再検討すべきである。

また、今回コーパスとして書籍ドメインのテキストを使用した。付与されたラベルの多くが [event-過去][event-完了][state-過去][state-完了] であり、ムードに関する [タ] はあまり現れなかったため、アノテーションに使用するコーパスについても再度検討する必要がある。

5 まとめ

本稿では複数の出来事間の関係を、形式意味論を用いて判定する枠組みを提案した。出来事間の関係に影響をもたらす要素としてテンス、アスペクト、量化、モダリティ、従属節・談話関係が挙げられるが、これらの表現に存在する曖昧性が問題となっている。本研究ではテンスのうち過去の助動詞 [～タ] に焦点をあて曖昧性解消のためのリソース構築を試みた。[～タ] の曖昧性を解消するために、eventuality に対する event/state の分類が重要となるが、我々は分類表を作成し、述語語幹に様々な要素が接続するような複雑なフレーズに対しても、

表 3: 助動詞 [～タ] へのアノテーション結果

カテゴリ	作業者 A	作業者 B	一致数
event-過去			362 ∩ 354 = 328
event-完了			20 ∩ 38 = 4
event-動作開始時			1 ∩ 0 = 0
event-ムード-仮定			2 ∩ 0 = 0
event-ムード-命令			0 ∩ 0 = 0
state-過去			77 ∩ 89 = 65
state-完了			40 ∩ 43 = 30
state-ムード-発見			0 ∩ 0 = 0
state-ムード-判明			4 ∩ 3 = 1
state-ムード-思い出し			9 ∩ 5 = 2
state-ムード-修正補強			2 ∩ 2 = 0
state-ムード-仮定			2 ∩ 0 = 0
state-ムード-正解			1 ∩ 0 = 0
state-他			0 ∩ 0 = 0
判定不能			8 ∩ 19 = 3
合計			533

event/state を指定することを可能にした。さらに、助動詞 [～タ] について、曖昧性解消のためのガイドラインを作成し、アノテーションを実施した。

参考文献

- Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Proceedings of SemEval 2013*, pp. 1–9, 2013.
- 戸次大介. 日本語文法の形式理論・活用体系・統語構造・意味合成 -, 第 24 巻. くろしお出版, 2010.
- Daisuke Bekki. Representing anaphora with dependent types. In *Logical Aspects of Computational Linguistics*, pp. 14–29. Springer, 2014.
- 川添愛, 齊藤学, 片岡喜代子, 崔崇殊, 戸次大介. 言語情報の確実性に影響する表現およびそのスコープのためのアノテーションガイドライン ver. 2.4. Technical report, Ochanomizu University, OCHA-IS 10-4, 2011.
- 田中リベカ, 小池惠里子, 戸次大介, 川添愛. 言語学的テストに基づく意味アノテーションのガイドライン設計 確実性判断に関わる表現を中心に. *NLP2012*, pp. 401–404, 2012.
- Kaneko Kimi and Bekki Daisuke. Toward a discourse theory for annotating causal relations in japanese. *The 28th Pacific Asia Conference on Language, Information and Computing*, 2014.
- Hans Kamp and Uwe Reyle. *From discourse to logic*. Dordrecht, 1993.
- 中村ちどり. 日本語の時間表現, 第 14 巻. くろしお出版, 2001.
- 井上優. 現代日本語の「タ」—主文末の「...タ」の意味について—. 「た」の言語学. ひつじ書房, 2001.