

# 弱教師あり学習によるイベントの意志性・主語有生性の分類の同時学習

清丸 寛一<sup>†</sup>・黒橋 禎夫<sup>†</sup>

意志性と主語有生性はイベントの基本的な属性であり、密接な関係にある。これらの認識は文脈を考慮したテキスト理解を必要とし、その学習には大量のラベル付きデータを要する。本論文では、人手でラベル付きデータを構築することなく、意志性と主語有生性を同時学習する手法を提案する。提案手法では生コーパス中のイベントにヒューリスティクスを用いてラベルを付与する。意志性のラベルは「わざと」や「うっかり」といった意志性を示す副詞を頼りに付与する。主語有生性のラベルは知識ベースに登録されている有生名詞・無生名詞を頼りに付与する。こうして集めたイベントから手がかり語を含まないイベントに汎化する分類器を構築する。本研究ではこの問題をバイアス削減ないしは教師なしドメイン適応の問題とみなして解く。日本語と英語の実験で、提案手法により、人手でラベル付きデータを構築することなく、意志性・主語有生性の高精度な分類器を構築できることを示した。

キーワード：イベント，意志性，有生性，弱教師あり学習

## Minimally-Supervised Joint Learning of Event Volitionality and Subject Animacy Classification

HIROKAZU KIYOMARU<sup>†</sup> and SADA O KUROHASHI<sup>†</sup>

Volitionality and subject animacy are fundamental and closely related properties of an event. Their classification is challenging because it requires contextual text understanding and a huge amount of labeled data. This paper proposes a novel method that jointly learns volitionality and subject animacy at a low cost, heuristically labeling events in a raw corpus. Volitionality labels are assigned using a small lexicon of volitional and non-volitional adverbs such as “deliberately” and “accidentally”; subject animacy labels are assigned using a list of animate and inanimate nouns obtained from ontological knowledge. We then consider the problem of learning a classifier from the labeled data so that it can perform well on unlabeled events without the words used for labeling. We regard the problem as a bias reduction or unsupervised domain adaptation problem and apply the techniques. We conduct experiments with crowdsourced gold data in Japanese and English and show that our method effectively learns volitionality and subject animacy without manually labeled data.

**Key Words:** *Event, Volitionality, Animacy, Weakly-supervised Learning*

<sup>†</sup> 京都大学大学院 情報学研究科, Kyoto University, Graduate School of Informatics

## 1 はじめに

**意志性 (volitionality)** はイベントの基本的な属性であり、イベントに何者かの意志的な関与があるかどうかを表す。本研究では特にイベントの主語が表すエンティティがイベントに意志的に関与しているか否かに着目する。例えば、主語のエンティティの観点から見て、「食べる」や「書く」といったイベントはふつう意志的 (volitional) であり、「泣く」や「怪我をする」、「怒られる」といったイベントは非意志的 (non-volitional) である。イベントの意志性分類は、因果関係知識の類型化 (Lee and Jun 2008; Inui et al. 2003; Abe et al. 2008a, 2008b) に用いられてきたほか、条件付きイベント予測 (Du et al. 2019)、スクリプト抽出 (Chambers and Jurafsky 2008)、顧客フィードバック分析 (Liu et al. 2017) などへの応用がある。

一方、有生性 (animacy) は名詞の属性であり、名詞が表すエンティティに人間のような意志的な行為が可能かどうかを表す。本研究ではイベントの主語が表すエンティティがイベントに意志的に関与しているか否かに着目するため、イベントの主語が有生名詞であることはイベントが意志的であることの必要条件となる。この密接な関係に着目し、本研究では**主語有生性**というイベントの属性を考える。意志性の学習では主語有生性の同時学習が助けになると期待される。

意志性を認識する難しさは、言語資源の不足と文脈理解が必要なことにある。意志性は多くの場合、イベントの述語によって同定できる。冒頭の「食べる」や「泣く」などがそうである。しかし、意志的 (あるいは非意志的) な行為を表す述語を網羅したリストは存在しない。また、たとえそうした言語資源があったとしても、述語だけではなく、その文脈も考慮しなければ意志性を同定できない場合が存在する。例えば、例 (1-a) と例 (1-b) の述語はどちらも同じ「浴びる」であるが、前者は意志的、後者は非意志的である<sup>1</sup>。

- (1) a. シャワーを浴びる. (V)  
b. 非難を浴びる. (NV)

また、例 (2-a) は非意志的であるが、例 (2-b) は「深く」という副詞を伴うことで意志的となる。

- (2) a. 息をする. (NV)  
b. 深く息をする. (V)

文脈理解の問題は言語資源の整備によって解決するのは困難である。あらゆる文脈—述語の項、項への連体修飾、述語への修飾 (副詞句) の組み合わせ—に対して意志性のラベルをアノテ

<sup>1</sup> 意志的なイベントの例は「V (volitional の略)」, 非意志的なイベントの例は「NV (non-volitional の略)」を付記して示す。

ションすることは非現実的だからである。この問題に対する有望な解決策は、イベントを構成する語句の意味とそれらの関係性を柔軟に捉えて意志性を認識する分類器を構築することである。そうした柔軟な分類器は深層学習モデルを訓練することで得られると期待されるが、その訓練には通常、大量のラベル付きデータが必要となる。

主語有生性の認識についても、意志性の認識と同様の難しさがある。まず、言語資源の不足の問題がある。主語有生性は、たいていの場合、主語の名詞が通常有生名詞 (animate noun) か無生名詞 (inanimate noun) かによって同定できる。有生名詞・無生名詞は ConceptNet (Speer et al. 2017) などの知識ベースから一定量のリストが得られるが、網羅的とは言いがたい。また、主語有生性の認識においても文脈理解が必要となる場合がある。例えば、例 (3-a) の主語「白バイ」は通常無生名詞であるが、例 (3-b) の主語「白バイ」は警察官の換喩であり、この文脈においては有生名詞である<sup>2</sup>。

- (3) a. 白バイが停まっている。(IA)  
b. 白バイが追いかけてくる。(A)

こうした現象に対処するには、やはり柔軟な文脈理解が可能な分類器を構築するのが有望であり、その訓練には大量のラベル付きデータが必要となる。

本研究では、イベントの意志性と主語有生性を同時学習する弱教師あり学習手法を提案する。提案手法の概要を図 1 に示す。提案手法ではまず、ヒューリスティクスを用いて生コーパス中のイベントにラベルを付与する。意志性のラベルは「わざと」などの意志的な行為を表す副詞 (意志的副詞) と「うっかり」などの非意志的な行為を表す副詞 (非意志的副詞) を手がかりに付与する。例えば、例 (4) は意志的副詞「あえて」が述語に係っているため、意志的であるとみなす。例 (5) は、非意志的副詞「うっかり」が述語に係っているため、非意志的であるとみなす。

- (4) あえて真実を話す。(V)  
(5) うっかり携帯を落とす。(NV)

主語有生性のラベルは既存の言語資源に登録されている有生名詞・無生名詞を手がかりに付与する。生コーパスの量は際限なく増やすことが可能であるため、この方法で大量のラベル付きデータを低コストで収集することができる。

例 (4)、例 (5) が示唆するように、意志的副詞・非意志的副詞を除いたとしても、多くの場合、イベントの意志性は保持される。これは主語有生性に関しても同様である。しかし、そうでな

<sup>2</sup> 主語が有生名詞であるイベントの例は「A (animate の略)」、主語が無生名詞であるイベントの例は「IA (inanimate の略)」を付記して示す。

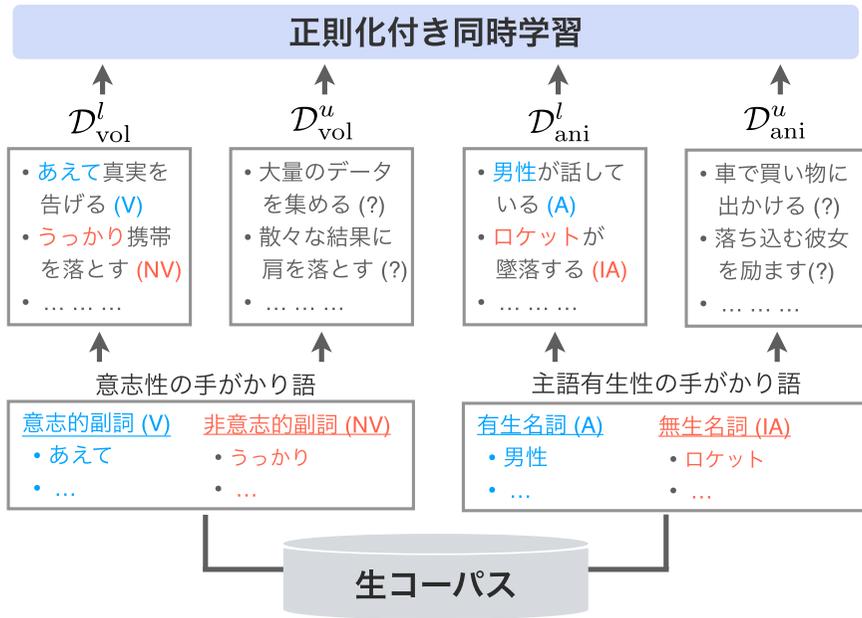


図 1 提案手法の概要. 生コーパス中のイベントにヒューリスティクスを用いて意志性・主語有生性のラベルを付与し, 意志性・主語有生性それぞれのラベル付きデータセット ( $D_{vol}^l \cdot D_{ani}^l$ ) とラベルなしデータセット ( $D_{vol}^u \cdot D_{ani}^u$ ) を得る. その上で, 意志性と主語有生性の分類を同時学習する. その際, 手がかり語だけに着目した分類に陥ることを防ぐための正規化を導入する.

い場合もある. 例えば, 例 (6-a) は意志的であるが, そこから「あえて」を除いた例 (6-b) は非意志的である.

- (6) a. あえてこける. (V)
- b. こける. (NV)

このように, 意志的副詞と共起するイベントが必ずしも意志的なイベントであるとは限らない.

主語有生性に関してもこうした例が存在する. 例えば, 例 (7-a) の主語である「衝撃」は無生名詞であるが, 「衝撃」を除いた例 (7-b) の主語は有生名詞として捉えるのが妥当である.

- (7) a. 衝撃が走る. (IA)
- b. 走る. (A)

手がかり語を常に含むラベル付きデータから, 手がかり語を含まないラベルなしイベントに汎化する分類器を得るには, 原則として手がかり語に頼らず, それと共起するテキストからラベルを予測することを学習しつつ, 手がかり語を除くことでラベルが変化する例に関しては, 手

がかり語と共起するテキストからラベルを予測することを学習しないことが重要である。本研究では、分類器を学習する際にラベル付けの手がかり語だけに着目して分類することを抑制する正則化を導入することで前者の原則を学習しつつ、汎用言語モデル (Devlin et al. 2019) が作り出すイベントの汎化ベクトル表現の上で分類器を構築することで、予測のために手がかり語に着目せざるを得ないケースがデータから学習されることを期待する。

本研究は、手がかり語に着目した分類を抑制する問題をバイアス削減あるいは教師なしドメイン適応の問題と捉え、その手法を活用する。バイアス削減はデータセット中に存在する特定のバイアスが予測に濫用されることを防ぐ手法である (Bolukbasi et al. 2016; Zhao et al. 2017, 2019; Kennedy et al. 2020)。本研究ではヘイトスピーチ認識器の学習において利用されているバイアス削減手法を転用する (Kennedy et al. 2020; Jin et al. 2020)。ここで提案されているバイアス削減手法は、分類器が「gay」といったヘイトスピーチに特徴的な単語 (バイアス) だけに着目した分類に陥ることを抑制し、文脈を考慮した分類を促すものである。本研究では、ラベル付けに用いる手がかり語をバイアスとみなして、単純なバイアス削減手法である word removal (WR)、より高度なバイアス削減手法で有効性が知られている sampling and occlusion (SOC) の2つを利用する。

教師なしドメイン適応は、ソースドメインのラベル付きデータとターゲットドメインのラベルなしデータを用いて、ターゲットドメインに汎化するモデルを構築する手法である。本研究の設定は、手がかり語を含むラベル付きデータをソースドメインのデータ、手がかり語を含まないラベルなしデータをターゲットドメインのデータとみなすことで、教師なしドメイン適応の問題として定式化できる。本研究では、深層学習モデルを利用したテキスト分類器の学習において有効性が確認されている教師なしドメイン適応手法 adversarial domain adaptation (ADA) を利用する (Ganin et al. 2016; Ganin and Lempitsky 2015; Shah et al. 2018; Shen et al. 2018)。ADA では、ラベル付きデータのもとで分類を学習しつつ、敵対的学習の枠組みで、ラベル付きデータとラベルなしデータが判別できなくなるようにイベントのベクトル表現を学習する。この学習により、ラベル付きデータにだけ現れる手がかり語になるべく頼らない分類が学習されると期待される。

提案手法の有効性を確認するため、日本語と英語で実験を行った。分類器の性能を評価するため、各言語についてクラウドソーシングで評価データを新たに構築した。実験を通して、提案手法により、人手でラベル付きデータを構築することなく、イベントの意志性・主語有生性の高精度な分類器を構築できることを示した<sup>3</sup>。

<sup>3</sup> 本研究で構築した評価データおよびモデルの実装は公開予定である。

## 2 関連研究

関連研究として、イベントの意志性分類、バイアス削減、教師なしドメイン適応について述べる。

### 2.1 イベントの意志性分類

イベントの意志性分類に関する過去の研究は、視点となる名詞句が与えられる設定と与えられない設定に大別できる。視点となる名詞句が与えられる設定では、述語と名詞句が与えられ、その名詞句が表すエンティティが述語が表すイベントに意志的に関与しているか否かを予測する。この設定は semantic proto-role labeling のサブタスクとして取り組まれている (Reisinger et al. 2015; White et al. 2016; Teichert et al. 2017)。

視点となる名詞句が与えられない設定では、イベントが与えられ、その主語が表すエンティティがイベントに意志的に関与しているか否かを予測する (Abe et al. 2008a, 2008b; Inui et al. 2003)。本研究はこの設定に取り組む。Abe et al. (2008a) と Abe et al. (2008b) は、意志的・非意志的な述語のリストを構築し、それを参照することでイベントの意志性を予測している。この手法は、構築する述語のリストの網羅性に性能が依存するほか、例 (1-a) と例 (1-b) にあるような、述語単独では意志性を同定できない場合に対応できない点に問題がある。Inui et al. (2003) はデータ駆動の手法を取っている。ここでは、少量のラベル付きデータを構築し、手作りの言語特徴量の上で SVM を学習している。しかし、意志性は述語、項、副詞などの修飾要素から非構成的に定まる属性であり、少量の事例から頑健な分類器を学習することは困難である。本研究では、ヒューリスティクスを用いて大量のラベル付きデータを収集し、そのもとで強力な汎用言語言語モデル (Devlin et al. 2019) に基づく分類器を構築することで、意志性の認識に関わる広範な言語現象と世界知識を学習する。

### 2.2 バイアス削減

バイアス削減はデータセット中に存在する特定のバイアスが予測において濫用されることを防ぐ手法である。バイアス削減は機械学習の公平性の分野で研究されている (Bolukbasi et al. 2016; Zhao et al. 2017, 2019; Kennedy et al. 2020)。本研究では、バイアス削減の手法を転用し、手がかり語だけに反射的に反応して分類を行うのではなく、手がかり語の文脈を考慮して分類を行うモデルを学習する。具体的には、バイアス削減の手法として、Kennedy et al. (2020) が提案している word removal と sampling and occlusion (Jin et al. 2020) の2つを利用する。これらの手法は、頑健なヘイトスピーチ分類器の学習のために提案されたものであり、「gay」などのヘイトスピーチに特徴的な単語のみを手がかりに分類を行うことを抑制し、文脈を考慮した予測を促す目的で利用されている。手法の詳細は 4.3 節に譲る。

## 2.3 教師なしドメイン適応

教師なしドメイン適応は、ターゲットドメインのラベルなしデータのみを用いて行うドメイン適応である (Ramponi and Plank 2020). 本研究の問題は、手がかり語を含むラベル付きデータをソースドメインのデータ、手がかり語を含まないラベルなしデータをターゲットドメインのデータとみなすことで、教師なしドメイン適応の問題として定式化できる。教師なしドメイン適応の手法は多数提案されているが、タスク横断的な評価を欠いており、どの手法が優れているとも言えないのが現状である (Ramponi and Plank 2020). 本研究で扱う問題は、広くはテキスト分類に分類されるものである。そこで、本研究では、ニューラルモデルに適用可能でかつテキスト分類タスクで多くの成功を収めている adversarial domain adaptation (ADA) を用いる (Ganin et al. 2016; Ganin and Lempitsky 2015; Shah et al. 2018; Shen et al. 2018). ADA では、ソースドメインのラベル付きデータでタスクを学習しつつ、ソースドメインとターゲットドメインのデータの区別がつかないような潜在空間を学習することを通して、ターゲットドメインにおけるモデルの汎化性能を向上させる。手法の詳細は 4.3 節に譲る。

## 3 問題設定

本研究で扱うイベントの表現、スコープ、意志性・主語有生性のアノテーションについて述べる。

### 3.1 表現

イベントは1つの主たる述語を含むテキスト (節) として表す。イベントは齋藤 他 (2018) の基準にならって認定する。齋藤 他 (2018) は、イベントを述語句構造を基本とする情報構造として定義している。齋藤 他 (2018) の基準では、述語であっても、意味の重要性や事象性が希薄なものはイベントとして認定されない。たとえば、例 (8) の「大きく」は副詞的な形容詞であり、単独のイベントとして認定される代わりに、述語「表示する」が構成するイベントの一部とみなされる。

(8) 大きく表示する。

例 (9) の「思う」は推量・伝聞のモダリティであり、単独のイベントとして認定される代わりに、述語「美しい」が構成するイベントの一部とみなされる。

(9) 美しいと思う。

また、齋藤 他 (2018) の基準では、項が複合名詞である場合や連体修飾されている場合は、それ

らの句・節の全体が項として認定される。齋藤 他 (2018) の基準でイベントを抽出した後、イベントを構成する語句を元のテキストにおける表示順で結合し、イベントのテキスト表現を得る。

### 3.2 スコープ

本研究では、単純な言語素性、具体的には述語の品詞と態 (voice) によって意志性が同定できないイベントを扱う。述語の品詞が例 (10) のように形容詞あるいは例 (11) のように判定詞であるイベントは状態を表し、非意志的であることが明らかであるため、本研究では扱わない。

(10) 空が綺麗だ. (NV)

(11) 彼は学生だ. (NV)

述語が例 (12) のように受動態または例 (13) のように可能態であるときも、イベントは主語の観点から見て非意志的であることが明らかであるため、本研究では扱わない。

(12) 先生に叱られる. (NV)

(13) 私は走れる. (NV)

加えて、モダリティを伴うイベントも本研究では扱わない。モダリティは、イベントに対する筆者の意見や態度を表す言語表現である。例 (14) は蓋然性を表すモダリティ「はずだ」を伴うイベントである。

(14) 彼は来るはずだ.

本研究の主眼はイベントそのものの意志性を認識することにあるため、こうしたイベントも本研究では扱わない。本研究では、蓋然性を含むモダリティ 10 種を考慮し、それらを伴うイベントを除いた。モダリティのリストと例を付録の表 9 に示す。

最後に、本研究では、主節に相当するイベントのみを採用し、従属節に相当するイベントは扱わない。例えば、以下の文からは「雨が強まったので」と「試合は中止になった」の 2 つのイベントが抽出されるが、前者 (従属節) は除外し、後者 (主節) のみを利用する。

(15) 雨が強まったので、試合は中止になった.

これは 5.2 節で説明するクラウドソーシングによる評価データ構築の際、イベントが接続表現で

終わっていること等を理由に不自然な言語表現であると判断されることを避けるためである<sup>4</sup>。従属節のイベントの意志性を分類する際は、節末の言語表現を主節らしく適当に整形すれば良い。

### 3.3 アノテーション

イベントには意志性と主語有生性の2つのラベルを付与する。

**意志性** 主語の表すエンティティがイベントに意志的に関与しているなら、「意志的 (volitional)」のラベルを付与する。そうでないなら、「非意志的 (non-volitional)」のラベルを付与する。

**主語有生性** 主語が表すエンティティに人間のような意志的な行為が可能であるなら、「有生名詞 (animate noun)」のラベルを付与する。そうでないなら、「無生名詞 (inanimate noun)」のラベルを付与する。主語有生性のラベルはそのイベントにおいて主語が有生名詞であるか無生名詞であるかを表すもので、名詞だけを取り出して、それが一般に有生名詞であるか有生名詞であるかを表すものではないことに注意されたい。

## 4 提案手法

本研究では、ある単独のイベント  $x$  を入力として、その意志性  $y_{vol}$  と主語有生性  $y_{ani}$  を予測する問題を考える。 $y_{vol}$  と  $y_{ani}$  はポジティブ (意志的/有生名詞) であるなら1, ネガティブ (非意志的/無生名詞) であるなら0である。

提案手法ではまず、ヒューリスティクスを用いて生コーパスから収集したイベントに意志性と主語有生性のラベルを付与する。そうして得たデータの上で、意志性と主語有生性の分類を同時学習する。その際、分類器が手がかり語だけに着目した分類に陥ることを抑制する正則化を導入する。意志性と主語有生性の分類の損失および正則化のペナルティを合わせたものを目的関数とし、それを最小化するように分類器とエンコーダを最適化する。

### 4.1 訓練データセットの構築

生コーパス中のイベントに対し、ヒューリスティクスを用いて意志性・主語有生性のラベルを付与し、意志性のラベル付きデータセット  $\mathcal{D}_{vol}^l$  とラベルなしデータセット  $\mathcal{D}_{vol}^u$ 、主語有生性のラベル付きデータセット  $\mathcal{D}_{ani}^l$  とラベルなしデータセット  $\mathcal{D}_{ani}^u$  の4種類のデータセットを得る。

まず、3.2節にて説明した条件を満たすイベントを生コーパスから抽出する。ここでは、既存の構文解析器および品詞タグ付けモデルを利用する。抽出したイベントそれぞれについて、ヒューリスティクスを用いて意志性および主語有生性のラベルを付与する。ラベル付与の成否

<sup>4</sup> クラウドワーカに対して不自然な言語表現と判断する基準を教示し、従属節のイベントもアノテーションの対象とすることも考えられるが、意志性・主語有生性の判断という本質的な作業に注力させるため、本研究では主節のイベントのみを採用することにした。

にしたがって、イベントを各データセットに振り分ける。

意志性のラベルを付与するため、意志的副詞・非意志的副詞の小規模な辞書を用意する。辞書中の副詞がイベントの述語を修飾していれば対応するラベルをそのイベントに付与し、イベントを  $\mathcal{D}_{\text{vol}}^l$  に追加する。ラベルを付与できなかったイベントは  $\mathcal{D}_{\text{vol}}^u$  に追加する。

主語有生性のラベルを付与するため、まず既存のオントロジーから得た有生名詞・無生名詞のリストを得る。意味役割解析器の出力を用いてイベントの主語の名詞句を同定し、主語の名詞句が見つかったら、有生名詞・無生名詞のリストを参照する。そこに主語の名詞句が含まれていたら、対応するラベルを付与した後、イベントを  $\mathcal{D}_{\text{ani}}^l$  に追加する。ラベルを付与できなかったイベントは  $\mathcal{D}_{\text{ani}}^u$  に追加する。

## 4.2 モデル

モデルは3つの要素から構成される。1つ目はエンコーダ  $E$  である。エンコーダはイベント  $x$  をベクトル表現に変換する。エンコーダは事前学習済みの汎用言語モデル BERT (Devlin et al. 2019) で初期化する。2つ目は意志性分類器  $C_{\text{vol}}$  である。意志性分類器は、エンコーダが出力するイベントのベクトル表現を入力として、そのイベントが意志的である確率を予測する。3つ目は主語有生性分類器  $C_{\text{ani}}$  である。主語有生性分類器は、エンコーダが出力するイベントのベクトル表現を入力として、そのイベントの主語が有生名詞である確率を予測する。意志性分類器と主語有生性分類器のパラメータはランダムに初期化する。エンコーダと各分類器のパラメータは後述の目的関数を最小化するように最適化される。

## 4.3 正則化付き分類学習

モデルは意志性の分類と主語有生性の分類を同時学習する。その際、手がかり語だけに着目して分類することを抑制する正則化を導入する。意志性と主語有生性はどちらも統一的な方法で学習するため、意志性・主語有生性の区別を取り払い、ラベル付きデータセットを  $\mathcal{D}^l$ 、ラベルなしデータセットを  $\mathcal{D}^u$ 、ラベルを  $y$ 、分類器を  $C$  とする表記を導入する。以降、上記の表記に関しては、意志性の学習では「vol」の下付き文字、主語有生性の学習では「ani」の下付き文字を伴うものとして読み替えるものとする。

### 4.3.1 分類 (CLS; classification)

主たる目的関数はラベル付きデータセット  $\mathcal{D}^l$  の上で分類を学習するものである。形式的には以下のように書ける。

$$\mathcal{L}_{\text{CLS}} = \mathbb{E}_{(x,y) \sim \mathcal{D}^l} \text{BCE}(y, C(E(x))) \quad (1)$$

BCE はバイナリ交差エントロピーである。

### 4.3.2 正則化 (REG; regularization)

手がかり語に頼らない分類を学習するための正則化として3つの手法を検討する. 実験では開発データに対する性能をもとに最適な正則化手法を選択する.

**Word Removal (WR)** WR はバイアス削減の最も単純な手法である. WR は訓練データ中のテキストから特定の語句を削除することで, 分類器がその語句に頼って予測を学習することを抑制する (Kennedy et al. 2020). 本研究ではラベル付けに利用した手がかり語に対して WR を適用し, 手がかり語が予測において注目されることを抑制する. 形式的には, 以下の目的関数の最小化を解く.

$$\mathcal{L}_{\text{WR}} = \mathbb{E}_{(x,y,w) \sim \mathcal{D}^l} \text{BCE}(y, C(E(x \setminus w))) \quad (2)$$

ここで  $w$  はイベント  $x$  中に含まれる手がかり語,  $x \setminus w$  は  $x$  から  $w$  を除いたテキストである.

**Sampling and Occlusion (SOC)** SOC (Kennedy et al. 2020) はバイアス削減の手法であり, ある語句の予測における寄与度を計算し, それを最小化することで, その単語が予測において注目されることを抑制する. SOC ではまず, 予測における寄与度を計算する語句の周りの文脈を学習済み言語モデルを用いてサンプリングする. そして, サンプリングした各文脈においてその語句をマスクトークンに置き換えたときと置き換えないうちの予測の差を計算し, その平均をその語句の予測における寄与度とする. SOC によって計算される寄与度は微分可能であるため, 通常の勾配法によって, 特定の語句が予測において注目されることを抑制できる. 本研究ではラベル付けに利用した手がかり語に対して SOC を適用する. 形式的には以下の目的関数の最小化を解く.

$$\mathcal{L}_{\text{SOC}} = \mathbb{E}_{(x,w) \sim \mathcal{D}^l} [\phi(x, w)]^2 \quad (3)$$

$$\phi(x, w) = \frac{1}{|S|} \sum_{x' \in S} [C(E(x')) - C(E(x' \setminus w))]^2 \quad (4)$$

ここで  $w$  はイベント  $x$  中に含まれる手がかり語,  $S$  は  $x$  を初期値として事前学習済みの言語モデルを用いて  $w$  の文脈の単語をサンプリングして得たイベントの集合,  $x' \setminus w$  は  $x'$  に現れる  $w$  を padding トークンで置き換えたテキストである.

**Adversarial Domain Adaptation (ADA)** ADA は教師なしドメイン適応の一手法である. 本研究では, ラベル付きデータ  $\mathcal{D}^l$  をソースドメインのデータ, ラベルなしデータ  $\mathcal{D}^u$  をターゲットドメインのデータとみなして ADA を適用する. 学習時には discriminator と呼ばれるニューラルネットワーク  $D$  を追加で学習する. discriminator はエンコーダが出力するイベントのベクトル表現を入力として, それがソースドメイン由来であるなら 1, ターゲットドメイン由来であるなら 0 を出力するように学習する. エンコーダは discriminator がその分類に失敗するよ

うに、すなわちエンコーダがソースドメイン由来の特徴に対して 0、ターゲットドメイン由来の特徴に対して 1 を出力するように学習する。形式的には以下の目的関数の最小化を解く。

$$\mathcal{L}_{\text{ADA}} = \mathbb{E}_{x \sim \mathcal{D}^t} \text{BCE}(0, D(E(x))) + \mathbb{E}_{x \sim \mathcal{D}^u} \text{BCE}(1, D(E(x)))$$

discriminator はこの目的関数の最大化を学習する。

### 4.3.3 一貫性 (CON; consistency)

手がかり語だけに着目した分類を抑制する正則化に加えて、意志性分類と主語有生性分類の一貫性を学習することを考える。主語が有生名詞であることはイベントが意志的であることの必要条件である。したがって、主語が無生名詞であるという予測とイベントが意志的であるという予測は両立しない。この関係を以下の目的関数を最小化することで学習する。

$$\mathcal{L}_{\text{CON}} = \mathbb{E}_{x \sim \mathcal{D}_{\text{vol}}^u + \mathcal{D}_{\text{ani}}^u} \max(0, C_{\text{vol}}(E(x)) - C_{\text{ani}}(E(x))) \quad (5)$$

実験では、意志性・主語有生性の一方の分類と一貫性の学習だけで他方の分類を解く設定も検証する。このとき、一貫性の学習は分類を学習している側の性能に影響しないことに注意されたい。一方の分類が学習されていない状況では、分類を学習している側が一貫性を学習することで得られる情報はなからである。

### 4.3.4 目的関数

上記の目的関数を重み付きで足し合わせ、最終的な目的関数を得る。

$$\mathcal{L} = \mathcal{L}_{\text{CLS}} + \alpha \mathcal{L}_{\text{REG}} + \beta \mathcal{L}_{\text{CON}} \quad (6)$$

$\alpha$  と  $\beta$  はハイパーパラメータとして設定される正則化の重み、 $\mathcal{L}_{\text{REG}}$  は  $\mathcal{L}_{\text{WR}}$ 、 $\mathcal{L}_{\text{SOC}}$ 、 $\mathcal{L}_{\text{ADA}}$  のいずれかである<sup>5</sup>。

## 5 実験

提案手法の有効性を確認するため、日本語と英語で実験を行った。

### 5.1 訓練データセット

4.1 節に記載の手続きにしたがって訓練データセットを構築した。

<sup>5</sup> これらの正則化手法を組み合わせることは技術的には可能であるが、計算資源の制約から、これらの組み合わせについては検証しない。

### 5.1.1 日本語

生コーパスとして CC-100 (Conneau et al. 2020; Wenzek et al. 2020) 中の 3 千万文書を利用した。イベントの抽出には構文解析器 KNP (Kawahara and Kurohashi 2006) を用いた。イベント抽出では、まず極端に出現頻度の低い  $D_{vol}^l$  のイベントを生コーパス全体から取り出した<sup>6</sup>。この際、副詞の文字列を含まない文を構文解析の対象から除外し、計算コストを削減した。その後、 $D_{vol}^u$ ,  $D_{ani}^l$ ,  $D_{ani}^u$  のイベントを  $D_{vol}^l$  と同じ件数だけ生コーパスのサンプルから取り出した。イベントの頻度の情報を保存するため、重複するイベントが得られても削除しなかった。

意志性のラベル付与のため、意志的副詞・非意志的副詞をそれぞれ 15 件ずつ整備した。表 1 に頻度上位 5 件の副詞、付録 B にすべての副詞を示す。主語有生性のラベルを付与するため、JumanDic<sup>7</sup> を知識源として用いた<sup>8</sup>。JumanDic に登録されている語にはカテゴリ情報が付与されており、カテゴリが人または組織・団体のものを有生名詞、それ以外を無生名詞とした。JumanDic に登録されている語の数はおよそ 3 万であった。主語は KNP の格解析によって同定した。具体的には、ガ 2 格の句があればそれを、ガ 2 格の句がなくガ格の句があればそれを主語として認定し、どちらもなければ主語有生性のラベルを付与しなかった。

### 5.1.2 英語

生コーパスとして CC-100 に収録されている 3 千万文書を利用した。イベントの抽出には spacy<sup>9</sup> を用い、日本語と同様の手続きで  $D_{vol}^l$ ,  $D_{vol}^u$ ,  $D_{ani}^l$ ,  $D_{ani}^u$  のイベントを抽出した。

意志性のラベルを付与するため、意志的副詞・非意志的副詞をそれぞれ 10 件ずつ整備した。表 1 に頻度上位 5 件の副詞、付録 B にすべての副詞を示す。主語有生性のラベルを付与するた

日本語		英語	
意志的	非意志的	意志的	非意志的
あえて (5,293)	思わず (18,115)	carefully (13,594)	unfortunately (13,070)
急いで (4,187)	つい (15,897)	thoroughly (12,468)	automatically (12,824)
じっくり (4,017)	自動的に (14,212)	actively (10,379)	accidentally (5,272)
慎重に (3,743)	ふと (12,050)	deliberately (3,366)	unexpectedly (3,106)
わざわざ (3,262)	つつい (10,054)	intentionally (2,713)	luckily (1,894)

表 1 意志的副詞・非意志的副詞それぞれの高頻度 5 件。括弧内は頻度。

<sup>6</sup> 意志的副詞・非意志的副詞を伴うイベントは日英ともに約 5,000 件に 1 件の割合で得られた。意志的副詞・非意志的副詞のリストは付録 B を参照されたい。

<sup>7</sup> <https://github.com/ku-nlp/JumanDIC>

<sup>8</sup> 実装としては、KNP が付与する「SM-主体」の素性を利用した。これは JumanDic のカテゴリ情報にもとづき付与される素性である。

<sup>9</sup> <https://spacy.io>

め、ConceptNet (Speer et al. 2017) から有生名詞・無生名詞のリストを得た。具体的には、有生名詞として「person」と「organization」の下位語、無生名詞として「object」、「item」、「thing」、「artifact」、「location」の下位語を用いた。結果として、2,604件の有生名詞、430件の無生名詞が得られた。

## 5.2 評価データセット

$D_{vol}^l$ ,  $D_{vol}^u$ ,  $D_{ani}^l$ ,  $D_{ani}^u$  における分類器の性能を調べるため、それぞれについて評価用に人手で正解ラベルを付与した。なお、本研究の主眼である意志性分類に関しては、手がかり語の副詞を伴うイベントは出現する割合が低く、実テキストにおける性能は  $D_{vol}^u$  に対する性能で近似できることに注意されたい。

まず、各データセットから1,200件ずつユニークなイベントを抽出した。それらにクラウドソーシングで正解ラベルを付与した。意志性に関して、クラウドワーカーは各イベントに対して以下のラベルの中から1つを選んで付与した。

- 主語がイベントに意志的に関与している。
- 主語がイベントに意志的に関与していない。
- どちらとも言えない。
- 理解不能。

主語有生性に関して、クラウドワーカーは各イベントに対して以下のラベルの中から1つを選んで付与した。主語有生性のラベルは、3.3節で説明した通り、そのイベントにおいて主語が有生名詞であるか無生名詞であるかを表すものであり、主語だけを取り出してそれが一般に有生名詞であるか無生名詞であるかを表すものではないことに注意されたい。

- 主語が人または組織を表す。
- 主語が人も組織を表さない。
- どちらとも言えない。
- 理解不能。

各イベントに5人のワーカーを割り当て、ラベル付けさせた。クラウドワーカーは一度の作業で10件のイベントにラベルを付与した。日本語のクラウドソーシングプラットフォームとして、Yahoo!クラウドソーシング<sup>10</sup>を利用した。品質管理のため、Yahoo!クラウドソーシングのチェック設問機能を利用した。この機能を用いて、事前にわれわれが正解を用意している簡単な設問（チェック設問）を1問追加した。チェック設問に誤答したワーカーは不適格として、報酬を支払わず、回答も破棄した。かかった費用は24,000円であった。英語のクラウドソーシングプラットフォームとして、Amazon Mechanical Turk (MTurk) を利用した。品質管理のため、

<sup>10</sup> <https://crowdsourcing.yahoo.co.jp/>

一般に取られている慣行 (Berinsky et al. 2012) にならった. 具体的には, ワーカをタスクの合格率が 95%以上, アメリカ在住, 1,000 タスク以上を完了している者に限定した. かかった費用は 288USD であった.

表 2 にアノテータ間の一致率および Fleiss' kappa, 表 3 に各ラベルごとの最大得票数を得たイベントの件数を示す. 日本語に関しては妥当な相関を確認できたが, 英語に関してはわずかな相関しか確認できなかった. アノテーションの平均的な一致率が高かった日本語のイベントに関して, アノテータ間で票が割れたものの一部を示す.

- (16) また, 持久走も行いました。(V) (V: 3 名, NV: 2 名)  
 (17) その他にかあれば追記で書いていきます。(V) (V: 3 名, NV: 2 名)  
 (18) ちょっと緊張してしまう。(NV) (V: 2 名, NV: 3 名)  
 (19) 実は, 驚くほど様々な効果・効能があります。(NV) (V: 2 名, NV: 3 名)

票の割れたイベントの多くは判断に困る例とは考えにくく, 票が割れた原因は不真面目なクラウドワーカーがラベル付けを担当したことにあると考えるのが妥当であった. 英語ではこの傾向がさらに顕著であった.

	$D_{vol}^l$	$D_{vol}^u$	$D_{ani}^l$	$D_{ani}^u$
日本語	0.782/0.351	0.769/0.332	0.772/0.357	0.743/0.239
英語	0.624/0.027	0.628/0.029	0.668/0.057	0.671/0.059

表 2 各データセットにおけるアノテータ間の一致率 (左) と Fleiss' kappa (右). 一致率は各イベントに関して最大の得票数を得た解答の割合の平均である.

ラベル		日本語		英語	
		$D^l$	$D^u$	$D^l$	$D^u$
意志性	主語がイベントに意志的に関与している.	488	628	599	583
	主語がイベントに意志的に関与していない.	700	553	594	614
	どちらも言えない.	0	1	3	2
	理解不能.	12	18	4	1
主語有生性	主語が人または組織である.	566	802	819	507
	主語が人でも組織でもない.	595	361	376	689
	どちらも言えない.	9	12	4	4
	理解不能.	30	25	1	0

表 3 各ラベルごとの最大得票数を得たイベントの件数.

	分割	ラベル	日本語	英語
$D_{\text{vol}}^l$	学習	意志的	31,812	47,926
		非意志的	81,002	40,564
	開発	意志的	149	67
		非意志的	233	92
	テスト	意志的	149	68
		非意志的	233	93
$D_{\text{vol}}^u$	学習	ラベルなし	112,814+	88,490+
	開発	意志的	206	62
		非意志的	164	104
	テスト	意志的	206	63
		非意志的	164	104
	$D_{\text{ani}}^l$	学習	有生名詞	29,344+
無生名詞			83,470+	17,233+
開発		有生名詞	175	170
		無生名詞	199	59
テスト		有生名詞	176	170
		無生名詞	200	60
$D_{\text{ani}}^u$	学習	ラベルなし	112,814+	88,490+
	開発	有生名詞	246	78
		無生名詞	93	152
	テスト	有生名詞	246	78
		無生名詞	93	153

表 4 データセットの統計. + がついているものは, 最も件数が少なかった  $D_{\text{vol}}^l$  のイベント数に合わせて, 同量のイベントを抽出したことを表す.

この観察から, 信頼できるクラウドワーカーによってラベル付けが行われたことが期待される, アノテータ間の一致率が80%以上 (5名中4名の回答が一致) のイベントを抽出し, 評価データとして利用した. この条件のもとで採用されるイベントと除かれるイベントの間に定性的な違いは確認できず, この条件で除かれたイベントに対しても, 本研究で構築した評価データにおける性能と同等程度の性能で分類が可能と期待する. 評価データのうち, 半分は開発データ, 残りの半分はテストデータとした. 分割はラベルごとにランダムに行った. 表 4 にデータセットの統計を示す.

### 5.3 実装の詳細

エンコーダには BERT<sub>BASE</sub> (Devlin et al. 2019) を用いた. 日本語の実験では NICT が提供

している事前学習済みモデル<sup>11</sup>を用いた。英語の実験ではオリジナルの BERT<sub>BASE</sub> の cased モデルを用いた。BERT の事前学習時の入力形式にならない、イベントのトークン列の先頭に CLS トークン、末尾に SEP トークンを追加して入力テキストを構成した。CLS トークンの最終層のベクトル表現をイベントのベクトル表現として利用した。意志性・主語有生性の分類器は ReLU 関数を非線層とする 3 層パーセプトロンとした。分類器の出力はシグモイド関数を用いて 0 から 1 の範囲の実数値に変換した。SOC におけるサンプル数  $|S|$  は 3 とした。文脈のサンプリングは、エンコーダに用いた BERT を使ってギブスサンプリングを行うことで実現した (Wang and Cho 2019)。ADA では discriminator を ReLU 関数を非線層とする 3 層パーセプトロンとして構成・学習した。 $\alpha$  と  $\beta$  の値は  $\{0.0, 0.01, 0.1, 1.0\}$  の中から選択した。モデルはバッチサイズを 256 とし 3 エポック学習した<sup>12</sup>。モデルのパラメータは Adam (Kingma and Ba 2015) で最適化した。学習率は  $3e-5$  とし<sup>13</sup>、10%の学習が完了した時点で学習率がピークになるように線形のスケジューリングを行った。モデル選択では、100 ステップごとに  $D_{vol}^u$  の開発データにおける精度を計算し、最高スコアを達成した時点のパラメータを採用した。 $D_{vol}^u$  の性能に基づいてモデル選択を行うのは、意志的・非意志的行為を表す副詞を伴うイベントの出現頻度が低く、実テキストにおける意志性分類の性能が  $D_{vol}^u$  に対する性能で近似できるからである。評価指標には AUC を用いた。モデルは異なるランダムシードで 3 回学習し、その平均・分散を計算した。モデルの実装には PyTorch を用いた。

## 5.4 実験結果

表 5 に日本語における実験結果を示す。主たる関心である  $D_{vol}^u$  を含む、全てのデータセットで正則化と同時学習を併せて利用したモデルが最高精度を達成した。 $D_{vol}^u$  に対する最高精度は、意志性を CLS+REG (SOC)、主語有生性を CLS+REG (WR) で学習したモデルが達成した (AUC: 96.7, 正解率: 92.2, 精度: 94.9, 再現率: 90.8, F 値: 92.8)。これはベースラインである意志性を CLS で学習したモデルの精度 (AUC: 89.5, 正解率: 83.5, 精度: 91.0, 再現率: 78.2, F 値: 84.1) を大きく上回っている。主語有生性を WR で学習するのが有効なことには主語の省略が関係していると考えられる。主語有生性のラベルを付与する際、主語の同定を格解析によって行ったため、主語が省略されているイベントには主語有生性のラベルが付与されず、それらはラベルなしデータセット  $D_{ani}^u$  に追加された。WR によって主語をテキストから除外した上でその有生性を予測することを学習したことで、主語が省略されているイベントに汎化する推論が学習され、それが精度の向上に繋がったのだと考える。一貫性の学習は有効に機能しなかった。

<sup>11</sup> <https://alaginrc.nict.go.jp/nict-bert/index.html>

<sup>12</sup> エポックには小規模の予備実験の結果をもとに十分大きな値を設定した。

<sup>13</sup>  $D_{vol}^u$  を正則化なし・主語有生性の同時学習なしで学習する設定において、 $\{5e-5, 3e-5, 2e-5\}$  の範囲を探索し、分類精度に影響しないことを確認した。ハイパパラメータの探索範囲を減らすため、本実験では学習率は  $3e-5$  で固定とした。

意志性	主語有生性		$D_{vol}^l$ (開発)	$D_{vol}^l$ (テスト)	$D_{vol}^u$ (開発)	$D_{vol}^u$ (テスト)
CLS	-		94.5 ± 0.5	91.7 ± 1.0	87.3 ± 0.8	89.5 ± 1.1
CLS	CLS		92.8 ± 1.0	91.8 ± 1.3	87.8 ± 0.4	90.6 ± 0.1
CLS	CLS	+CON	95.0 ± 0.4	92.1 ± 0.5	84.9 ± 5.4	89.6 ± 1.9
CLS	CLS+REG (WR)		95.3 ± 0.3	93.9 ± 0.6	90.8 ± 0.4	92.5 ± 1.2
CLS	CLS+REG (SOC)	+CON	93.6 ± 0.9	91.5 ± 1.0	88.5 ± 1.2	93.5 ± 0.6
CLS+REG (SOC)	-	+CON	<b>97.1 ± 0.0</b>	94.4 ± 0.6	90.4 ± 0.2	92.9 ± 0.1
CLS+REG (SOC)	CLS		96.8 ± 0.2	<b>94.6 ± 0.4</b>	91.8 ± 0.3	94.0 ± 0.5
CLS+REG (SOC)	CLS	+CON	96.8 ± 0.1	<b>94.6 ± 0.4</b>	92.1 ± 0.4	94.7 ± 0.5
CLS+REG (SOC)	CLS+REG (WR)		97.0 ± 0.1	94.3 ± 0.1	<b>92.6 ± 0.2</b>	<b>96.7 ± 0.7</b>
CLS+REG (SOC)	CLS+REG (SOC)	+CON	96.9 ± 0.1	94.4 ± 0.4	92.1 ± 0.2	96.0 ± 0.5
-	CLS	+CON	70.2 ± 2.6	65.3 ± 2.6	72.3 ± 0.9	77.3 ± 0.9
-	CLS+REG (WR)	+CON	76.1 ± 0.9	73.5 ± 1.4	81.4 ± 0.2	85.1 ± 1.0

(a) 意志性分類の結果.

主語有生性	意志性		$D_{ani}^l$ (開発)	$D_{ani}^l$ (テスト)	$D_{ani}^u$ (開発)	$D_{ani}^u$ (テスト)
CLS	-		89.9 ± 0.8	92.0 ± 0.7	77.4 ± 0.5	81.4 ± 0.8
CLS	CLS		90.9 ± 0.6	91.3 ± 0.3	77.7 ± 2.2	81.7 ± 1.7
CLS	CLS	+CON	84.3 ± 1.7	87.7 ± 3.3	79.3 ± 1.2	83.0 ± 1.2
CLS	CLS+REG (SOC)		89.0 ± 1.7	92.3 ± 1.8	81.0 ± 0.8	86.1 ± 0.8
CLS	CLS+REG (SOC)	+CON	87.7 ± 0.8	90.0 ± 1.0	79.3 ± 1.0	84.5 ± 0.4
CLS+REG (WR)	-	+CON	91.3 ± 0.6	94.3 ± 0.6	85.9 ± 0.3	86.4 ± 0.2
CLS+REG (WR)	CLS		92.0 ± 0.1	94.0 ± 0.0	86.5 ± 0.6	86.4 ± 0.4
CLS+REG (SOC)	CLS	+CON	87.1 ± 0.8	89.6 ± 1.5	79.7 ± 1.2	83.7 ± 0.8
CLS+REG (WR)	CLS+REG (SOC)		<b>92.2 ± 0.4</b>	<b>95.3 ± 1.0</b>	<b>87.8 ± 1.5</b>	<b>89.9 ± 0.6</b>
CLS+REG (SOC)	CLS+REG (SOC)	+CON	87.7 ± 0.5	90.0 ± 0.4	78.9 ± 0.6	84.6 ± 0.8
-	CLS	+CON	73.7 ± 1.2	72.6 ± 2.4	66.0 ± 3.3	70.7 ± 2.7
-	CLS+REG (SOC)	+CON	68.9 ± 2.0	67.3 ± 2.2	63.8 ± 1.1	67.1 ± 0.7

(b) 主語有生性分類の結果.

表 5 日本語における意志性・主語有生性分類の結果. 評価指標は AUC である. 数値は異なるランダムシードを用いて 3 回実験を行った際の平均と分散である. 太字のスコアはすべてのモデルのうち最も平均スコアが高いことを表す. 正則化 (REG) には開発データに対する性能をもとに選択された手法を併記している. ハイフン (-) はその分類をラベル付きデータを用いて学習していないことを示す.

表 6 に英語における実験結果を示す. 英語においても全てのデータセットで正則化と同時学習を併せて利用したモデルが最高精度を達成した. 主たる関心である  $D_{vol}^u$  に対する最高精度は, 意志性を CLS+REG (SOC), 主語有生性を CLS で学習し, 加えて一貫性を学習したモデルが達成した (AUC: 75.8, 正解率: 74.3, 精度: 66.1 再現率: 65.1, F 値: 65.6). これはベースライン

意志性	主語有生性		$D_{vol}^l$ (開発)	$D_{vol}^l$ (テスト)	$D_{vol}^u$ (開発)	$D_{vol}^u$ (テスト)
CLS	-		68.3 ± 2.2	73.7 ± 1.8	63.2 ± 3.4	66.2 ± 0.9
CLS	CLS		68.4 ± 2.0	<b>74.4 ± 0.5</b>	68.9 ± 0.9	70.7 ± 3.1
CLS	CLS	+CON	69.3 ± 1.3	74.0 ± 1.6	69.4 ± 1.3	69.8 ± 3.0
CLS	CLS+REG (WR)		69.2 ± 1.3	72.4 ± 4.3	69.8 ± 0.1	69.7 ± 0.2
CLS	CLS+REG (WR)	+CON	<b>71.0 ± 0.2</b>	71.9 ± 2.6	70.2 ± 0.3	70.8 ± 0.6
CLS+REG (SOC)	-	+CON	70.0 ± 0.5	73.3 ± 0.4	71.6 ± 0.2	72.2 ± 1.3
CLS+REG (SOC)	CLS		69.3 ± 0.6	73.9 ± 0.5	73.8 ± 0.5	75.4 ± 0.8
CLS+REG (SOC)	CLS	+CON	70.6 ± 1.6	73.7 ± 0.3	75.4 ± 0.6	<b>75.8 ± 0.5</b>
CLS+REG (SOC)	CLS+REG (ADA)		69.3 ± 0.6	73.1 ± 0.4	74.4 ± 0.6	74.0 ± 1.8
CLS+REG (SOC)	CLS+REG (WR)	+CON	70.2 ± 0.2	73.6 ± 0.2	<b>75.6 ± 0.4</b>	75.7 ± 0.2
-	CLS	+CON	59.9 ± 0.5	64.0 ± 0.9	69.0 ± 0.6	69.3 ± 0.7
-	CLS+REG (WR)	+CON	63.8 ± 0.6	65.1 ± 0.6	70.1 ± 0.2	70.7 ± 0.2

(a) 意志性分類の結果.

主語有生性	意志性		$D_{ani}^l$ (開発)	$D_{ani}^l$ (テスト)	$D_{ani}^u$ (開発)	$D_{ani}^u$ (テスト)
CLS	-		90.0 ± 0.5	83.5 ± 1.0	81.2 ± 2.1	82.4 ± 1.3
CLS	CLS		91.0 ± 1.1	82.3 ± 2.1	78.3 ± 3.5	81.5 ± 1.0
CLS	CLS	+CON	91.6 ± 0.2	84.3 ± 1.6	78.6 ± 3.5	81.9 ± 1.3
CLS	CLS+REG (SOC)		90.4 ± 0.7	83.2 ± 1.9	82.3 ± 0.8	83.3 ± 0.2
CLS	CLS+REG (SOC)	+CON	90.9 ± 0.2	<b>85.4 ± 0.3</b>	<b>83.0 ± 0.4</b>	<b>85.1 ± 0.3</b>
CLS+REG (WR)	-	+CON	89.6 ± 0.9	84.2 ± 0.4	82.5 ± 0.5	81.7 ± 0.4
CLS+REG (WR)	CLS		91.3 ± 0.4	83.6 ± 0.7	82.3 ± 0.8	81.6 ± 0.1
CLS+REG (WR)	CLS	+CON	<b>91.7 ± 0.4</b>	84.1 ± 0.8	82.3 ± 0.9	81.9 ± 0.6
CLS+REG (ADA)	CLS+REG (SOC)		90.8 ± 1.2	83.9 ± 0.4	80.6 ± 1.3	83.1 ± 0.2
CLS+REG (WR)	CLS+REG (SOC)	+CON	91.1 ± 0.7	84.6 ± 0.9	80.9 ± 0.5	84.7 ± 0.4
-	CLS	+CON	69.3 ± 0.3	67.2 ± 3.6	59.7 ± 4.0	66.0 ± 1.9
-	CLS+REG (SOC)	+CON	69.8 ± 0.8	66.2 ± 1.8	66.9 ± 2.6	73.0 ± 1.3

(b) 主語有生性分類の結果.

表 6 英語における意志性・主語有生性分類の結果. 評価指標は AUC である. 数値は異なるランダムシードを用いて 3 回実験を行った際の平均と分散である. 太字のスコアはすべてのモデルのうち最も平均スコアが高いことを表す. 正則化 (REG) には開発データに対する性能をもとに選択された手法を併記している. ハイフン (-) はその分類をラベル付きデータを用いて学習していないことを示す.

である意志性を CLS で学習したモデルの精度 (AUC: 66.2, 正解率: 69.5, 精度: 65.8, 再現率: 39.7, F 値: 49.5) を大きく上回っている. 日本語のスコアと比較して, 英語のスコアは全体的に低い傾向が見て取れた. この主たる原因は評価データセットの品質にあった. 5.2 節で述べたように, 英語の評価データのラベル付けにあたったクラウドワーカーには不真面目な者が多

かった。英語の評価データの中にそうしたクラウドワーカーの票が偶然誤ったラベルに集まったことで採用されたと考えられるイベントが一定数含まれており、それによって分類器の性能が過小評価されていた。日本語に関しては、そうした例はほとんど見られなかった。日本語のラベル付けを行ったクラウドワーカーには不真面目な者が少なく、複数の不真面目なクラウドワーカーが特定のイベントのラベル付けに割り当てられかつ誤ったラベルに票が集まるケースが少なかったからだと考えられる。高品質な英語データセットの構築は今後の課題である。

## 6 分析

### 6.1 定性分析

提案手法で学習したモデルが1節で言及した文脈依存性を捉えられるか定性的に分析した。分析には  $D_{\text{vol}}^u$  において最高精度を達成したモデルを用いた。ランダムシードを変えて学習したモデルが3つ存在するが、 $D_{\text{vol}}^u$  の開発データでの性能に関して2番目のモデルを用いた。

日本語に関して、 $D_{\text{vol}}^u$  における最良のモデルは意志性を CLS+REG (SOC)、主語有生性を CLS+REG (WR) で学習したものであったため、これを分析に用いた。まず、例 (1-a) と例 (1-b) を用いてモデルが述語の項の違いを認識して意志性を分類できるか調べ、これらが正しく分類できることを確認した。次に、例 (2-a) と例 (2-b) を用いてモデルが副詞の意味を捉えて意志性を分類できるか調べ、こちらについても正しく分類できることを確認した。最後に、例 (3-a) と例 (3-b) を用いて主語有生性を文脈を考慮して予測できるか調べ、こちらについてもやはり正しく分類できることを確認した。分析に用いた例のように、イベントの一部が共通で、残りの部分の違いによって意志性・主語有生性が変化する例を集めたベンチマークを構築すれば、こうした文脈理解がどの程度可能なのか定量的に調べることができる。これは今後の課題としたい。

一方で、モデルが意志性の認識に失敗する例もあった。最も顕著だったのは述語が「いる」のイベントであった。「いる」は状態動詞であり、基本的に状態を表すが、主語が有生名詞である場合には意志的な動作を表すことがある。この認識誤りはイベントの主語が省略されている場合に特に顕著であり、省略された主語の有生性を予測する難しさが「いる」の意味理解をさらに難しくしていると推測される。

この問題に対する解決策として、学習時に前後のイベント、すなわちより広い文脈を考慮することが考えられる。イベントの意味が異なれば、その前後に現れるイベントの分布も異なることが期待できる。そうした分布の異なりを学習に利用することで、表層的に類似しているイベント同士の意味の異なりをより精緻に捉えられるモデルが構築できると期待できる。

英語に関して、 $D_{\text{vol}}^u$  における最良のモデルは意志性を CLS+REG (SOC)、主語有生性を CLS で学習し、加えて予測の一貫性を考慮したものであり、これを用いて定性分析を行った。結果、英語についても、モデルが文脈を考慮した分類を行えていることを確認した。たとえば、以下

の例は述語はどちらも同じ「made」であるが、項の違いを認識して、前者を意志的、後者を非意志的と正しく分類できた。

- (20) a. I made pancakes. (V)  
b. I made a mistake. (NV)

以下の例もまた「for him」という副詞句の有無によるイベントの意味の違いを認識し、前者を非意志的、後者を意志的と正しく分類できた。

- (21) a. I tumbled. (NV)  
b. I tumbled for him. (V)

## 6.2 意志性のラベル付けに用いる副詞の種類数が性能に与える影響

提案手法は、生コーパス中のイベントに意志性のラベルを付与するために意志的副詞・非意志的副詞を手手で準備する必要がある。分析として、日本語のデータセットを用いて、副詞の種類数が主たる関心である  $D_{vol}^u$  の分類精度に与える影響を調査した。具体的には、意志的・非意志的な行為を表す副詞をそれぞれ頻度順に並べたときの上位1件、2件、4件、8件、15件を用いる場合の精度を比較した。副詞の種類数が増えると同じ大きさの生コーパスから得られるラベル付きデータの数も増えるが、ここでは副詞の種類数が性能に与える影響を調べるため、頻度上位1件の副詞を用いたときに得られるラベル付きデータの件数 23,408 件（「あえて」にマッチしたイベント 5,293 件と「思わず」にマッチしたイベント 18,115 件の合計）に合わせてサンプリングした。

モデルとして、意志性のラベル付きデータのみを正則化なしで学習する素朴な設定（ベースライン）と、 $D_{vol}^u$  における最高精度を達成した、意志性を CLS+REG (SOC)、主語有生性を CLS+REG (WR) で学習する設定（提案手法）の2つを用いた。学習時の設定は5.3節に記載のものと同じである。

表7に結果を示す。両方のモデルにおいて、ラベル付けに利用する副詞の種類を増やすことで  $D_{vol}^u$  の性能が向上することが見て取れた。これは、副詞の種類を増やすことでラベル付きデータの多様性が増し、多くのイベントに汎化する推論が学習できたからだと考えられる。また、驚くべきことに、提案手法は選定した副詞の中で最も高頻度であった「あえて」と「思わず」だけをラベル付与に用いた設定でも 87.4 ポイントの AUC という高い分類精度を達成した。これは、「あえて」と「思わず」が多様な文脈で現れる副詞であり、これらの副詞を含むイベントからだけでも、意志性分類がかなり学習できることを示唆している。

	上位 1 件	上位 2 件	上位 4 件	上位 8 件	上位 15 件	上位 15 件*
ベースライン	74.3 ± 3.7	83.9 ± 0.5	85.9 ± 5.9	87.3 ± 2.0	88.0 ± 0.8	89.5 ± 1.1
提案手法	87.4 ± 0.7	92.3 ± 0.2	94.2 ± 0.6	95.1 ± 0.4	95.4 ± 0.7	96.7 ± 0.4

表 7 日本語の  $D_{vol}^u$  のテストデータにおける意志性のラベル付けに用いる副詞の種類数による分類性能の変化. 評価指標は AUC であり, 異なるランダムシードを用いて 3 回実行した際の平均と分散を記載している. 最も右の結果は, 表 5 からの再掲であり, 頻度上位 1 件の副詞を用いたときに得られるラベル付きデータの件数に合わせてサンプリングを行わず, 表 4 に記載の数のデータを用いたときの結果である.

	ラベル	日本語	英語
$D_{vol}^l$	意志的	88%	94%
	非意志的	92%	80%
$D_{ani}^l$	有生名詞	81%	96%
	無生名詞	72%	76%

表 8 正しいラベルが付与されていたイベントの割合.

### 6.3 ヒューリスティクスの精度

ラベル付きデータにはヒューリスティクスを用いてラベルを付与しているため, 誤ったラベルが付与されるケースもある. しかし, あまりにラベル付けが誤っていると, そこから期待する分類結果が得られるような分類器を学習することは困難である.

本研究で構築したデータセットのもとでかなりの精度の分類器を学習できたという事実を踏まえ, 他言語や別ドメインのデータに提案手法を適用する際の参考になるように, ラベル付きデータの品質を調べた.  $D_{vol}^l$  と  $D_{ani}^l$  からポジティブ・ネガティブのイベントをそれぞれ 100 件ずつ抽出し, それらに正しいラベルが付与されているかどうか, 著者 1 名が人手で確認した.

表 8 に結果を示す. ほとんどのイベントに正しいラベルが付与されていることが確認できた. 日本語の  $D_{ani}^l$  のネガティブ (無生名詞) のイベントは比較的ラベル付けの精度が悪かった. これは主に主語の認識に失敗に起因するものであった. また, 英語の  $D_{ani}^l$  のネガティブ (無生名詞) のイベントも比較的ラベル付けの精度が悪かった. これは, 「location」の下位語を無生名詞として利用していたが, 国名など一部の名詞は組織や集団を表す有生名詞としての用法もあることに起因していた.

## 7 おわりに

本研究では, 意志性分類と主語有生性の密接な関係に注目し, ヒューリスティクスを利用し

て自動構築した訓練データのもとでこれらの正則化付き同時学習を行う, 弱教師あり学習手法を提案した. 日本語・英語の実験を通して, 提案手法によりラベル付きデータを人手で構築することなく高精度の分類器を学習できることを確認した.

本研究では単独のイベントを入力としてそれが意志的であるかどうか分類する設定を考えたが, 文や段落, 文章などイベントより大きな言語単位を入力とし, 入力テキスト中の各イベントの意志性を分類する設定では, イベントの前後の文脈を考慮しなければならないケースがある. 例えば, 例 (22) の下線部のイベントは, 前の文脈を考慮しないなら非意志的とみなすのが自然であるが, 前の文脈を考慮すれば意志的とみなすのが自然である.

(22) 私は彼女を庇って非難を受けた. (V)

このようなケースに対応できないことは本研究の限界であり, より広い文脈を考慮した意志性分類は今後の課題である.

意志性はイベントの基本的な属性であり, その応用は広い. しかし, 意志性に着目した研究はほとんど存在しない. これは, 容易に利用可能な意志性分類器も意志性分類器を低コストで構築する手法もなく, 意志性に基づく処理をするにはそれなりのコストをかけて意志性分類器を構築することから始める必要があったからである. その意味で, 本研究は意志性に着目した研究を促進する可能性を持っている. 今後は, 本研究で構築した意志性分類器をテキストマイニング等の諸タスクで活用することに取り組み, そこで得られた観察をもとに手法のさらなる改善に取り組みたい.

## 謝 辞

本研究の一部は The Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI 2022) で発表したものである (Kiyomaru and Kurohashi 2022). 本研究はヤフー株式会社の支援を受けた. ここに感謝の意を表す.

## 参考文献

- Abe, S., Inui, K., and Matsumoto, Y. (2008a). “Acquiring Event Relation Knowledge by Learning Cooccurrence Patterns and Fertilizing Cooccurrence Samples with Verbal Nouns.” In *Proceedings of the 3rd International Joint Conference on Natural Language Processing: Volume-I*, pp. 497–504.
- Abe, S., Inui, K., and Matsumoto, Y. (2008b). “Two-Phased Event Relation Acquisition: Cou-

- pling the Relation-Oriented and Argument-Oriented Approaches.” In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp. 1–8, Manchester, UK. Coling 2008 Organizing Committee.
- Berinsky, A., Huber, G., and Lenz, G. S. (2012). “Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk.” *Political Analysis*, **20**, pp. 351–368.
- Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., and Kalai, A. T. (2016). “Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings.” In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R. (Eds.), *Advances in Neural Information Processing Systems*, Vol. 29. Curran Associates, Inc.
- Chambers, N. and Jurafsky, D. (2008). “Unsupervised Learning of Narrative Event Chains.” In *Proceedings of ACL-08: HLT*, pp. 789–797, Columbus, Ohio. Association for Computational Linguistics.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). “Unsupervised Cross-lingual Representation Learning at Scale.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8440–8451, Online. Association for Computational Linguistics.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Du, L., Ding, X., Liu, T., and Li, Z. (2019). “Modeling Event Background for If-Then Commonsense Reasoning Using Context-aware Variational Autoencoder.” In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2682–2691, Hong Kong, China. Association for Computational Linguistics.
- Ganin, Y. and Lempitsky, V. (2015). “Unsupervised Domain Adaptation by Backpropagation.” In Bach, F. and Blei, D. (Eds.), *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37 of *Proceedings of Machine Learning Research*, pp. 1180–1189, Lille, France. PMLR.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., and Lempitsky, V. (2016). “Domain-Adversarial Training of Neural Networks.” *Journal of Machine Learning Research*, **17** (59), pp. 1–35.

- Inui, T., Inui, K., and Matsumoto, Y. (2003). “What Kinds and Amounts of Causal Knowledge Can Be Acquired from Text by Using Connective Markers as Clues?” In *International Conference on Discovery Science*, pp. 180–193. Springer.
- Jin, X., Wei, Z., Du, J., Xue, X., and Ren, X. (2020). “Towards Hierarchical Importance Attribution: Explaining Compositional Semantics for Neural Sequence Models.” In *International Conference on Learning Representations*.
- Kawahara, D. and Kurohashi, S. (2006). “A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis.” In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pp. 176–183, New York City, USA. Association for Computational Linguistics.
- Kennedy, B., Jin, X., Mostafazadeh Davani, A., Dehghani, M., and Ren, X. (2020). “Contextualizing Hate Speech Classifiers with Post-hoc Explanation.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5435–5442, Online. Association for Computational Linguistics.
- Kingma, D. P. and Ba, J. (2015). “Adam: A Method for Stochastic Optimization.” In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.
- Kiyomaru, H. and Kurohashi, S. (2022). “Minimally-Supervised Joint Learning of Event Volitionality and Subject Animacy Classification.” *Proceedings of the AAAI Conference on Artificial Intelligence*, **36** (10), pp. 10921–10929.
- Lee, H.-Y. and Jun, S. (2008). “Constructing an Ontology of Coherence Relations: An Example of ‘Causal Relation’.” In *Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation*, pp. 245–252, The University of the Philippines Visayas Cebu College, Cebu City, Philippines. De La Salle University, Manila, Philippines.
- Liu, C.-H., Moriya, Y., Poncelas, A., and Groves, D. (2017). “IJCNLP-2017 Task 4: Customer Feedback Analysis.” In *Proceedings of the IJCNLP 2017, Shared Tasks*, pp. 26–33, Taipei, Taiwan. Asian Federation of Natural Language Processing.
- Ramponi, A. and Plank, B. (2020). “Neural Unsupervised Domain Adaptation in NLP—A Survey.” In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 6838–6855, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Reisinger, D., Rudinger, R., Ferraro, F., Harman, C., Rawlins, K., and Van Durme, B. (2015). “Semantic Proto-Roles.” *Transactions of the Association for Computational Linguistics*, **3**, pp. 475–488.
- 齋藤純, 坂口智洋, 柴田知秀, 河原大輔, 黒橋禎夫 (2018). 述語項構造に基づく言語情報の基本

- 単位のデザインと可視化. 言語処理学会第 24 回年次大会発表論文集, pp. 93–96, 岡山. [J. Saito et al. (2018). Jutsugo Ko Kozo ni Motozuku Gengo Joho no Kihon Tan'i no Dezain to Kashika. Proceedings of the 24th Annual Meeting of the Association for Natural Language Processing, pp. 93–96.].
- Shah, D., Lei, T., Moschitti, A., Romeo, S., and Nakov, P. (2018). “Adversarial Domain Adaptation for Duplicate Question Detection.” In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1056–1063, Brussels, Belgium. Association for Computational Linguistics.
- Shen, J., Qu, Y., Zhang, W., and Yu, Y. (2018). “Wasserstein Distance Guided Representation Learning for Domain Adaptation.” *Proceedings of the AAAI Conference on Artificial Intelligence*, **32** (1), pp. 4058–4065.
- Speer, R., Chin, J., and Havasi, C. (2017). “ConceptNet 5.5: An Open Multilingual Graph of General Knowledge.” In *Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI’17*, pp. 4444–4451. AAAI Press.
- Teichert, A., Poliak, A., Van Durme, B., and Gormley, M. (2017). “Semantic Proto-Role Labeling.” *Proceedings of the AAAI Conference on Artificial Intelligence*, **31** (1), pp. 4459–4465.
- Wang, A. and Cho, K. (2019). “BERT has a Mouth, and It Must Speak: BERT as a Markov Random Field Language Model.” In *Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation*, pp. 30–36, Minneapolis, Minnesota. Association for Computational Linguistics.
- Wenzek, G., Lachaux, M.-A., Conneau, A., Chaudhary, V., Guzmán, F., Joulin, A., and Grave, E. (2020). “CCNet: Extracting High Quality Monolingual Datasets from Web Crawl Data.” In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 4003–4012, Marseille, France. European Language Resources Association.
- White, A. S., Reisinger, D., Sakaguchi, K., Vieira, T., Zhang, S., Rudinger, R., Rawlins, K., and Van Durme, B. (2016). “Universal Decompositional Semantics on Universal Dependencies.” In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1713–1723, Austin, Texas. Association for Computational Linguistics.
- Zhao, J., Wang, T., Yatskar, M., Cotterell, R., Ordonez, V., and Chang, K.-W. (2019). “Gender Bias in Contextualized Word Embeddings.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 629–634, Minneapolis, Minnesota. Association for Computational Linguistics.
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., and Chang, K.-W. (2017). “Men Also Like Shop-

ping: Reducing Gender Bias Amplification using Corpus-level Constraints.” In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2979–2989, Copenhagen, Denmark. Association for Computational Linguistics.

## 付録

### A モダリティのリスト

本研究ではモダリティを伴うイベントは対象としなかった。表9に本研究で扱ったモダリティのリストとその例を示す。

### B 意志的副詞・非意志的副詞のリスト

意志性のラベル付与のために準備した意志的副詞・非意志的副詞のリストは以下の通りである。括弧内の数字は頻度を表す。

日本語の意志的副詞 あえて (5,293), 急いで (4,187), じっくり (4,017), 慎重に (3,743), のんびり (3,262), わざわざ (3,222), さっさと (1,945), 集中して (1,194), 意図的に (920), わざと (880), 意識的に (786), 念入りに (766), 気楽に (591), ちゃっかり (510), 注意深く (496)。

日本語の非意志的副詞 思わず (18,115), つい (15,897), 自動的に (14,212), ふと (12,050), ついつい (10,054), 自動で (5,058), 気づいたら (1,414), うっかり (1,333), 幸いにも (950), 幸運にも (571), 思いがけず (546), あいにく (422), 幸運なことに (285), 不運にも (63), 不運なこ

モダリティ	例
意志	行くつもりだ
勧誘	行こうよ
命令	行きなさい
禁止	行くな
評価	行ってもいい
認識-推量	行くだらう
認識-蓋然性	行くかもしれない
認識-証拠性	行くようだ
依頼	行って下さい
推量・伝聞	行くと思う

表9 モダリティのリストとその例。

とに (32).

英語の意志的副詞 carefully (13,594), thoroughly (12,468), actively (10,379), deliberately (3,366), intentionally (2,713), consciously (1,846), purposely (1,391), hurriedly (942), attentively (839), proactively (388).

英語の非意志的副詞 unfortunately (13,070), automatically (12,824), accidentally (5,272), unexpectedly (3,106), luckily (1,894), instinctively (1,321), unconsciously (1,059), inadvertently (999), unintentionally (635), carelessly (384).

## 略歴

**清丸 寛一**：2022年京都大学大学院情報学研究科博士課程修了。博士（情報学）。同年より京都大学大学院情報学研究科特定研究員。自然言語処理の研究に従事。言語処理学会，ACL，各会員。

**黒橋 禎夫**：1994年京都大学大学院工学研究科電気工学第二専攻博士課程修了。博士（工学）。2006年4月より京都大学大学院情報学研究科教授。自然言語処理，知識情報処理の研究に従事。言語処理学会10周年記念論文賞，同20周年記念論文賞，第8回船井情報科学振興賞，2009 IBM Faculty Award 等を受賞。2014年より日本学術会議連携会員。

(2021年12月1日 受付)

(2022年3月11日 再受付)

(2022年4月20日 採録)