情報処理学会論文誌 Vol.65 No.12 1735–1748 (Dec. 2024)

# 単語の関連性に着目した複合語型隠語の検出

羽田 拓朗[1,2,a)] 清 雄一[1] 田原 康之[1] 大須賀 昭彦[1]

受付日 2024年3月8日, 採録日 2024年9月9日

**概要**：近年，マイクロブログにおける違法薬物取引などの犯罪行為の勧誘が増加の一途をたどっており，社会的な問題となっている．こういった犯罪を取り締まるためにサイバーパトロールが行われている一方，犯罪へと誘導する投稿を行う者たちは，監視される対象となるキーワード（「援助交際」，「大麻」など）を避けるため，犯罪の意図をカモフラージュした用語，いわゆる「隠語」を駆使して，監視の目をかいくぐりながら巧妙にやり取りを続けている．これらの隠語は，監視側が一度把握したとしても，一般的に広まれば陳腐化し，新たな隠語が使われ始めるため，つねに最新の隠語を把握する労力が必要となる．これまでに，犯罪の意図で用いられる隠語を検出することを目的とし，投稿内の単語の用途の差異から隠語を検出する手法を提案し，一定の成果を上げてきた．本稿では，既存研究ではこれまで検出できなかった2つ以上の単語を結合させた複合語型の隠語を，単語の関連性を活用することで検出する手法を提案する．そして，提案手法の効果を確認するため，複合語検出実験および隠語検出実験を行った．その結果，既存手法より7パーセントポイント高い精度で複合語型隠語を含めた隠語を検出できた．効果を確認するため，7名の現役警察官へヒアリングをしたところ，検出した複合語型隠語のうち93.2%が過半数の警察官にとって未知のものであることが確認できた．

キーワード：隠語，複合語，マイクロブログ，Twitter，Word Embedding，Word2vec

# Detection of Compound-Type Dark Jargons Using Similar Words

Takuro Hada[1,2,a)] Yuichi Sei[1] Yasuyuki Tahara[1] Akihiko Ohsuga[1]

Received: March 8, 2024, Accepted: September 9, 2024

***Abstract:*** Recently, drug trafficking on microblogs has increased and become a social problem. While cyber patrols are being conducted to combat such crimes, those who post messages that lead to crimes continue to communicate skillfully using so-called “dark jargon,” a term that conceals their criminal intentions, to avoid using keywords (“drug,” “marijuana,” etc.) of the target of monitoring. Evading detection by the eyes of monitoring, they continue to communicate with each other skillfully. Even if the monitors learn these dark jargons, they become obsolete over time as they become more common, and new dark jargons emerge. We have proposed a method for detecting dark jargons with criminal intent based on differences in the usage of words in posts and have achieved a certain level of success. In this study, by using similar words, we propose a method for detecting compound-type dark jargons that combines two or more words, which have been difficult to detect using existing methods. To confirm the effectiveness of the proposed method, we conducted a detection experiment with compound words and a detection experiment with dark jargons. The experimental results showed that not only the proposed method was able to detect codewords with an accuracy of 7 persentages points higher than existing methods, but also 10 compound-type dark jargons that could not be detected by the existing method were detected. Furthermore, interviews with seven police officers confirmed that 93.2% of the dark jargons detected in the experiment were unknown to the majority of the police officers.

***Keywords:*** dark jargon, compound word, Microblog, Twitter, Word Embedding, Word2vec

## 1. はじめに

ソーシャルネットワーキングサービス（以下，「SNS」という）の急激な普及にともない，SNSを利用した犯罪も増加の一途をたどっているところ，SNS上での勧誘が起因となる犯罪については，違法薬物売買以外にも，児童買春や

[1] 電気通信大学大学院情報理工学研究科
UEC, Chofu, Tokyo 182–8585, Japan
[2] 警察庁刑事局組織犯罪対策部組織犯罪対策第一課
First Organized Crime Countermeasures Division, Organized Crime Department, Criminal Investigation Bureau, National Police Agency, Chiyoda, Tokyo 100–8974, Japan
[a)] hadatakuro@gmail.com

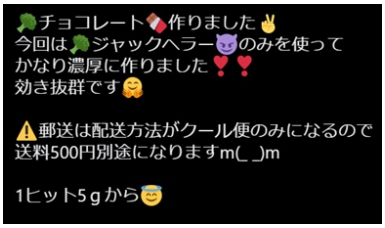


**図 1** 隠語が用いられた文章例：違法薬物の入荷を隠語を使って表現している

**Fig. 1** Example sentences with dark jargons from Twitter.

闇バイトの募集，自殺幇助などにも及んでいる．犯罪を未然に防ぐためにも，SNS 上における犯罪を唆す書き込みは早急に検知すべきであり，警察や SNS の運営会社によるサイバーパトロールが行われている．一方このような違法な取引を目的とした投稿者は，サイバーパトロールから逃れるため，犯罪に直接関係する単語（「大麻」,「覚醒剤」など）を避け，**図 1** のように隠語を用いて，言葉の意味を知っている者同士だけで違法な取引を実施する傾向にある．

隠語の例をあげると，違法薬物売買においては，大麻の場合，「マリファナ」,「ガンジャ」，覚醒剤には「エス」,「シャブ」といった単語が用いられていることが一般に知られている．これらの隠語を定期的にキーワード検索により検知する対策をとったとしても，効果は限定的である．なぜなら，隠語の特徴として，一般的に認知されると監視を回避するために新しい隠語が作られたり，今まで使われていなかった一般的な言葉に隠語の意味が付与されたりするようになるからである [1], [2], [3], [4].

たとえば，大麻の場合，「草」,「雑草」,「ジョイント」，覚醒剤の場合，「アイス」,「クリスタル」といった隠語が新たに使われるようになっている．また，直接手渡しを意味する隠語として「手押し」が従来使われてきたが，認知度が高くなってきたために「ハンドプッシュ」という隠語に変わってきたという例がある．その結果，監視側は継続して新しい隠語を把握し続け，それらを検知対象として追加していく必要があるため，負担が非常に大きい．このようなことから，犯罪を誘発する投稿の検出や，犯罪者が用いる言語を理解するために，新しい隠語の検出を目指す．

隠語検出の手法は近年いくつか存在しているが [4], [5], [6]，事前処理として文章中の単語を正しく分けることを前提としている．こうして得られた単語集合の各単語に対して，隠語であるかどうかを判定する．空白で単語が区切れる英語などと異なり，日本語や韓国語などは切れ目が明確でないため，まず単語の分かち書きが必要である．この分かち書きは，一般に事前に用意されている辞書を用いて行われる．しかし隠語検出のシナリオでは，そもそも対象の言葉が辞書に登録されていないことが想定されるため，正しく分かち書きを行うことができず既存手法をそのまま用いることができない．実際に羽田らの手法 [7] において分割され 1 つの単語として検出できなかった例としては，大麻の品種名である「グリーン・クラック」や「ブルー・ドリーム」などがあった（「·」は文節が区切られていた位置を表す）．

このような複合語でかつ隠語であるものを本研究では，複合語型隠語と呼ぶこととする．

既存手法において，複合語型隠語が検出できない理由として，日本語を意味解析していく際に，分かち書きの工程が発生するが，この分かち書きにおいて，複合語型隠語に該当する単語の性質として，一般的な認識度が低い可能性も高く，分かち書きの際に文節が短く分割されてしまうからである．この課題について，分かち書き用の内部辞書に複合語型隠語となるような単語を登録すれば，複合語単位の文節で区切られることが期待できるが，単語の認知が低いことが懸念されることから，自動的に内部辞書が充実することは期待できない．そのため，隠語かどうかを考慮せずに幅広く複合語を内部辞書に登録し，そのあと，事前に複合語型隠語の辞書登録ができれば，複合語型隠語の文節で分かち書きが行われ，既存手法により，複合語型隠語の検出が期待できる．

そこで，本稿では，自動的に複合語を検出する手法を提案し，そこで検出した複合語を事前に分かち書き用の内部辞書に登録させ，既存手法と組み合わせることで，より隠語検出の検出範囲を広げ，かつ精度を向上させることを目指す．

なお，複合語の辞書的な意味に従うと，複合語とは，本来独立した単語が 2 つ以上結合して，新たに 1 つの単語としての意味・機能を持つようになったものと定義[*1]されている．そのため日本語の分かち書きツールが 100%の精度で分かち書きを行えるのであればすべての（辞書的な意味における）複合語も正しく 1 つの単語として分割されることが期待される．しかし，最新の複合語や隠語の複合語など，分かち書きツールの内部辞書に掲載されていない単語が対象の場合，誤って 2 つ以上の単語に分割されてしまう場合があり，この誤りにより隠語の検出精度が低下してしまう．

本研究においては，「分かち書きで分割されてしまうが，本来は分割されるべきではない語」を複合語と定義する．この複合語の検出を目的とすることで，隠語検出の精度向上をめざす．なお，ここで「単語」とは「文法上，意味・職能を持った最小の言語単位[*2]」という辞書的な意味における単語のことを指すため，分かち書きの内部辞書によって何が複合語となるかが異なる．本研究では，分かち書きの内部辞書として，Sudachi（version1.1）[8] を用いた．辞書については雑多な固有名詞まで収録された Full 辞書を選択した．

*1 デジタル大辞泉（小学館）より引用
*2 weblio.jp より引用

本稿の一部は，SMASH（Symposium on Multi Agent Systems for Harmonization）22 Summer Symposium（電子情報通信学会「人工知能と知識処理」研究会共催）[9] および 15th International Conference on Agents and Artificial Intelligence（ICAART 2023）[10] で発表したものである．

本稿の構成は，以下のとおりである．2 章では，本研究の背景について記載する．続いて，3 章では本研究で提案する手法について述べ，4 章では複合語の検出実験および結果を示し，5 章では 4 章の実験をふまえた複合語型隠語の実験および結果を示す．そして 6 章では，実験を通じて提案手法に関する考察について述べる．7 章では，関連研究を紹介する．最後に，本研究の結論を 8 章で述べる．

## 2. 背景

### 2.1 SNS に起因した犯罪の増加とその対策について

**図 2** は，年代別の大麻事犯における検挙人数の推移を表すものであり，特に 10 代と 20 代において年々増加傾向にあることが見て取れる．Twitter は，若年層の利用が多いことから若年層の目に触れることを狙って悪用されることが多いと考えられる．

こういったことを受け，たとえば平成 30 年 8 月には薬物乱用対策推進会議において，「第五次薬物乱用防止五か年戦略」を策定し，その中でも巧妙化・潜在化する薬物の密売についても危惧し，対策を講じている [11]．

近年，SNS 上で闇バイトと呼ばれる犯罪に加担する者を募集する投稿が問題視されている．実際に令和 4 年 12 月～令和 5 年 1 月に広域で発生した一連の強盗事件の実行犯が Twitter 上で闇バイトとして募集されたものであったと発覚した [12]．

これを受けて，省庁横断で対策が開始されることとなり，「SNS で実行犯を募集する手口による強盗や特殊詐欺事案に関する緊急対策プラン」を策定し，各省庁に対応を求めている [13]．その中でも，「実行犯を生まない」ための対策として，「闇バイト等情報に関する情報収集，削除，取締り等の推進」や「サイバー空間からの違法・有害な労働募集の排除」などが推進されることとなり，たとえば，「インターネット・ホットラインセンター（IHC）」では，警察庁から委託を受けインターネット上の監視をしているところ，近年，インターネット上において闇バイト募集情報が氾濫している状況をふまえ，令和 5 年 9 月 29 日より「犯罪実行者募集情報（いわゆる闇バイトの募集）」に関する通報の受理・分析などを新たに開始し，12 月末までの約 3 カ月間で 4,411 件の検知が報告されている [14]．なお，2,979 件について対応依頼がなされており，そのうち 2,136 件で削除が完了していると報告されている（令和 6 年 1 月末時点）．

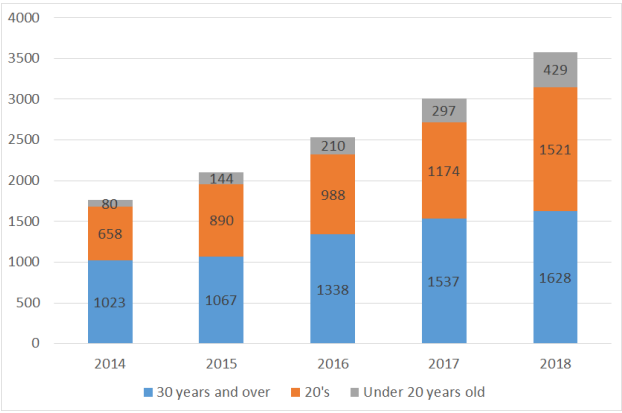


**図 2** 年代別大麻事犯の検挙人員の推移（警察庁資料を元に作成[*3]）
**Fig. 2** Based on the number of arrests for marijuana offenses by age group (data from the National Police Agency).

*3 平成 30 年における組織犯罪の情勢

また，令和 5 年 8 月 8 日には，第六次薬物乱用防止五か年戦略が決定され，その中では，SNS などにより「闇バイト」として安易に密輸に加担させられることについても言及されており，さらなる薬物対策が推進されている [15]．

このような状況をふまえるとともに，SNS の中では Facebook や Twitter が主に利用されていることと [16]，SNS の中でも Twitter は，利用者の多さに加え不特定多数が閲覧できるため，違法な取引が行われやすい環境にあること，その際，隠語でやり取りされる傾向にあることなどから，本研究では特に Twitter に着目することとした．Twitter における隠語に関するツイートは以下の 4 種類に分類される [7]．

(1) 既知の隠語（および犯罪に直接関係する単語）だけが利用されたツイート
(2) 未知の隠語だけが利用されたツイート
(3) 既知の隠語（および犯罪に直接関係する単語）と未知の隠語が混在したツイート
(4) 既知の隠語（および犯罪に直接関係する単語）も未知の隠語も利用されていないツイート

このうち，(3) のツイートが存在することを前提として，既知の隠語（および犯罪に直接関係する単語）を基に，未知の隠語を検出することを目指す．

なお，2023 年に Twitter は X に改称されたが，本研究では便宜上，2023 年以前の用語で呼称することとする．また今回の実験やデータ収集については，仕様変更前に実施したものである．

### 2.2 隠語および犯罪関連語について

隠語は，特定の社会・集団内でだけ通用する特殊な語と定義されており[*4]，我々の身の回りでも様々な場面で使用されているが，その中でも本研究における隠語の対象は，警察などの目をかいくぐり，犯罪などに用いられる単語，特に違法薬物売買に関係する隠語とした．それ以外に，そ

*4 デジタル大辞泉（小学館）より引用

表 1 隠語・犯罪関連語の整理
**Table 1** Clarification of dark jargons and related words.

| | | 認知度（言葉としての認知度） | | | |
|---|---|---|---|---|---|
| | | 高 | | 低 | |
| | | 分類 | 例 | 分類 | 例 |
| 造語 | | 隠語 | シャブ，チャリ，パパ活，神待ち | 隠語 | 円光，UD |
| 造語以外 | 一般語の意味を変化（転用） | 隠語 | アイス，野菜，チョコ，レモン | 隠語 | ゴリラグルー，カリフォルニアオレンジ |
| | 一般語の意味のそのまま | 犯罪関連語 | 高収入，営業中，都内，大麻，覚せい剤 | 隠語 | ホワイトクッシュ，グリーンクラック |

の単語自体だけでは隠語として成立しない，犯罪行為を指さないが，隠語と一緒に出現する傾向が高い，もしくは，複数の同様の単語から，ある特定の犯罪を想起する単語について，「犯罪関連語」と定義した．

本研究で対象とする単語について，具体的に，**表 1** のとおり，大きく 2 つの観点で分類した．

(1) 造語かどうか
違法な取引を行うため，意図的に造られた単語が該当する．中には，知らない言葉であったとしても，音が一緒であったり，漢字などから連想できるようなものもある．たとえば，援助交際関連の隠語として用いられる，「円光」や「えん」などがこれにあたる．一方，造語でないものについては，さらに一般語の意味を変化しているかどうか，すなわち転用（カモフラージュ）しているかどうかの観点を追加した．転用については，一般的に使用されている単語に隠語の意味を付与し，カモフラージュさせて使われる単語がこれに該当する．たとえば，大麻を表す隠語として，「野菜」や「草」といった隠語がある．

(2) 認知度が高いかどうか
対象の単語自体が一般的に認知されていないため，そのまま用いたとしても特定の人々にしか分からないことから，特段別の単語に置き換えなくとも隠語と同等の効果が認められる単語が該当する．

### 2.3 複合語型隠語について

複合語とは，本来独立した単語が 2 つ以上結合して，新たに 1 つの単語としての意味・機能を持つようになったものと定義されている[*5]．

[*5] デジタル大辞泉（小学館）より引用

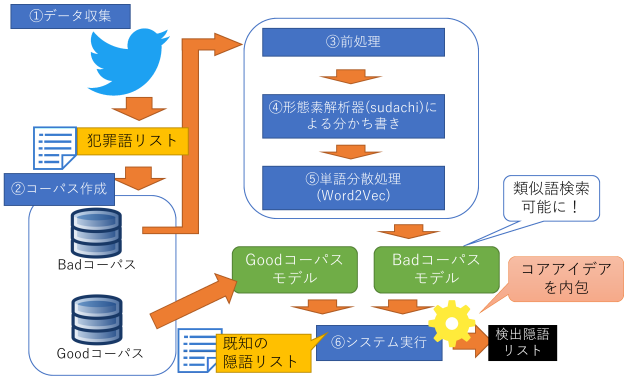


図 3 隠語検出手法のプロセス
**Fig. 3** Process of dark jargons detection method.

複合語の定義については，1 章で説明したところ，複合語の例として，「ほん・ばこ（本箱）」「やま・ざくら（山桜）」などがある．なお，「·」は単語の連結を表す．

複合語の隠語の例として，大麻の隠語としては，「レモンスカンク」，「ゴリラグルー」，「ホワイトウィドー」などがこれまで確認したツイートの中から確認できている．ただし，これまでの隠語検出の既存手法 [4], [5], [6] では，前処理における分かち書きの文節単位に依存することから，前述の単語については，それぞれ，「レモン・スカンク」，「ゴリラ・グルー」，「ホワイト・ウィドー」というように単一単語の文節で区切られてしまう．そのため，羽田らによる既存研究の手法 [7], [17] では，「ゴリラ」が検出されることはあっても「ゴリラグルー」という 1 単語として検出できなかった．

このようなことから，1 章でも述べたが，事前に複合語型隠語を分かち書き用の辞書に登録することができれば，複合語型隠語の文節で分かち書きが行われ，既存手法により，複合語型隠語の検出が期待できる．

## 3. 複合語型隠語の検出について

### 3.1 複合語の検出のアプローチ

これまでに著者らは，犯罪を誘導する隠語を検出するため，2.1 節で記載したツイートの分類のうち (1)，(3) を対象としたツイート群（以下，「Bad コーパス」という）と上記の (4) を対象としたツイート群（以下，「Good コーパス」という）の 2 つのツイート群に分類した．そして，2 つのコーパスのそれぞれで Word2vec [18] を用いて単語分散表現モデルを構築し，同じ単語におけるコーパス間の類似語の差異に基づき，既知の隠語から未知の隠語を検出する方法について提案した（**図 3**）[7], [17].

この既存手法においても，他の多くの隠語検出研究と同様に分かち書きされた単語に基づき隠語を検出するため，文節で区切られる単語で構成される複合語型隠語は検出ができないという課題があった．本課題においては，まず複合語の検出が課題となる．たとえば，大麻の隠語として用

いられる「グリーンクラック」は，文節が「グリーン」と「クラック」の間で区切られる．一方で，単語分散表現モデルを構築した際，設定パラメータも関与するが，複合語となるような 2 単語は，本来類似語でなくても分散表現モデル上では非常に近い単語であると認識されると考えられる．なぜなら，複合語特に複合語型隠語については，同じ文脈で出る頻度が高いことが想定され，単語どうしの関連が強いことが考えられることから，複合語の文節で区切られた単語どうしの類似語として互いの単語が上位に出現する．ここで，「グリーンクラック」という単語について，「グリーン」と「クラック」のそれぞれの単語の Bad コーパスで構築された単語分散表現モデルにおける類似語を調べたところ，「グリーン」の類似語 1 位は「クラック」であり，一方の「クラック」の類似語 1 位は「グリーン」であり，双方の類似語 1 位どうしという結果が得られた．この際の Word2vec における設定パラメータである Windows Size は 2 としている．このことから，ともに互いの類似語上位の単語が一致する単語を検出することで，複合語を自動的に検出できる可能性がある．

この複合語検出について，Algorithms 1，2 で示す手法を提案する．提案手法の概要は以下のとおりである．Bad コーパスで構築した単語分散表現モデル内の単語群 $\mathcal{A}$ 内の単語 $\alpha$ の上位 $M$ 位までの類似語を検索し得られた $M$ 個の集合を $\mathcal{S}(\alpha)$ とおく．$\mathcal{S}(\alpha)$ の各要素 $S(\alpha)_1 \ldots S(\alpha)_M$ についても，同様に上位 $M$ 位までの類似語を検索し（*Get similar words*），それぞれの $\mathcal{S}(\alpha)$ の類似語上位 $M$ 位の単語群を求める．

もし $\mathcal{S}(\alpha)$ の類似語の中に $\alpha$ と同一の文字があれば，「$\alpha \cdot S(\alpha)$」と，「$S(\alpha) \cdot \alpha$」という複合語候補を作成する（*Concat*）．そして，「$\alpha \cdot S(\alpha)$」と「$S(\alpha) \cdot \alpha$」のそれぞれについて元の Bad コーパスにおける出現回数を確認し，一定回数以上のものは複合語と判定した（*Count*）．

ここで複合語について前後の単語を総あたりで登録した場合，「アイス食べたい」のような明らかに複合語ではない文章も登録されることとなる．そのため，類似語を利用することは複合語を検出する効果的な方法である．

なお，本稿では，単語の意味が異なっていても，分散表現モデル上で近い単語は「類似語」と表現することとする．

### 3.2 複合語登録までのプロセス

以下の流れで複合語の検出を行った．

(1) Bad コーパスを分かち書きをし，Word2vec [18] を用いて複合語検出用の単語分散表現モデルを構築し，構築した単語分散表現モデルから，モデル内の単語を単語群 $\mathcal{A}$（$\mathcal{A} = \{\alpha_1, \ldots, \alpha_n\}$（$\alpha_i$ は $i$ 番目の単語））として作成する（CreateList 関数）．

(2) $\alpha_i$ の上位 $M$ 位までの類似語を検索する（$\mathcal{S}(\alpha_i)_1$ ... $\mathcal{S}(\alpha_i)_M$．$\mathcal{S}(\alpha_i)_{jth}$ は類似語上位 $j$ 番目の単語）．

**Algorithm 1** ***Main_Detect_Com_words***

**Input:** M, Bad_Corpus,X
**Output:** All Compound words List(Z)

```
Z ⇐ {}
Bad_Corpus_Word_List ⇐ CreateList(Bad_Corpus)
for all Word in Bad_Corpus_Word_List do
  Comp_word-candidate_List ⇐ SIMILAR_Comp_
  words(Word, M, Bad_Corpus)
  for all Comp_word-candidate in Comp_word-candidates_
  List do
    if (Count(Comp_word-candidate, Bad_Corpus) >= X)
    then
      Z ⇐ Z ∪{Comp_word-candidate}
    end if
  end for
end for
return(Z)
```

**Algorithm 2** Function ***SIMILAR_Comp_words***

**Input:** Word, M, Corpus
**Output:** Compound words list(Y)

```
Y ⇐ {}
Sim_Comp_words_list ⇐ Corpus.Get_similar_words(Word,
M)
for all Sim_word in Sim_Comp_words_list do
  Sim_Sim_Comp_words_list ⇐ Corpus.Get_similar_
  words(Sim_word, M)
  for all Sim_Sim_word in Sim_Sim_Comp_words_list do
    if Sim_Sim_word = Word then
      Y ⇐ Y ∪ Concat{Sim_word, Word}
      Y ⇐ Y ∪ Concat{Word, Sim_word}
    end if
  end for
end for
return(Y)
```

(3) 次に，$\mathcal{S}(\alpha_i)_j$ のうち，$j = 1 \sim M$ までのそれぞれの単語について，$M$ 位までの類似語検索を行う．（$\mathcal{S}(\mathcal{S}(\alpha_i)_j)_1$ ... $\mathcal{S}(\mathcal{S}(\alpha_i)_j)_M$．$\mathcal{S}(\mathcal{S}(\alpha_i)_j)_k$ は類似語上位 $k$ 番目の単語）．

(4) もし $\alpha_i = \mathcal{S}(\mathcal{S}(\alpha_i)_j)_k$ となれば，“$\alpha_i \cdot \mathcal{S}(\alpha_i)_j$” と “$\mathcal{S}(\alpha_i)_j \cdot \alpha_i$” を複合語候補として作成する．$\alpha$ について，複数の複合語候補が出る可能性はあるが，すべて対象とした．

(5) それぞれの単語について，元の Bad コーパスにおける出現回数を確認し，別で定める $\mathcal{X}$ 回以上のものは複合語と見なすこととした．

### 3.3 精度向上についての検討

3.2 節で説明した手法に加え，不要な単語の検出を抑制し，隠語検出の精度を向上させるため，いくつかの機能を検討および検証し，機能追加を行い，提案手法とした．

#### 3.3.1 Good コーパスとの比較

一般的な複合語の検出により，複合語型隠語の検出機会が減少することを回避するため，2 つのコーパス間（Good コーパス，Bad コーパス）間で同じ単語 $\alpha$ の類似語を検索し，同じ単語が出現した単語 $\alpha$ は，隠語としての文脈で出現していないと考えられる．そこで，類似語検索時に Good コーパスの類似語も検索し，上位 20 位までに出現した場合，その単語は一般的な単語として扱い，複合語としての辞書登録を行わないこととした．

#### 3.3.2 品詞によるフィルタ

著者らによってツイートを調査したところ，複合語型隠語は，主に名詞どうしから構成されているものが多かったことから，複合語を結合させる前に，品詞分類を実施するようにし，名詞だけを抽出するようにした．

#### 3.3.3 文字数による制限

複合語型隠語は，1 文字の平仮名，カタカナ，記号が別の単語と結合して複合語型隠語となるケースは見あたらなかったことから，1 文字の平仮名，カタカナ，記号は複合語の候補からは除外し，候補数を絞るようにした．ただし，漢字では「紙」や「罰」といった，一語で成立するものや，アルファベットとしては「L グミ」があったため，漢字やアルファベットは対象とすることとした．

#### 3.3.4 辞書内の単語の削除

複合語検出プログラムにより検出された複合語候補の中には，元々の分かち書き用の辞書に登録されている「質量」といった単一単語一語として認識される「援助交際」といった複合語などがあったが，これらの単語については，複合語の候補から除外することとした．

## 4. 複合語検出実験

### 4.1 概要

提案手法を用いて，複合語検出の検証実験を行った．ここで作成した分かち書き用のユーザ辞書に基づき，後に 5 章で説明する複合語型隠語検出実験を行った．

### 4.2 実験のプロセス（複合語検出）

**図 4** のとおり，複合語検出実験を行った．

#### 4.2.1 データ収集

TwitterAPI を利用し，Twitter のデータを約 1 年間（2019/07/19 から 2020/07/27）収集したうち，本文データのみを使用した．続いて，違法薬物売買関連のダブルミーニングの隠語を除いた既知の隠語群（以下，「犯罪語リスト」）のうち，いずれかの単語が含まれるツイートを抽出した後，その投稿アカウントに着目し，違法な取引のツイートをしたアカウントリストとして作成した．その後，そのリストをキーに再度同じ期間におけるツイートを収集し，Bad コーパスを作成した．

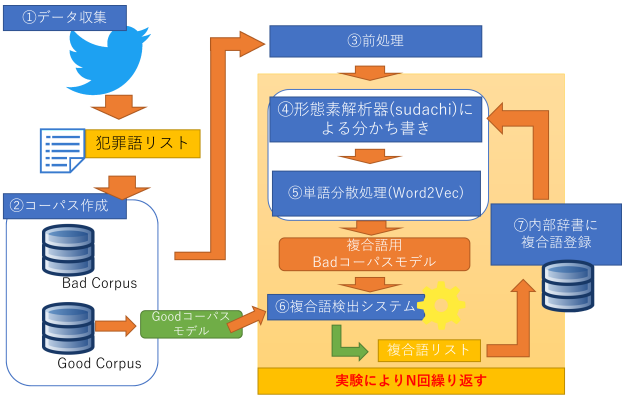


**図 4** 複合語検出手法のプロセス

**Fig. 4** Process of compound word detection method.

#### 4.2.2 コーパス作成

用意したコーパスについては，以下の 2 つである．

(1) Bad コーパス

4.2.1 節のとおり．

(2) Good コーパス

一般的なコーパスとして，大規模なツイートコーパスを利用する方法を選択し，株式会社ホットリンクの作成した日本語大規模 SNS+Web コーパスを利用した [19]．

#### 4.2.3 前処理

隠語検出に無関係な単語については，事前に削除した．削除した項目は以下のとおりである．

(1) URL

(2) 改行文字

(3) 絵文字

(4) Twitter に定型でよく現れる単語

「RT」「まとめ」「お気に入り」

#### 4.2.4 分かち書き

今回の検出対象が隠語であるため，中には造語に近いものもあることが考えられ，また中には，複合語型隠語は一般的な単語の組み合わされたものがあることから，複合語型隠語より短い文節で分かち書きされ，正しく分かち書きされないおそれがある．そのため，4.2.6 項による処理により，分かち書き用の内部辞書に複合語を登録すれば，複合語型隠語がより短い文節で分かち書きされないことが期待できる．

なお，本研究では形態素解析器として Sudachi [8] を用いている．

#### 4.2.5 単語分散表現モデルの構築

Word2vec を用いて実行した．複合語検出用の Word2vec のパラメータは**表 2** のとおり．パラメータのうち Window Size については，複合語という隣り合う単語を検出するこ

表 2 複合語検出用の Word2vec のパラメータ
**Table 2** Parameters of Word2vec for compound word detection.

| パラメータ項目 | 設定値 |
|---|---|
| Size | 300 |
| Min-Count | 3 |
| Window Size | 2 |
| Negative | 20 |
| 手法 | Skip-Gram [20] |

とから 2 としている．なお，予備実験として Window Size を 2 から 4 まで変更し実験を行ったところ，2 が一番検出精度が良かった．

#### 4.2.6 複合語の検出と辞書登録

(1) コーパスから複合語の検出
構築した単語分散表現モデル内の単語群を抽出し，3.2 節で示したアルゴリズムで作成した複合語検出プログラムを実行し，複合語候補を作成する．なお，アルゴリズムの上位 $M$ 位の $M$ の値については，$M$ の値を大きくすれば，照合する対象の単語が増えるが，一方で再帰処理となるため，処理時間が大幅に増えることとなる．そこで，本実験では，予備実験の結果をふまえ，$M = 5$ とした．

(2) 出現回数の確認
分かち書き前の元のコーパスの文章群において，できる限り多くの単語が隠語である可能性に鑑み $\mathcal{X} = 2$ 回以上出現した単語をピックアップする．

(3) 形態素解析器のユーザー辞書に単語を登録
分かち書き用の形態素解析器のユーザー辞書に検出した複合語を登録することで，当該複合語が分かち書き時に文節が分かれなくなる．

なお，本実験では，2 語が結合する複合語だけでなく，3 語以上から構成される複合語検出を行うために複合語検出と辞書登録を繰り返し実行する．そのため，4.2.4 項から 4.2.6 項までの処理を 10 回繰り返した．

### 4.3 実験条件

提案手法に加え，以下の 3 つの手法で比較評価を実施した．

#### 4.3.1 ベースライン手法（Bi-gram 生成）

提案手法による効果を検証するため，分かち書き後の文章について，すべてのバイグラムを取得したものをベースライン手法として用意した．たとえば，「バブル OG 入り ました」という分かち書き後の文章の場合，「バブル·OG」,「OG·入り」,「入り·まし」,「まし·た」となる．提案手法としては，前後を入れ替えたものも複合語候補として用意したことから，上記の単語についても前後を入れ替えたものも用意した．

本手法で Bad コーパスから複合語候補を作成したところ，686,561 語が作成され，提案手法に合わせて，前処理前の Bad コーパスから参照した際の出現頻度の回数の閾値を 4.2.6 項の考えを踏襲し，$\mathcal{X} = 2$ 回としたところ，105,266 語となり，他の条件に比べ膨大な数となり，アノテーションが困難な数となった．そのため，ベースラインの精度が最も高くなるように調整を行った結果，出現頻度の閾値を 14 回に緩和し，対象単語数を 14,218 語と選定した．

表 3 複合語候補の分類
**Table 3** Classification of compound word candidates.

| 手法 | 複合語候補 | | 複合語 | Precision[b] | Recall (max) | F1-Score (max) |
|---|---|---|---|---|---|---|
| | 閾値設定適用前 | 閾値設定適用後 | | | | |
| 提案手法[a] | 61,731 | 295 | 264 | 0.895 | 0.357 | 0.510 |
| 提案手法（機能追加なし）[a] | 101,573 | 521 | 308 | 0.591 | 0.416 | 0.489 |
| BL 手法 | 686,561 | 14,218 | 388 | 0.027 | 0.524 | 0.052 |
| Small ら [21] | - | 453 | 72 | 0.159 | 0.097 | 0.121 |

[a] 4.2.4 項から 4.2.6 項までのプロセスの 10 回の試行の合計値
[b] 複合語 / 複合語候補数（閾値設定適用後）

#### 4.3.2 Small らの手法（N-gram 生成）

Small らによるフレーズ生成方法を参考に，比較手法として用意した [21]．

具体的な手法としては，以下の流れで N-gram の複合語を生成した．

(1) 用意した文章群から N = 2〜8 までの N-gram を生成する
(2) 生成したすべての N-gram について元の文章における出現頻度をカウントする
(3) 各 N-gram は出現頻度が十分に高ければ，主フレーズと見なす．

これに基づき，Bad コーパスから N-gram（2〜8）を生成し，それを元の Bad コーパスにおける出現回数を確認し，出現回数上位 100 位までの単語のうち，複合語として成立しているかどうかで評価した．その結果，それぞれ 100 位まで抽出の後に，重複を除外したところ，453 語が候補となり，それぞれ確認したところ，72 語が複合語として成立しており，割合としては 15.9%であった．

#### 4.3.3 機能追加なし条件

提案手法から，3.3 節の 4 つの機能を除いたものを比較手法として用意した．精度向上機能により精度が向上したと期待できる提案手法の比較手法として用意した．

### 4.4 実験結果（複合語検出実験）

3 つの手法で複合語候補を検出し，複合語であるか否かの 2 つに分類した．

複合語検出の結果は**表 3** のとおりであった．

なお，複合語候補を抽出するための出現回数条件（3.2 節 (3)）については，提案手法は $\mathcal{X} = 2$，ベースラインは 4.3.1 項でも説明したとおり，$\mathcal{X} = 14$ とした．

表 3 より，提案手法はベースライン手法，既存手法に比べ，複合語検出の精度において大きく上回る結果が得られた．さらには，提案手法の中でも機能追加しなかった

場合に比べ，提案手法の方が 30.1%ポイント上回る結果となった．

対象の単語ペアが 686,561 単語ペアと膨大であるため，これらすべてに対してアノテーションを行うことが困難であった．そのため，Recall および F1-Score を導出することができなかった．したがって以下のように Recall および F1-Score についての考察を行った．実験に用いた 4 手法で正しく検出できた複合語は重複を除くと 740 個であった．仮にこれ以外に複合語が存在しないと仮定すると，利用したデータセットにおける Recall と F1-Score の理論上の最大値を算出できる．この結果を表 3 の “Recall (max)” と “F1-Score (max)” に示す．提案手法の F1-Score は 0.510 であり，他の手法の F1-Score を大幅に上回っていることが分かる．なお，ベースライン手法（閾値設定前）については，アノテーションが困難であるため Precision を算出することはできないが，Precision はベースライン手法（閾値設定後）を下回ると想定される．しかし，ベースライン手法（閾値設定後）と同じ 0.027 であると想定し，Recall を最大の 1.00 とすると，F1-Score は 0.053 である．やはり提案手法の F1-Score が大幅に上回っていることが分かる．当然，740 個以外にも複合語は存在する可能性がある．データセットに含まれる複合語の数の想定値を横軸にとり，縦軸にその際の各手法の F1-Score をとるグラフを**図 5** に示す．データセットに含まれる複合語の数が多いほど提案手法とベースラインや Small らの手法との差は縮まるが，複合語の数が 5,000 程度であっても提案手法の F1-Score は 0.100 程度であり，これらの手法を大きく上回っていることが確認できる．なお，提案手法と提案手法（追加機能なし）については，複合語の数が 1,000 程度で F1-Score の大小が入れ替わるが，その差はわずかである．

### 4.5 考察（複合語検出実験）

表 3 より，複合語の検出精度については，ベースライン手法に比べ提案手法が大きく上回っていたことから，提案手法は複合語検出に有効な手法であるといえる．さらに提案手法と追加した機能を除いた条件と比較したところ，提案手法の方が 30.1%ポイントも上回る結果が得られたことから，複合語検出については追加した機能が有効に機能していたといえる．

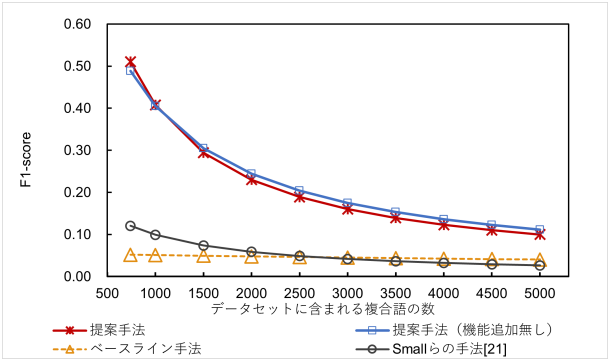


**図 5** データセットに含まれる複合語の数の想定値と F1-Score の関係

**Fig. 5** Relationship between the F1-Score and the assumed number of compound words in the data-set.

## 5. 複合語型隠語検出実験

### 5.1 実験の概要

コーパス内にある単語群を用いて，隠語，さらには複合型隠語を検出する実験を行った．具体的には，21,220 語に含まれるうち，照合用の既知隠語として，既知の隠語のうちランダムに選択した 18 語を用意し隠語の検出を目指す．

### 5.2 実験環境

複合語検出実験において用いた提案手法により複合語辞書を作成し，既存手法である隠語検出プログラムを実行した．

### 5.3 実験のプロセス（複合語型隠語検出実験）

実験において図 3 の流れで処理を実施する．データ収集，コーパス作成，前処理については 4.2.1〜4.2.2 項で述べた処理を行う．続いて，複合語を分かち書き用の内部辞書に追加する処理として，4.2.4 項から 4.2.6 項までのプロセスを実施し，検出されたすべての複合語を内部辞書へ追加する．本実験では，4.4 節で検出された 264 語を登録する．

#### 5.3.1 分かち書き

分かち書き用の内部辞書に複合語を追加登録後，これに基づき分かち書きを行う．これにより，既存手法では分割されていた複合語型隠語が分割されないことが期待できる．

なお，本研究では形態素解析器として，既存手法と同様に，Sudachi [8] を用いる．

#### 5.3.2 単語分散表現処理

形態素解析処理の実施後，Word2vec を用いて，単語分散表現処理を実施した．

パラメータは**表 4** のとおり設定した．パラメータのうち Window Size については，複合語検出の表 2 と異なり，予備実験の結果から 4 としている．

#### 5.3.3 隠語検出プログラムの実行

既存手法 [7] のプログラムを実行する．

**表 4** Word2vec のパラメータ
**Table 4** Parameters of Word2vec.

| パラメータ項目 | 設定値 |
|---|---|
| Size | 300 |
| Min-Count | 3 |
| Window Size | 4 |
| 手法 | Skip-Gram [20] |

表 5 実験結果
Table 5 Classification results.

| 評価手法 | 提案手法 | | 既存手法 [7]（複合語処理なし） | |
|---|---|---|---|---|
| | 個数 | 割合 | 個数 | 割合 |
| 隠語 | 73 | 64.0% | 53 | 57.0% |
| 犯罪関連語 | 20 | 17.5% | 15 | 16.1% |
| 複合語型隠語の一部 | 14 | 12.2% | 18 | 19.4% |
| 無関係 | 7 | 6.1% | 7 | 7.5% |
| 合計 | 114 | | 93 | |

隠語検出プログラムについては，単語分散表現モデルから，両コーパスで共通して出現する単語および Bad コーパスのみに出現する単語を抽出して隠語かどうか判定する単語群とする仕組みとしているところ，本実験において両コーパスで共通して出現した単語数は 19,068 語であり，Bad コーパスのみに出現した単語数は 2,152 語の合計 21,220 語を対象とした．

なお，提案手法において事前に調整する必要があるパラメータは，隠語判定の閾値である．予備実験を行い，一番良い値を閾値として決定した．既存手法におけるパラメータの調整についても同様に実施した．

### 5.4 実験結果（複合語型隠語検出実験）

実験の結果，隠語候補として検出された単語は 114 語であり，分類結果は**表 5** のとおりであった．

既存手法 [7] ではこれまで検出できなかった複合語型隠語を今回の提案手法により確認することができた．このうち，たとえば，大麻を指すパイナップルチャンク，ジャックヘラー，ブルードリームなどといった複合語型隠語が 10 語含まれていた．

また Precision についても既存手法に比べ，7.0%ポイント精度の良い結果が得られた．また既存手法に比べ，精度だけでなく検出した隠語の数についても 21 個多く検出できた．

## 6. 考察

### 6.1 複合語としては検出されなかった単一単語について

隠語として検出された単語で誤検出として分類したもののうち，表 5 の「複合語型隠語の一部」と分類したものにあっては，悪意のある文脈で使用された単語ではあったものの分かち書き時に文節が切れてしまい，一語では隠語としての意味を持たないものであった．例として，「ビック」（「ビックバッツ」，「ビックバット」などを確認）や「スーパー」（「スーパーレモンヘイズ」，「スーパーレモンスカンク」などを確認）などが該当する．これらは複合語として検出される際に，たとえば「スーパー」は一般的な他の単語との出現頻度が高かったため，類似語上位の単語も隠語と無関係の単語が出現し，類似語に基づく提案手法では複合語として検出されなかったものと考えられる．そのため，一般的な単語との出現頻度の高い単語が含まれた複合語を検出する方法について今後の検討課題とする．

それ以外にも，「ドクター」（「ドクターグリーンスプーン」や「ドクタージャマイカ」などを確認）など，複合語隠語の 1 文節が隠語として出現しており，複合語型隠語として検出できなかった単語が確認されたことから，これらについても引き続き検討を行う．

表 6 3 連複合語の出現数の結果
Table 6 Number of occurrences of triple compound words.

| 試行回数 | 登録した複合語の数 | 3 連複合語の数 | 割合 |
|---|---|---|---|
| 1 回目 | 159 | 0 | 0.0% |
| 2 回目 | 40 | 8 | 20.0% |
| 3 回目 | 18 | 1 | 5.6% |
| 4 回目 | 17 | 0 | 0.0% |
| 5 回目 | 12 | 0 | 0.0% |
| 6 回目 | 11 | 0 | 0.0% |
| 7 回目 | 11 | 1 | 9.1% |
| 8 回目 | 9 | 0 | 0.0% |
| 9 回目 | 6 | 0 | 0.0% |
| 10 回目 | 12 | 0 | 0.0% |
| 合計 | 295 | 10 | 3.4% |

### 6.2 3 単語以上が結合する複合語について

本手法の 10 回の試行の中で，辞書に登録した単語に基づき，さらに単語どうしが結合した，いわゆる 3 連複合語が検出されていたことも分かった．10 回の試行の結果は**表 6** のとおり．

具体的に 3 連複合語は，10 語出現していたがその中に隠語は含まれていなかった．また 4 連以上の複合語は検出されなかった．これは 4 つ以上の単一単語が結合した複合語型隠語があったとしても非常に数が少ないことが考えられる．

### 6.3 未知の隠語の検出について

本提案手法により検出できた未知の隠語が，用意した照合用の隠語リスト以外の隠語という意味での未知の隠語ではなく，本当の意味での未知の隠語であったのかについて，警察官にヒアリングを実施し，確認を行った．具体的には，5.4 節で検出した 73 語の隠語について，「暴力団員による不当な行為の防止一般に関すること」をつかさどる警察庁刑事局組織犯罪対策部組織犯罪対策第一課に所属の捜査経験豊富な 40 代から 50 代の現役警察官 7 名に，「認知している」，「認知していない」の 2 択で評価を依頼した．なお，本調査結果の公表については，同意済みである．また評価対象の隠語については，実際のツイートとともに確認を依頼したところ，隠語との判断についての異論はなかった．そして，7 人中何人が認知しているかどうかで未知かどう

表 7 ヒアリング結果（未知の隠語か否か）
Table 7 Interview (verification of whether it is an unknown dark jargon or not).

| 未知率 | 個数 | 割合 |
|---|---|---|
| 100% | 45 | 61.6% |
| 85.7% | 15 | 20.5% |
| 71.4% | 6 | 8.2% |
| 57.1% | 2 | 2.7% |
| 42.9% | 1 | 1.4% |
| 28.6% | 2 | 2.7% |
| 14.3% | 1 | 1.4% |
| 0% | 1 | 1.4% |
| 合計 | 73 | |

隠語評価に用いた単語はすべて隠語であった

かの割合を求めた．つまり，7 人中 7 人が認知していない単語は未知率 100%ということとなる．

評価結果は，**表 7** のとおりである．ヒアリングの結果，過半数が認知していない単語，すなわち未知率が 57.1%以上である単語は合計で 68 語であり，93.2%の単語が認知されていない，すなわち未知の隠語であったといえる．それ以外の単語，すなわち比較的認知されていた隠語としては，キメセク，氷，クッキー，野菜，キャンディーといった単語であった．

このようなことから，本提案手法を用いることで未知の隠語を検出できるといえることが分かった．

### 6.4 閾値について

閾値をどう設定するかによって精度（Precision および Recall）は変化する．Precision と Recall はトレードオフの関係にあるため，どちらをより重視するかによっても設定すべき閾値は変化していく．たとえば，事前に目標となる Recall を 60%と設定したとする．事前に既知の隠語を 5 個用意し，このうち何個検出できるか見つけられるかについて予備実験を行い，5 個中 3 個見つけたときの設定値を閾値とすることができる．Precision についても同様の方法で閾値を設定することが可能である．このように，状況によって，Recall を重視したり，Precision を重視したり，フレキシブルに閾値を変更できる仕組みとなっている．

### 6.5 制限事項

予備実験において，対象とした範囲のツイートでは絵文字の効果が薄かったことから，前処理において削除することとした．ただし，範囲や使用する SNS などによっては，絵文字も効果がある可能性はある．絵文字を 1 つの単語と見なすことで，提案手法をそのまま適用できると考えられる．つまり，複数の絵文字によって表される 1 つの言葉があれば，それについても複合語として登録できると考えられる．絵文字を削除せずに提案手法を適用する実験の実施は将来課題とする．

同じく予備実験において，複合語候補作成時に，名詞句に限定することで無関係な単語を大幅に削減でき，精度向上に大きく貢献できたことから，複合語候補作成時に品詞分類を行い，名詞句以外を除外する機能を導入した．また隠語検出手法 [7] についても，名詞句に限定した品詞分類機能を導入することで精度をあげたことを報告している．一方で，複合語候補作成時に名詞句以外を除外することで，名詞句以外で構成させる複合語型隠語を検出できないリスクがある．そのため，精度を維持したまま名詞句以外の複合語型隠語の検出については，将来課題とする．

また提案手法は，複合語型隠語が同じ文脈・文章で出現する頻度が高いことを想定した手法である．複合語型隠語は単語どうしの関連性が強いと考えられることから，複合語の分節で区切られた単語どうしの類似語として，互いの単語が上位に出現することを利用している．マイクロブログ以外の文章についても，提案手法の前提が成立すると考えられることから，提案手法を，適用できると考えられるが，実際に掲示板などの文章に対しての検証実験の実施は将来課題とする．

また，ELMo [22] や BERT [23] のような文脈に基づく単語埋め込みモデルは，文脈に応じて異なる単語の分散表現を生成できる．たとえば 1 万件のツイートに出現する単語 A は，それぞれの文脈に応じて 1 万通りの分散表現が生成される．一方，Word2Vec のような静的な単語埋め込みモデルは，文脈に関係なく単語 A に対して 1 つの固定された分散表現が構築される．TinyBERT [24] のように軽量なモデルもあるが，文脈を考慮するモデルを利用する際には，ツイートごとに異なる分散表現を高速に取り扱う課題が生じる．ELMo や BERT を採用することでより良い結果が得られる可能性もあり，その場合は提案手法の優位性はさらに高まる．単語分散表現モデルの選択は将来課題とする．

## 7. 関連研究

自然言語処理（NLP）の分野において，近年，大規模言語モデル（LLM）を利用した手法などが取り組まれており，BERT，ChatGPT など様々な技術とともに実用性について報告されている．

Web 上で非公式な辞書として隠語を説明しているページが多数あれば，LLM はその単語を学習することができるため，LLM を用いてそのような隠語を検出できる可能性がある．しかし，隠語が SNS 上などで説明なく利用されているだけの場合は，それがどのような意味を持つかを LLM では知ることはできない．また，ChatGPT（2023 年 5 月現在）は 2021 年 9 月までのデータしか訓練に利用されておらず，1，2 年のタイムラグがある．そのため最新の隠語に対応できない．実際，「ホワイトクッシュ」や「円光」など，犯罪者の間では一般的になりつつある隠語に関

しては，ChatGPT は正しくその意味を解説することができた．しかし，「受け子」など新しい言葉については，性的な意味を持つスラングとして解説を行い，詐欺に関する意味を把握していない回答が得られた（質問内容（プロンプト）を工夫することにより，適切な回答を得られる可能性はある．しかし，何が隠語が分からないシナリオで，どのような犯罪に関連しているかも不明な状況では，適切なプロンプトを生成すること自体が困難であると思われる）．

近年の ChatGPT などの大規模言語モデル（LLM）は必要に応じてウェブから最新情報を引用できる．そのため，LLM の精度が向上した場合，LLM を使って隠語の有無を判定する使い方が想定される．しかしすべてのツイートに対して LLM を用いて隠語の有無を判定することはコストが非常に大きい．また，LLM は逐次学習が困難であり，リリースまでに現状は数カ月を要する．一方で提案手法の学習コストは LLM と比べると低く，任意のタイミングで実行することが可能である．また，LLM の補助的な使い方として，提案手法で抽出された単語が隠語であるかどうか，また隠語である場合はその意味は何かを確認する使い方が想定される．引き続き，LLM と提案手法の連携については検討していく必要がある．

このような隠語検出の分野において，これまでにも掲示板などのウェブサイトを対象とした隠語や煽り用語などの検出などに関する研究は，いくつか報告されている [4], [6], [25], [26], [27].

ただし，これらの研究は係り受けができることが前提であり，1 つの投稿につき文字数が短く限定される Twitter などの短文では係り受けがない場合が多く，そのままでの対応は難しい [27]．また単語が意図したとおりに分かち書きされることを前提としており，主に造語であったり認知度が低い単語であったりする複合語型隠語には，そもそも正しく分かち書きできないことが想定され対応できない．また Yuan らは，ダークウェブ上ではポップコーンやブルーベリーの名で大麻がやり取りされていたり，チーズピザという名でチャイルドポルノがやり取りされていたりすることから，ダークウェブから自動的に「隠語」を識別する手法について提唱している [3]．その際，Word2vec [18] による単一のコーパスでは，隠語が発見できないとのことから，複数のコーパスを用意し，そのうちの 2 つの異なるコーパスに現れる用語の意味的な矛盾から隠語を検出している．ただし，コーパス間で差が開けば隠語判定としているだけであるが，我々の手法は，品詞分類も取り入れ，精度を高めるだけでなく，再帰的な処理も実施している点が大きな差異と考える．

それ以外のプラットフォームを対象とした隠語検出の研究についても，いくつか報告されている [28], [29]．たとえば，安彦らは，短文で文脈性のない ID 掲示板を対象に，テキスト分類（教師あり学習）を用いて有害性を分類している [29]．ただし，本報告の中では，F 値などで有意差が確認できなかったことと，「野菜」,「アイス」などダブルミーニングなものへの対応が難しいことと，そして ID 掲示板については，違法な行為を目的とした投稿が多い一方，Twitter については，以前に調査した結果，一般的な投稿の方が圧倒的に多かったことから [7]，そういった点で ID 掲示板とは大きく異なる．

それ以外にも，隠語などの特定の目的の単語検出を目的とした研究とは異なり，隠語を認識している前提で投稿や文章が有害かどうか判定している研究もある [29], [30], [31]．これらの研究については，まず隠語を認識する必要があることから，我々の手法と組み合わせることでより相乗効果が期待できる．

ここまでの研究は，適切に分かち書きされることを前提に隠語の検出や有害な投稿の分類などが行われてきているが，フレーズ，すなわち 2 語以上で構成される単語の検出を目的とした研究もある．

たとえば，Zhu ら [32] は，単一単語の隠語検出はあるものの，「blue dream」（マリファナ）や「black tar」（ヘロイン）などの複数単語（multi-word euphemisms）の隠語を自動的に検出できる既存の研究はないと述べており，自動的に隠語を検出することを目指している．さらには，未知の隠語も検出したと報告している．ただし，Zhu らの手法により，隠語を検出するためには，良質なフレーズコーパスが必要であり，整然とした文章がある前提の手法であるため，Twitter のような誤字なども多い，単語の羅列も含まれる短文では適用できないと考える．

これに関連して，キーフレーズ抽出は，自然言語処理における基本的なタスクであり，文書を代表的なフレーズにマッピングすることを容易にするものでありこれまでにも様々な研究が進められてきている [33], [34].

たとえば，Small らは論文の意味把握を課題としており，論文のアブストラクトの中から，論文における主要なフレーズを抽出し，意味把握することを目指している [21]．このように，英語であっても 2 語以上で構成させる単語の検出に関心が高まっているといえる．また日本語においても，キーフレーズ抽出については，研究されている [35].

ただし，キーフレーズ抽出が対象とするものは，論文のアブストラクトやレビューなど，ある程度まとまった文章や文法的にも正しく記載されているものを対象とし，その中からキーとなるフレーズを抽出するものが多い．しかしながら，本研究が対象とする Twitter は，短文でかつ文法が構造化されておらず，誤字も多いことから [36], [37]，キーフレーズ抽出は難しく，また特に本研究で対象とするような隠語は出現頻度も低く，キーフレーズとなっていない可能性も高い．

このようなことから，これらの複合語についての手法では，隠語を含む複合語の検出は難しい．そのため，隠語を

対象に複合語型隠語を検出することは非常に有意義である.

## 8. おわりに

サイバーパトロールを支援するため，隠語を検出することを目的として，複合語を検出する手法を提案し，複合語辞書を作成したうえで隠語検出実験を実施し，隠語，特に複合語型隠語の検出を目指した．実験の結果，既存手法に比べ Precision が向上しただけでなく，既存手法で検出できなかった複合語型隠語を 10 語検出することができた．さらに，7 名の現役警察官へのヒアリングにより，実験で検出した隠語のうち 93.2%が過半数の警察官にとって未知のものであることが確認できた.

これらのことから，本提案手法を用いることでより効率的に隠語を検出することが期待できる.

**謝辞** 本研究は JSPS 科研費 JP21H03496, JP22K12157, JP23H03688 および公益財団法人住友電工グループ社会貢献基金の助成を受けたものです.

### 参考文献

[1] Mihalcea, R. and Nastase, V.: Word epoch disambiguation: Finding how words change over time, *Proc. 50th Annual Meeting of the Association for Computational Linguistics* (*Volume 2: Short Papers*), pp.259–263, Association for Computational Linguistics (2012).

[2] Wijaya, D. and Yeniterzi, R.: Understanding semantic change of words over centuries (2011).

[3] Yuan, K., Lu, H., Liao, X. and Wang, X.F.: Reading thieves' cant: Automatically identifying and understanding dark jargons from cybercrime marketplaces, *27th USENIX Security Symposium* (*USENIX Security 18*), pp.1027–1041, USENIX Association (2018).

[4] 大西 洋，田島敬史：語の出現の偏りに基づく新たな隠語の発見，*DBSJ Journal*, pp.103–108 (2013).

[5] 青木竜哉，笹野遼平，高村大也，奥村 学：ソーシャルメディアにおける単語の一般的ではない用法の検出，自然言語処理，Vol.26, No.2, pp.381–406 (2019).

[6] 三谷亮介，小野 守，松本裕治，隅田飛鳥，服部 元，小野智弘：有害性スコアリングによる web テキストにおける隠語の発見，言語処理学会第 19 回年次大会，pp.461–464 (2013).

[7] 羽田拓朗，清 雄一，田原康之，大須賀昭彦：コーパス間の類似語の差異に着目したマイクロブログにおける隠語検出，電気学会論文誌 C（電子・情報・システム部門誌），Vol.142, No.2, pp.177–189 (2022).

[8] Takaoka, K., Hisamoto, S., Kawahara, N., Sakamoto, M., Uchida, Y. and Matsumoto, Y.: Sudachi: A Japanese tokenizer for business, *Proc. 11th International Conference on Language Resources and Evaluation* (*LREC 2018*), European Language Resources Association (ELRA) (2018).

[9] 羽田拓朗，清 雄一，田原康之，大須賀昭彦：類似語を利用した複合語型隠語の検出，信学技報，Vol.122, No.186, pp.58–63, (2022).

[10] Hada, T., Sei, Y., Tahara, Y. and Ohsuga, A.: Detection of compound-type dark jargons using similar words, Rocha, A.P., Steels, L. and van den Herik, H.J. (Eds.), *Proc. 15th International Conference on Agents and Artificial Intelligence*, *CAART 2023*, Vol.1, pp.427–437, SCITEPRESS (2023).

[11] 第五次薬物乱用防止五か年戦略（平成 30 年 8 月），入手先 ⟨https://www.mhlw.go.jp/content/11126000/000341876.pdf⟩.

[12] NHK：広域強盗 東京 稲城の事件で幹部 4 人再逮捕 主要 8 事件で立件（2023 年 12 月 5 日），入手先 ⟨https://www3.nhk.or.jp/news/html/20231205/k10014278391000.html⟩ (2023).

[13] 犯罪対策閣僚会議：SNS で実行犯を募集する手口による強盗や特殊詐欺事案に関する緊急対策プラン（令和 5 年 3 月 15 日）(2023)，入手先 ⟨https://www.kantei.go.jp/jp/singi/hanzai/kettei/230317/honbun-1.pdf⟩.

[14] インターネット・ホットラインセンター：令和 5 年におけるインターネット・ホットラインセンターの運用状況について（令和 6 年 3 月 14 日）(2023)，入手先 ⟨https://www.internethotline.jp/file_preview/InfoContents/956/file/⟩.

[15] 第六次薬物乱用防止五か年戦略（令和 5 年 8 月），入手先 ⟨https://www.mhlw.go.jp/content/11120000/001237115.pdf⟩.

[16] Mangnejo, J., Khuhawar, A., Kartio, M. and Soomro, S.: Inherent flaws in login systems of facebook and twitter with mobile numbers, *Annals of Emerging Technologies in Computing*, Vol.2, pp.53–61 (2018).

[17] Hada, T., Sei, Y., Tahara, Y. and Ohsuga, A.: Codeword detection, focusing on differences in similar words between two corpora of microblogs, *Annals of Emerging Technologies in Computing* (*AETiC*), Print ISSN: 2516-0281, Online ISSN: 2516-029X, Vol.5, No.2, pp.1–7 (2021).

[18] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient estimation of word representations in vector space, Bengio, Y. and LeCun, Y. (Eds.), *1st International Conference on Learning Representations*, *ICLR 2013*, *Workshop Track Proceedings* (2013).

[19] Mizuki, S., Matsuno, S. and Sakaki, T.: Constructing of the word embedding model by Japanese large scale SNS + web corpus, *The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019*, Vol.JSAI2019, pp.4Rin113–4Rin113 (2019).

[20] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J.: Distributed representations of words and phrases and their compositionality, *CoRR*, Vol.abs/1310.4546 (2013).

[21] Small, E. and Cabrera, J.: Principal phrase mining (2022).

[22] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L.: Deep contextualized word representations, *CoRR*, Vol.abs/1802.05365 (2018).

[23] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding, *CoRR*, Vol.abs/1810.04805 (2018).

[24] Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., Wang, F. and Liu, Q.: TinyBERT: Distilling BERT for natural language understanding, *CoRR*, Vol.abs/1909.10351 (2019).

[25] Lee, W., Lee, S.S., Chung, S. and An, D.: Harmful contents classification using the harmful word filtering and SVM, Shi, Y., van Albada, G.D., Dongarra, J. and Sloot, P.M.A. (Eds.), *Computational Science – ICCS 2007*, pp.18–25, Springer Berlin Heidelberg (2007).

[26] 松本典久，上野 史，太田 学：BERT を利用した煽りツイート検出の一手法，第 13 回データ工学と情報マネジ

メントに関するフォーラム（DEIM2021）論文集，I14-2 (2021).
[27] 橋本広美，木下嵩基，原田 実：フィルタリングのための隠語の有害語意検出機能の意味解析システム sage への組み込み，情報処理学会研究報告：自然言語処理研究会報告，Vol.196, pp.N1–N6 (2010).
[28] Seyler, D., Liu, W., Wang, X.F. and Zhai, C.X.: Towards dark jargon interpretation in underground forums, Hiemstra, D., Moens, M.-F., Mothe, J., Perego, R., Potthast, M. and Sebastiani, F. (Eds.), *Advances in Information Retrieval*, pp.393–400, Springer International Publishing (2021).
[29] 安彦智史，長谷川大，プタシンスキミハウ，中村健二，佐久田博司：Id 交換掲示板における書きこみの隠語表記揺れを考慮した有害性評価，情報システム学会誌，Vol.13, No.2, pp.41–58 (2018).
[30] 住田 淳，亮 隆弘，菱田隆彰：児童被害を抑止するための SNS 上の不正コメント抽出方法，第 80 回全国大会講演論文集，Vol.2018, No.1, pp.117–118 (2018).
[31] Zhao, K., Zhang, Y., Xing, C., Li, W. and Chen, H.: Chinese underground market jargon analysis based on unsupervised learning, *2016 IEEE Conference on Intelligence and Security Informatics* (*ISI*), pp.97–102 (2016).
[32] Zhu, W. and Bhat, S.: Euphemistic phrase detection by masked language model (2021).
[33] Shang, J., Liu, J., Jiang, M., Ren, X., Voss, C.R. and Han, J.: Automated phrase mining from massive text corpora (2017).
[34] Liang, X., Wu, S. Li, M. and Li, Z.: Unsupervised keyphrase extraction by jointly modeling local and global context, *CoRR*, Vol.abs/2109.07293 (2021).
[35] 木村優介，楠 和馬，寺本優香，波多野賢治：単語埋め込みと名詞句の共起グラフを用いた教師なしキーフレーズ抽出手法の提案，Technical Report 2，同志社大学大学院文化情報学研究科，同志社大学文化遺産情報科学調査研究センター，同志社大学文化情報学部 (2020).
[36] Rosa, K.D. and Ellen, J.: Text classification methodologies applied to micro-text in military chat, pp.710–714 (2009).
[37] Sahinuç, F. and Toraman, C.: Tweet length matters: A comparative analysis on topic detection in microblogs, Hiemstra, D., Moens, M.-F., Mothe, J., Perego, R., Potthast, M. and Sebastiani, F. (Eds.), *Advances in Information Retrieval - 43rd European Conference on IR Research*, *ECIR 2021*, *Virtual Event, March 28 - April 1, 2021*, *Proceedings, Part II*, Lecture Notes in Computer Science, Vol.12657, pp.471–478, Springer (2021).

**羽田 拓朗** （学生会員）

1980 年生．2004 年静岡大学大学院情報学研究科情報学専攻博士前期課程修了．2021 年電気通信大学大学院情報理工学研究科情報学専攻博士前期課程修了．2024 年同博士後期課程修了．博士（工学）．2008 年より警察庁中部管区警察局．現在，警察庁刑事局組織犯罪対策部組織犯罪対策第一課および電気通信大学協力研究員．主として自然言語処理に関連した機械学習の研究等犯罪軽減に向けた研究に従事．

**清 雄一** （正会員）

1981 年生．2009 年東京大学大学院情報理工学系研究科博士後期課程修了．同年（株）三菱総合研究所入社．2013 年より電気通信大学．現在，同大学大学院情報理工学研究科教授．博士（情報理工学）．エージェント，プライバシ保護技術等の研究に従事．2016 年度土木学会水工学論文賞，情報処理学会論文賞，2021 年 IPSJ/IEEE Computer Society Young Computer Researcher Award，2022 年度日本冷凍空調学会学術賞受賞．電子情報通信学会，日本ソフトウェア科学会，IEEE Computer Society 各会員．

**田原 康之** （正会員）

1966 年生．1991 年東京大学大学院理学系研究科数学専攻修士課程修了．同年（株）東芝入社．1993〜1996 年情報処理振興事業協会に出向．1996〜1997 年英国 City 大学客員研究員．1997〜1998 年英国 Imperial College 客員研究員．2003 年国立情報学研究所着任．2008 年より電気通信大学准教授．博士（情報科学）（早稲田大学）．エージェント技術，およびソフトウェア工学等の研究に従事．2016 年度日本ソフトウェア科学会解説論文賞，2022 年度情報処理学会卓越研究賞（SE 研究会）受賞．電気学会，日本ソフトウェア科学会各会員．

**大須賀 昭彦** （正会員）

1958 年生．1981 年上智大学理工学部数学科卒業．同年（株）東芝入社．同社研究開発センター，ソフトウェア技術センター等に所属．1985〜1989 年（財）新世代コンピュータ技術開発機構（ICOT）出向．2007 年電気通信大学大学院情報システム学研究科教授．2016 年同大学大学院情報理工学研究科教授．2024 年より同大学産学官連携センター長・特任教授．博士（工学）（早稲田大学）．電子情報通信学会フェロー．電気通信大学名誉教授．ソフトウェア工学，エージェント，人工知能の研究に従事．1986 年度情報処理学会論文賞，2013 年度人工知能学会研究会優秀賞，2014 年度同学会功労賞，2016 年度情報処理学会論文賞，2018 年度電子情報通信学会 ISS 活動功労賞，2022 年度情報処理学会卓越研究賞（SE 研究会）受賞．IEEE Computer Society Japan Chapter Chair，人工知能学会理事，日本ソフトウェア科学会理事，同学会監事，同学会評議員等を歴任．電子情報通信学会，人工知能学会，日本ソフトウェア科学会，IEEE Computer Society 各会員．本会フェロー．