マイクロブログにおいて
犯罪を誘導する隠語の検出手法の提案

羽田　拓朗

電気通信大学大学院情報理工学研究科

博士（工学）の学位申請論文

2024年3月

マイクロブログにおいて
犯罪を誘導する隠語の検出手法の提案

博士論文審査委員会

主査　　清　雄一　教授

委員　　大須賀　昭彦　教授

委員　　柏原　昭博　教授

委員　　内海　彰　教授

委員　　稲葉　通将　准教授

A Proposed Method for Detecting Crime-Inducing Dark Jargons in Microblogs

Takuro Hada

Abstract

In recent years, the number of drug trafficking using microblogs has been increasing, which has become a social problem. While cyber patrols have been conducted to crack down on such crimes, those who post crime-inducing messages use terms that camouflage their criminal intentions so-called “dark jargons” to avoid keywords such as “enjo kosai,” “marijuana,” and “methamphetamine” that may be monitored and attract police attention. These dark jargons change once they become popular, so it is always necessary to keep track of the latest dark jargons. Therefore, we propose a new method for detecting the latest dark jargons. In this paper, we offer a new way of detecting code words from the differences in the words used in posts to detect dark jargons used in a crime. Specifically, we propose a new method in which we divide words into two corpora, depending on whether a post containing a word has a criminal intention and detect dark jargons from the differences between similar words of the same word between two corpora. To confirm the effectiveness of the proposed method, we conducted an experiment to detect dark jargons. The experimental-1 results showed that the proposed method was able to detect dark jargons with higher accuracy than that of the baseline method. Moreover, by using word associations, we propose a method for detecting compound-type dark jargons that combines two or more words, which have been difficult to detect using only the proposed methods. To confirm the effectiveness of the proposed method, we conducted a detection experiment with compound words and a detection experiment with dark jargons. As a result, we confirmed that the proposed method enabled to detect compound-type dark jargons that could not be detected by existing methods. The experiment shows that the proposed method can reduce the burden of continuously monitoring code words by rapidly

and automatically detecting new dark jargons that change with time; thus, it provides the possibility of showing clues for crimes.

マイクロブログにおいて
犯罪を誘導する隠語の検出手法の提案

羽田　拓朗

概要

近年，マイクロブログにおける違法薬物取引等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，取り締まりから逃れて取引をするために容易に監視対象となることが想像できる直接的な用語（「援助交際」,「大麻」等）を避け，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．隠語は，たとえば，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，また，たとえば野菜や草（大麻の隠語），氷やクリスタル（覚せい剤の隠語）等，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．そのため，監視側は，常に最新の隠語を把握する労力が必要となり負担が大きい．また隠語の特徴として，そもそも一般的な投稿に比べ出現頻度は大幅に低く，また一般的な単語にカモフラージュされたり，連想されたりして使用される．このような隠語検出を目的とした研究は，これまでウェブサイトなどではなされてきたが，本研究で対象とする Social Networking Service (SNS) においてはそのまま適用することは難しい．なぜなら，SNS の特徴として，短文であること，文体が不完全な投稿が多いなどの特性があるからである．一方で，大規模言語データベースを作成し，その中から隠語を見つけようとしても，隠語が一般的な文章に埋もれてしまうため，検出が難しい．このため，本研究では，犯罪の端緒を迅速に把握するため，犯罪を誘導する投稿に含まれる可能性が高い隠語を検出することを目的としているところ，単語の類似語を基づく手法を考案することにより，隠語の検出を目指す．

隠語検出については，使用される単語の周りには，類似した関連する単語が出現するとの仮説のもと，二つのコーパス間の同じ単語の類似語の差異に着目し，隠語を検出する新たな手法を提案する．具体的には，二つの単語コーパス（一般的なコーパス（Good コーパス），犯罪を誘導するための既知の隠語を含んでいる投稿が含まれるコーパス（Bad コーパス））を用意し，それぞれで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コサイン類似度上位に出現する単語の差異から隠語を検出する方法について提案する．本提案手法に基づきプログラムを実装し，実験を行った結果，比較手法と比べて提案手法は高い精度で隠語を検出することができた．

続いて，多くの隠語検出研究において，分かち書きされた単語に基づき隠語を検出するため，二つ以上の単語を結合させた，すなわち文節で区切られる単語で構成される複合語型の隠語（以下,「複合語型隠語」という.）は検出ができないという課題があった．実際の投稿を分析する中で，対象コーパス内には複合語型隠語が存在を確認できたものの，形態素解析器による分かち書きの段階で，文節が分割されていたことも確認した．

そのため，複合語型隠語を発見することを課題とし，単語の関連性を活用することで複合語型隠語を検出する手法を提案する．具体的には，分析の結果，複合語は分割された単語同士の出現頻度が共に高いため，単語の関連性が高く現れることが判明したことから，単語の関連性を利用し，まずは複合語を検出する手法を提案した．そして，複合語型隠語検出実験を行なった結果，隠語検出提案手法のみでは検出できなかった複合語型隠語の検出を確認できた．

本提案手法を用いて，時と共に変遷する新しい隠語や犯罪を誘導する投稿を自動的かつ迅速に検出することで，隠語の継続的把握の負担が軽減でき，犯罪の端緒を迅速に掴む可能性が示されたといえる．

目 次

図 目 次

表 目 次

第1章　はじめに

1.1　はじめに

世界中で，違法薬物売買，売春などの違法な取引が問題となっており，国連のレポートを元にしたニュース記事においても，Facebook,Twitter,Instagram を介したオンライン麻薬取引の増加について言及されている [11]．また近年，ソーシャルネットワーキングサービス（以下,「SNS」という.）の急激な普及に伴い，SNS 上での投稿が起因となる犯罪については，世界的にも同様の犯罪は広がっている [12]．たとえば HaoYang らはこれらの犯罪を 7 つのカテゴリー（違法薬物売買，危険物，ギャンブル，性，Blackhat SEO，代理出産，その他の犯罪）に分類した [13]．日本でも，違法薬物売買以外にも，児童買春や闇バイトの募集，自殺幇助等が特に問題となっている．

警察庁の委託を受けた民間団体「インターネット・ホットラインセンター（IHC）」では 2006 年から，ネット上の違法・有害情報に関する通報を受け付けているところ，令和 4 年中の受理した通報件数は，584,733 件となっており，ホットライン運用ガイドラインに基づいて IHC にて分析した結果として，そのうち 4.4%に当たる 25,895 件 (分析の結果，1 つの通報で 2 件含まれること等あり) が違法情報との報告であった（図 1.1）.

違法情報についても，前年度と比べ相対的に見れば減少しているように見えるものの，H25 年から年平均で約 30,000 件と件数については大きく減少はしていない（図 1.2）.

さらには，認知され通報された件数が全てではなく，氷山の一角に過ぎないと考えられる．なぜなら，巧妙な投稿については，監視の目をかいくぐり取引を行うために，その分野に興味を持っている特定のものにしか分からないような単語，いわゆる「隠語」を用いてやり取りされているからである．最近でも Twitter で隠語を使って覚醒剤の取引を呼びかけたとして，麻薬特例法違反 (あおり，唆し) の疑いで男性が逮捕される事案 [14] などが

[2]ICH 資料
[2]平成 30 年における組織犯罪の情勢

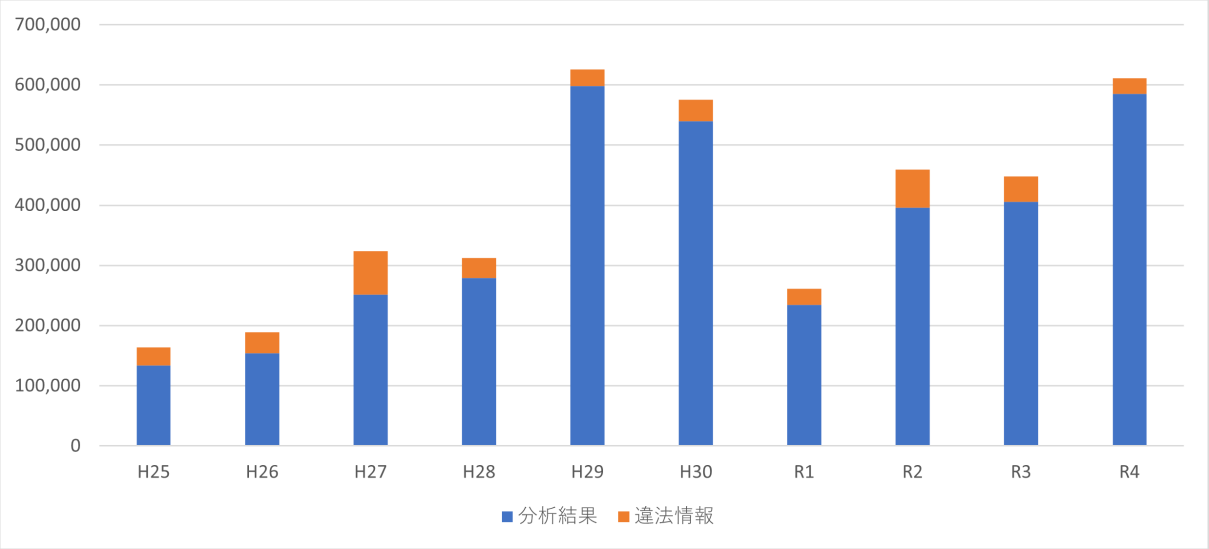


図 1.1: IHC における通報件数の推移 (ICH 資料 [1] を元に作成．)

発生している．また，麻薬特例法については SNS の投稿などにも適用しているところ，図 1.3 のとおり，近年急激に検挙件数が増加している．

また，援助交際については，特に最近では，東京では歌舞伎町の俗に言う「トー横キッズ」や大阪では道頓堀周辺の「グリ下キッズ」などの呼び名で若年層の若者がたむろする様が確認されており，彼らが性犯罪などに巻き込まれていることが問題となっている．立ちんぼと呼ばれる街角に立って客引きをする売春行為を行うことも確認されており，SNS で募集されている場合もある．

それ以外にも，最近では闇バイトと呼ばれる SNS 上で犯罪の人員を募ることに注目が集まっており今年になって，殺人にまで発展した凶悪犯罪として，世間の話題だけでなく国や警察が至急対策チームを立ち上げているところである [15].

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ながら取引をするため，監視されていることを想定した上で監視対象となることが容易に想像できる犯罪に直接関係する単語 (「大麻」,「覚醒剤」等) を避け，図 1.4 のように監視者からは見つかりにくく，ただし言葉の意味を知っている者同士は分かる隠語を用いて，違法な取引を実施する傾向にある．

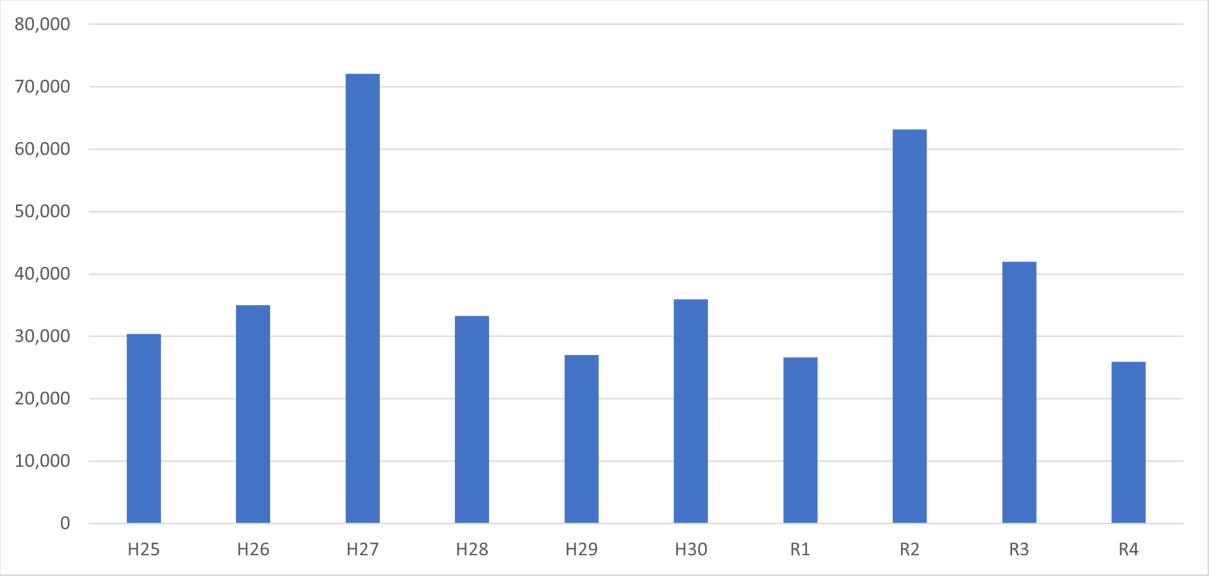


図 1.2: IHC における通報のうち，違法情報の件数の推移（ICH 資料 [2] を元に作成．）

隠語の例を挙げると，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」，児童買春については「援助交際」から「援交」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るようになるからである ([16]，[17]，[18]，[19])．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」，買春についても,「円光」と漢字を変えたり,「パパ活」,「P」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」という隠語も生まれてきたという例もある．

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．また，隠語の周辺には隠語とは言えないものの，隠語に準じた，隠語と共に出現する傾向が高い単語も確認されている．本研究では，このような単語を「犯罪関連語」と定義し，隠語と合

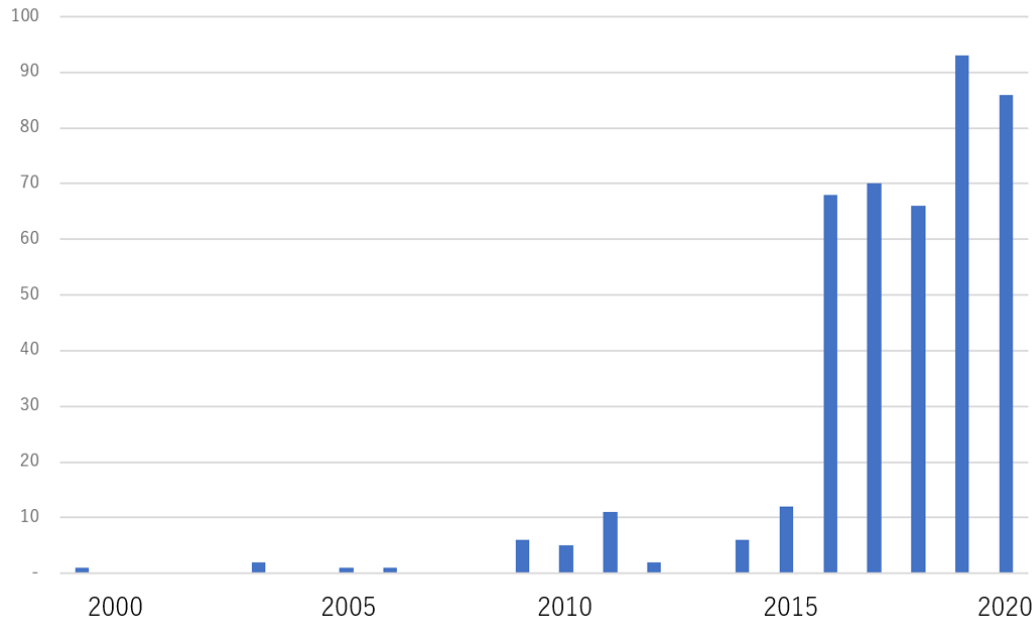


図 1.3: 麻薬特例法違反（あおり，唆し）による検挙件数の推移（法務省のデータ [3] を元に作成.)

わせて犯罪関連語の検出も目指す.

そこで本研究では，単語の類似語を基づく手法を考案することにより，隠語の検出を目指すと共に，その中で，単語の関連性に基づく手法を考案し用いることで，提案手法による隠語検出の幅をより広げることを目指した.

提案手法については，以下のとおり述べる.

1. 提案手法 1(隠語検出手法)

犯罪を誘導する隠語及び隠語と共に出現する傾向が高い単語 (犯罪関連語) を検出するため，不正な取引に使用される単語の周りには，類似した関連する単語が出現するとの仮説のもと，二つのコーパスにおける同じ単語の類似語の差異に着目した．具体的には，全く性質の異なる二つのコーパス（一般的なコーパス（Good コーパス），犯罪を誘導するための既知の隠語を含んでいる投稿が含まれるコーパス（Bad コーパス））のそれぞれで Word2vec[20] 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コサイン類似度上位に出現する単語（以下,「類似語」という.）の差異から隠語を検出する方法について提案する.

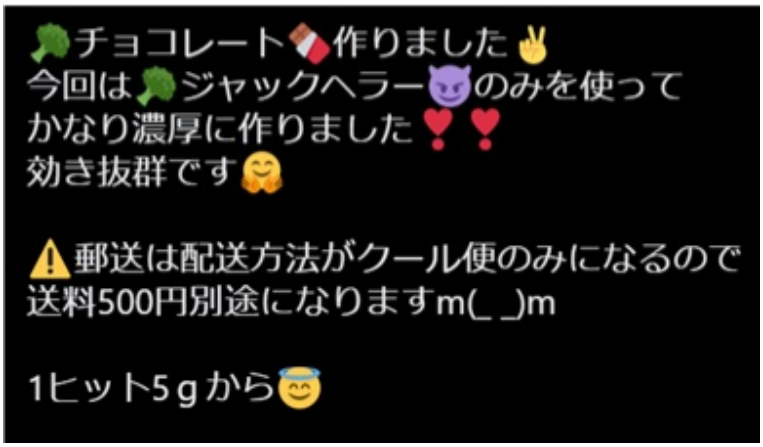


図 1.4: 隠語が用いられた文章例：違法薬物の入荷を隠語を使って表現している

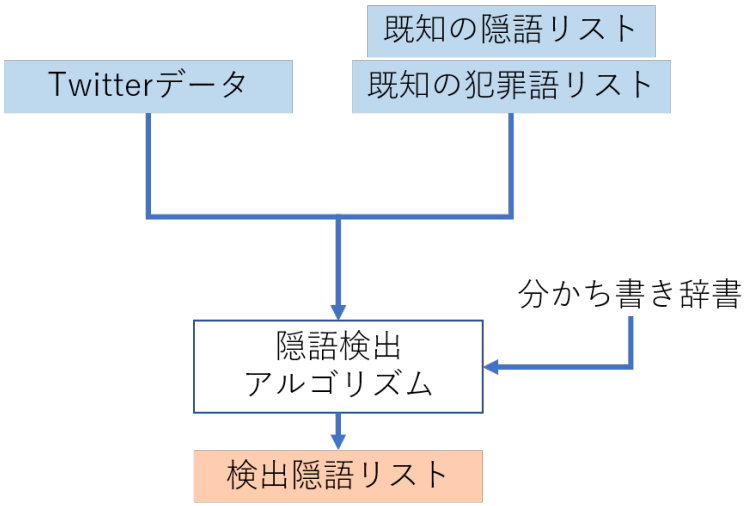


図 1.5: 提案手法 1 の概要図

なお，提案手法 1 の提案の概要図は，図 1.5 のとおり．

2. 提案手法 2(複合語型隠語検出手法)

隠語検出の手法は近年いくつか存在しているが，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．そして，こうして得られた単語集合の各単語に対して，隠語かどうかを判定するため，二つ以上の単語を結合させた複合語でかつ隠語については，検出できないという問題点がある．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．そのため，複合語型隠語ではない，単文節の隠語を「単一隠語」と呼ぶこととする．つまり，空白で単語が区

切れる英語等と異なり，日本語や韓国語等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，分かち書き用のツール内に組み込まれた内部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，意図しない文節で区切られることとなり，隠語検出提案手法だけでは複合語型隠語を検出することができない．

「隠語検出提案手法」において，複合語型隠語が検出できない理由として，単語分散表現モデル生成前の前処理において，複合語型隠語に該当するような単語は一般的な認識も低いことから，分かち書き用時に文節が分けられてしまっていたためである．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が，単語の認知も低いことから，自動的に内部辞書が充実することも期待できない．そのため，事前に隠語に限らず複合語を広く辞書登録できれば，複合語型隠語の文節で分かち書きが行われ，提案手法により複合語型隠語の検出が期待できる．

このようなことから，複合語型隠語の検出という課題に対し，提案手法 2 として，

単語の関連性を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，複合語型隠語の文節で分かち書きされるようにし,「隠語検出提案手法」と組み合わせることで，単一隠語だけでなく，複合語型隠語も検出可能とし，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，提案手法 2 の提案の概要図は，図 1.6 のとおり．

1.2　本論文の構成

本稿の構成は以下のとおりである．第 2 章では，本研究の背景について記載する．第 3 章では，本研究の関連研究について記載する．第 4 章では，隠語の特性に関する調査結果について述べる．そして，第 5 章では，隠語検出を目的とした提案手法について説明する．それを受けて，第 6 章では，第 5 章で説明した提案手法に基づき，実験を行った結果を示

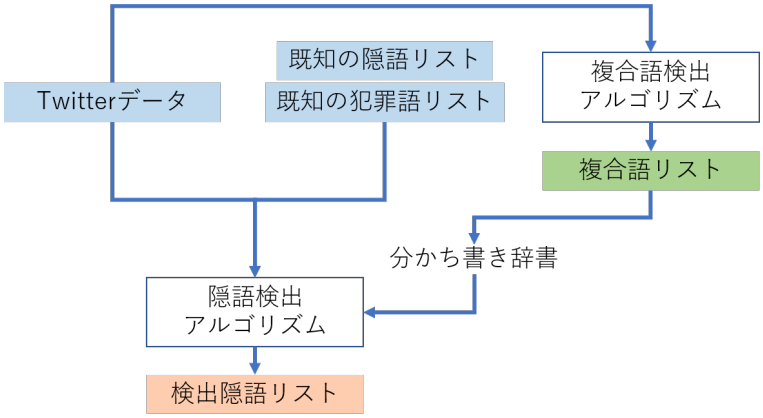


図 1.6: 提案手法 2 の概要図

す．第 7 章では，第 6 章の手法のみでは検出できなかった複合語型隠語の検出を目的とした提案手法 2 について説明する．そして，第 8 章では提案手法 2 で事前に検出することを想定した複合語が実際に検出されたかの確認実験及び結果を示し，第 9 章では第 8 章の結果を元に提案手法 2 による複合語隠語検出実験の結果を示す．第 10 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．最後に，本研究の結論を第 11 章で述べる．

第2章　背景

本章では，本研究の背景について説明する．まず，2.1 節で SNS に起因した犯罪の各種類ごとの状況と国としての対策の動きについて説明する．続いて，2.2 節では，本研究で対象とする隠語と本研究で新たに定義した「犯罪関連語」について説明する．また本研究で対象とする単語について体系的に整理し分類した．2.3 節では，複合語と本研究で定義した複合語型隠語について説明する．2.4 節では，本研究で対象とするマイクロブログの一種である Twitter について説明する．

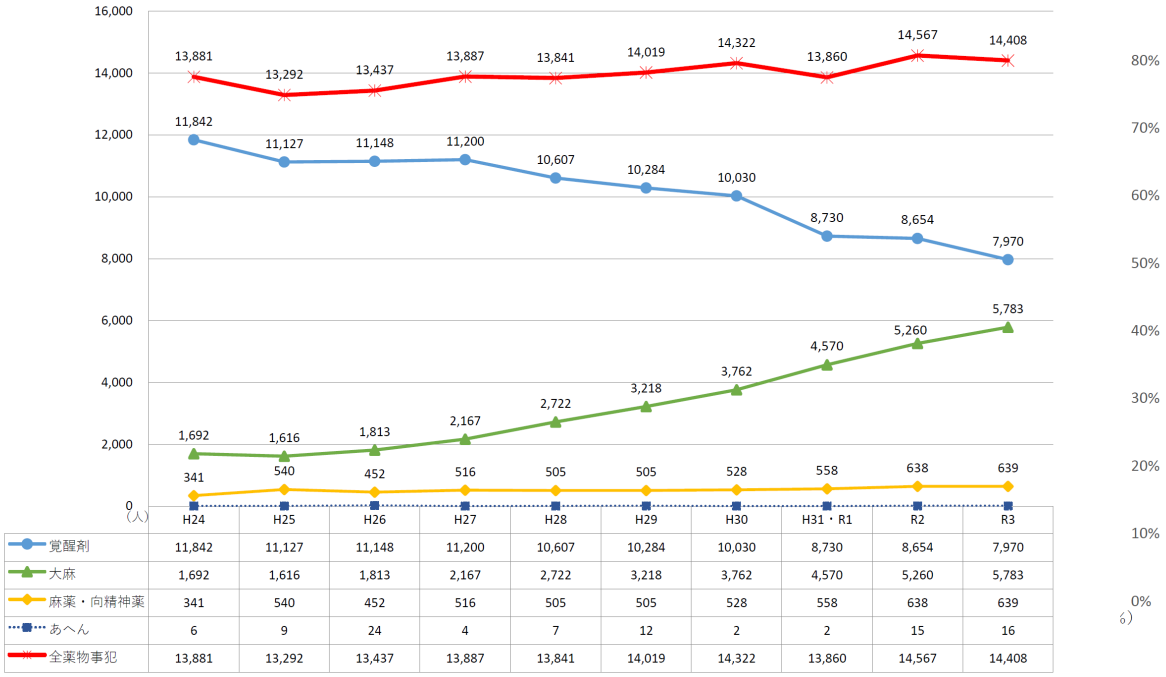
2.1　SNS に起因した犯罪の増加とその対策について

1.1 章でも言及したように，違法薬物の使用を唆す投稿をはじめとした違法投稿がインターネット上で増えており，それは IHC への通報件数にも現れている．これらの犯罪は，合法ドラッグやパパ活，闇バイトなどと比較的カジュアルで犯罪性を感じにくい言葉に置き換えられることで，若年層を中心に心理的抵抗感を下げ，知らず知らずのうちに違法薬物を売買させたり，犯罪に加担させている．

具体的に，主に犯罪が多発している違法薬物売買，援助交際，闇バイトについての現状とその対応策の現状について，以下のとおり述べる．

2.1.1　違法薬物売買

違法薬物について，図 2.1 のとおり，薬物ごとの検挙人員の推移について示しているところ，大麻については，8 年連続で増加し，過去最多を更新している．大麻は，ゲートウェイドラッグとも呼ばれ，その他のコカイン，ヘロイン，覚せい剤などの他の更に強い依存性の薬物使用の入口となる薬物と言われている．



	H24	H25	H26	H27	H28	H29	H30	H31・R1	R2	R3
覚醒剤	11,842	11,127	11,148	11,200	10,607	10,284	10,030	8,730	8,654	7,970
大麻	1,692	1,616	1,813	2,167	2,722	3,218	3,762	4,570	5,260	5,783
麻薬・向精神薬	341	540	452	516	505	505	528	558	638	639
あへん	6	9	24	4	7	12	2	2	15	16
全薬物事犯	13,881	13,292	13,437	13,887	13,841	14,019	14,322	13,860	14,567	14,408

図 2.1: 薬物事犯検挙人員の推移 (第五次薬物乱用防止五か年戦略フォローアップ 令和 5 年 8 月 8 日取りまとめより [4])

その大麻について，年代別の大麻事犯における検挙人数の推移を見たところ，特に 30 歳未満は，平成 25 年から 5.5 倍の増加，中でも 20 歳未満については，平成 25 年から 16.4 倍の増加と増加率が著しい (図 2.2)．この若年層を狙って，SNS の中でも若年層の利用が多い Twitter が悪用されることが多い．

このような犯罪の増加を危惧し，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 ([21])．本戦略の中では，5 つの目標が掲げられているところ，この中で，SNS に関するものについても,「巧妙化潜在化する密売事犯等への対応」として，以下の具体的な取り組みをすることが明記されている．

「インターネット等を利用した密売事犯への対応強化」項目として，

- インターネット・ホットラインセンター（ＩＨＣ），あやしいヤクブツ連絡ネット等からの通報及びサイバーパトロールにより，薬物密売に関する違法情報の収集を推進する．(警察庁，厚生労働省)

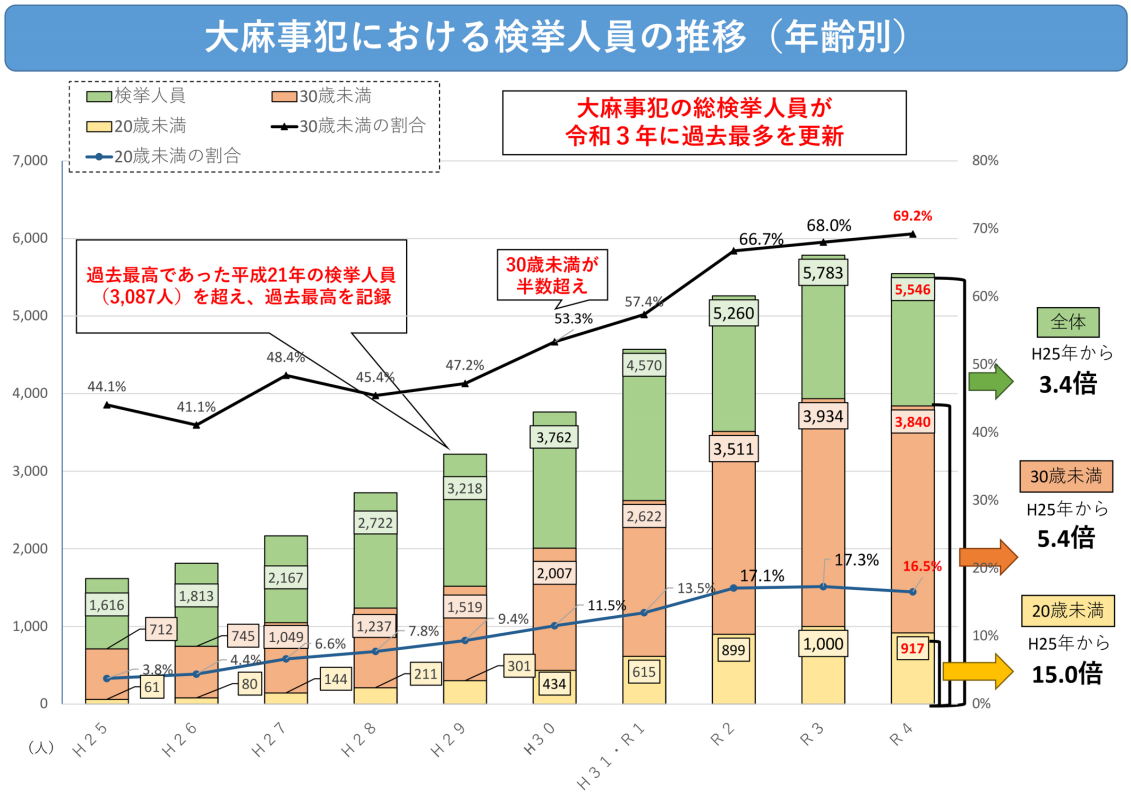


図 2.2: 年代別大麻事犯の検挙人員の推移 (第五次薬物乱用防止五か年戦略フォローアップ 令和 5 年 8 月 8 日取りまとめより [4])

- インターネット等を利用した薬物密売手口の解析・分析を強化するとともに，各種法令を駆使した取締りを推進する.（警察庁，厚生労働省）
- 違法情報に関する証拠保全や送信防止措置を進めるため，プロバイダ等との協力関係を強化する.（警察庁，厚生労働省）

「各国・地域における薬物密売手口と対策に関する情報収集の推進」項目として，

- 各国・地域の捜査機関から，課題となっている密売手口やその対策等に関する情報を収集する.（警察庁，厚生労働省，海上保安庁）

とあり，共にどの項目においても，警察庁に求められる責務が多くを占めることが見て取れる.

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等により「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 ([22]).

また，大きな取り組みとは別に，隠語リストなどの提供については，継続的に行われている．アメリカにおいては隠語のリストが国家麻薬管理政策局提供の元，公開されているが，この中でもドラッグの隠語は常に変化するため，リストが公開されるとすぐに多少時代遅れになると言及している [23]．一方で日本においては，警察庁による啓蒙活動 [24] や都道府県警察や自治体において，参考例程度の若干の隠語のリストが公開されているところはあるが，十分に網羅しているとは言えない [25].

そのため，最新の隠語を把握することは非常に意義がある．

2.1.2 援助交際

援助交際は，海外でも「enjo kosai」というローマ字で表記されるほど一般的となっており，Miller氏は,「若い女性が見知らぬ男性との，お金や贈り物と引き換えに，時にはセックスを含むデートをすること」と表現している [26]．援助交際，パパ活などの呼び名で違法性がオブラートに包まれ行われているが，未成年による援助交際については,「児童買春，児童ポルノに係る行為等の規制及び処罰並びに児童の保護等に関する法律」の中でも規定されているとおり，犯罪となり得る．そして援助交際をきっかけに犯罪に巻き込まれるなど，SNSに起因した事犯の被害児童数は年々増加し，特に令和元年には過去最高の被害数を記録し，その後も高止まりしている（図2.3）．また被害のきっかけとなったアクセス手段についても，年々スマートフォンの割合が増加し，近年ではほぼスマートフォンとなっているのが見て取れる．

さらには，学職別の被害児童の推移からは，中学生・高校生で全体の88.6%を占めることが分かる 2.4.

この13〜19歳までの年齢層は，図2.5からも分かるとおり，スマートフォンの個人保有率が8割を超え，被害に遭遇するリスクは非常に高いと言える．

その中でも，被害児童が最も多く利用していたサイトはTwitterであり，約4割を占めていた（図2.6）.

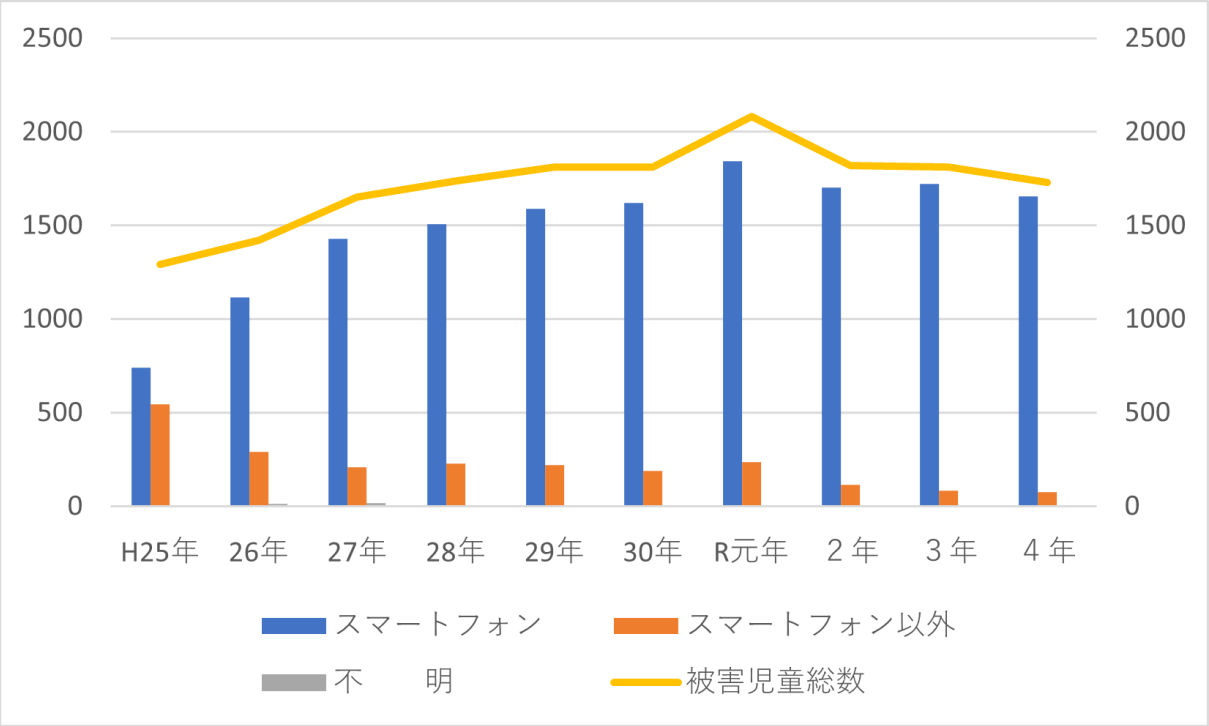


図 2.3: SNS 等に起因する被害児童数とアクセス手段の推移（警察庁資料 [5] を元に作成）

2.1.3 闇バイト募集

高額な報酬を掲げつつ，簡単さを謳い人を募ることで，特殊詐欺当の犯罪にに加担させるアルバイトのことを，いわゆる「闇バイト」と言われている．特殊詐欺とは，被害者に電話をかけるなどして対面することなく信頼させ，指定した預貯金口座への振込みその他の方法により，不特定多数の者から現金等をだまし取る犯罪をいい，オレオレ詐欺などの名前で紹介されている．警察としても全国的に周知しているもののその認知状況と被害額は依然として高いままである．具体的には，特殊詐欺の認知件数は，令和 3 年以降増加しており，その被害額は，令和 4 年に 8 年ぶりに増加に転じている．また検挙件数・人員についても，令和 4 年に増加に転じている (図 2.7).

令和 2 年 1 月 1 日から，特殊詐欺の手口について以下の 10 種類に分類された [27].

- オレオレ詐欺
- 預貯金詐欺

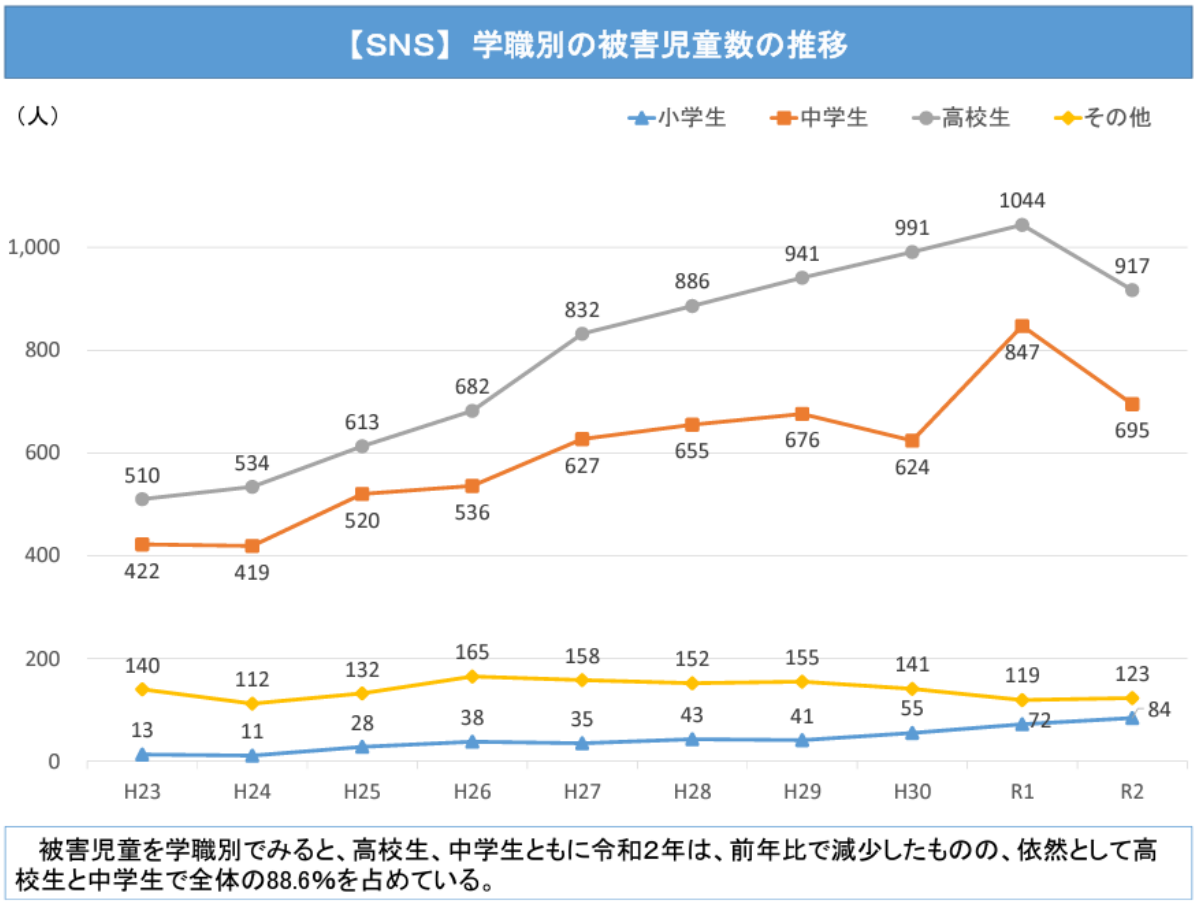


図 2.4: SNS 等に起因する学職別の被害児童数の推移（警察庁資料 [6]，[7] を元に作成）

- 還付金詐欺
- 架空料金請求詐欺
- キャッシュカード詐欺盗（窃盗）
- 融資保証金詐欺
- 交際あっせん詐欺
- 金融商品詐欺
- ギャンブル詐欺

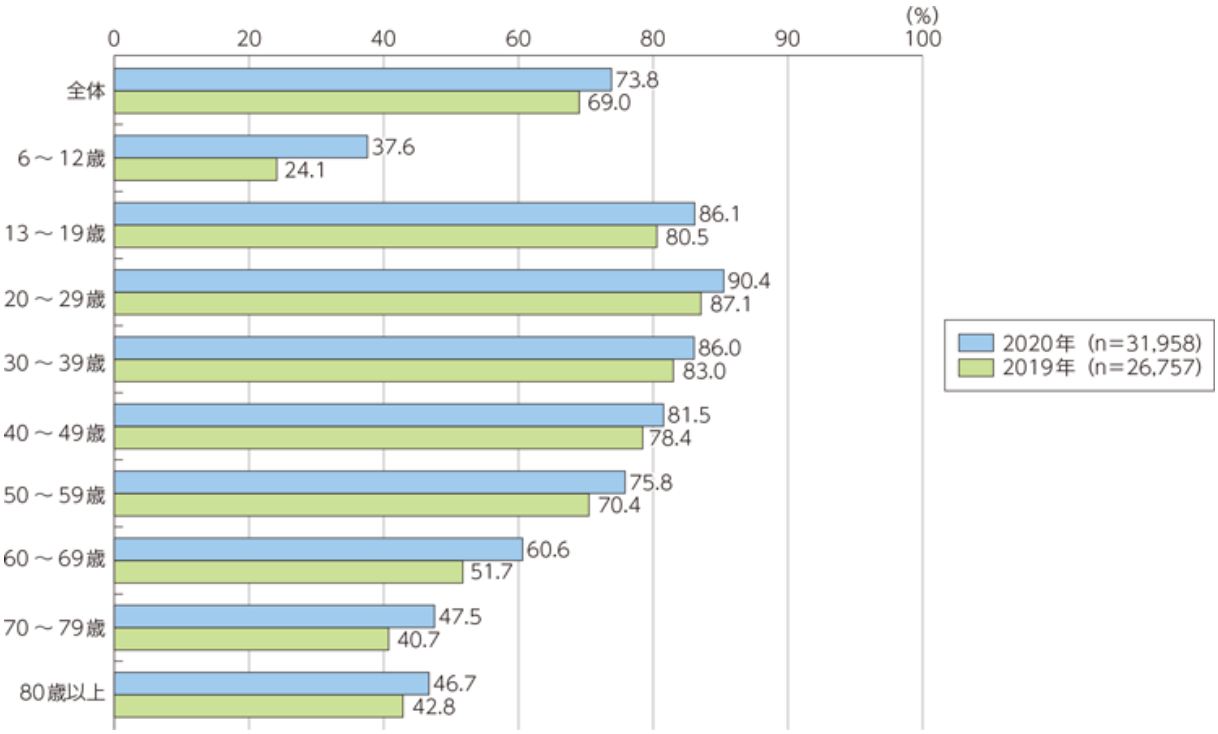


図 2.5: 年齢階層別ソーシャルネットワーキングサービスの利用状況 (総務省資料 [8] より抜粋)

このような特殊詐欺に加担するおそれのある,「闇バイト」には具体的には大きく 5 つの役割があり，それぞれの役割について，隠語を用いて募集される.

- 受け子

  特殊詐欺の手先として，被害者から現金やキャッシュカードを受け取る役のことである.

- 出し子

  特殊詐欺の手先として，ATM で他人の口座から現金を引き出す役のことである．簡単なアルバイトほど詐欺の可能性がある.

- 架け子

  特殊詐欺の手先として，被害者に電話を掛ける役のことである.「鞄なくした.」「女性を妊娠させた.」「会社の金を横領した.」など嘘の電話でお金を要求する.

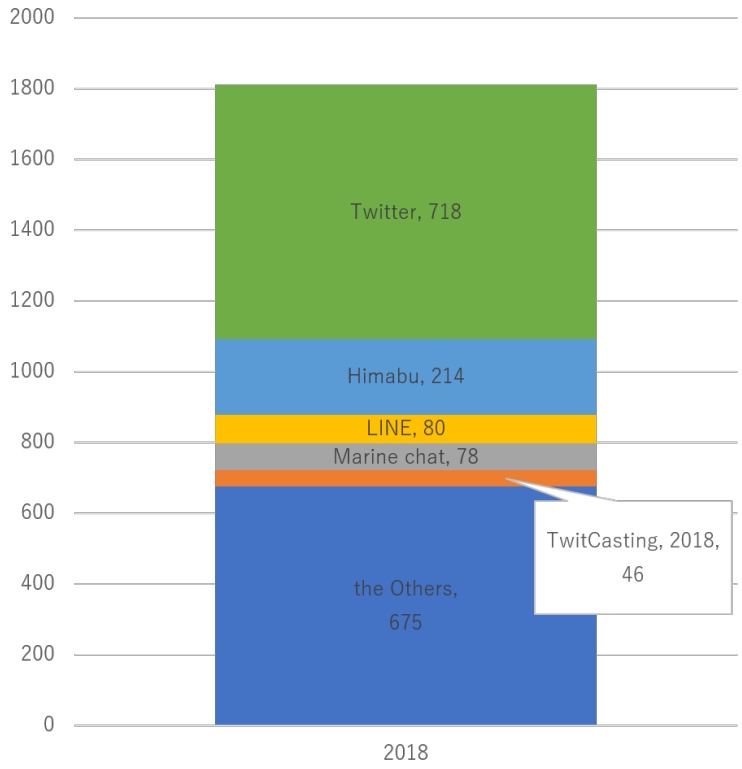


図 2.6: 被害児童が利用していたサイト (警察庁資料 [9] を元に作成)

- 運び屋

  特殊詐欺の手先として，目的も相手もわからないのにコインロッカーから，封筒に入った現金などを運ぶ役のことである．

- 道具屋

  特殊詐欺の手先として，他人名義の携帯電話や銀行口座を手配する役のことである．

これらの特殊詐欺とこれに起因する闇バイトについては，これまでも問題となっていたが，近年，特に「闇バイト強盗」と称される SNS 上で実行犯を募集する手口等を特徴とする一連の強盗等事件など，凶悪性，暴力性が増している．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．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 [15]．その中でも，「実行犯を生まない」ための対策として，「闇バ

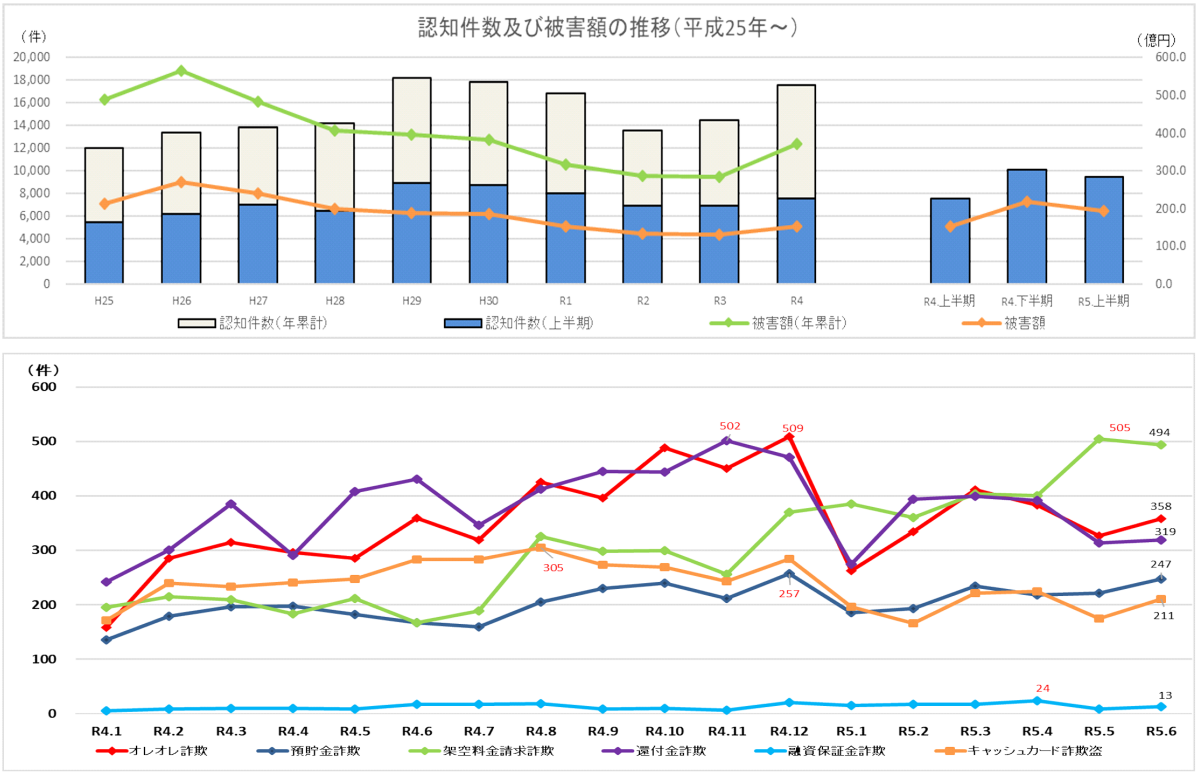


図 2.7: 特殊詐欺の認知件数及び被害額の推移（警察庁資料を元に作成．）

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，時事ニュースによると，警視庁犯罪抑止対策本部は令和 3 年に闇バイトなどの犯罪行為に引き込む疑わしい投稿を発見次第，リプライ（返信）を通じて警告する取り組みを本格的に開始し，警告件数は令和 3 年に 2,246 件であったところ，令和 4 年には 3,480 件と約 1.5 倍に急増した [28].

その数は図 2.8 のとおり，認知件数も被害額も年々増加している．特殊詐欺の被害が増えるということは，それだけ闇バイトも活発に募集されているということであるため，可及的速やかに対策する必要が求められている．

このような状況を踏まえると共に，SNS の中では Facebook や Twitter が主に利用されていることと [29]，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．

図 2.8: 特殊詐欺の認知件数の推移（警察庁資料 [10] を元に作成．）

2.2　隠語及び犯罪関連語について

2.2.1　隠語について

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いる[1]．たとえば，警察では犯人のことを「ホシ」，汚職事件を「さんずい」といったり，すし屋ではすし飯を「シャリ」，お茶を「上がり」といったりするなど，他人にわからないことばを使うことで仲間意識を強める，特別なことばを考え出して使うことで単調さを破る，といった効用もあり[2]，我々の身の回りでも様々な場面で使用されている．

またインターネットにおける違法な取引は世界中で問題視されており，日本だけでなく世界中でその際に隠語を用いて警察等からの取り締まりを回避している．たとえば，Yuan らは「Dark Jargon」[18]，Hao らは「Black keyword」[13] と表現しているが，具体的には，中国語の場合,「溜冰」は一般的にはアイススケートを意味するが，覚せい剤という隠語としての意味を持ち [13]，英語では，rat がブラックマーケットでは，不正プログラムの隠語として使用されている [18]．

なお，本研究における隠語の対象は，警察等の目をかいくぐり，犯罪に用いられる単語

[1]デジタル大辞泉 (小学館) より引用
[2]日本大百科全書 (小学館) より引用

とした.

2.2.2 犯罪関連語の定義

またそれ以外に，実際の違法な取引に係るツイートにおける隠語を確認する中で，たとえば，一万円の単位を表す「諭吉」や郵送ではなく手渡しを意味する「手押し」などが確認されたところ，これらの隠語は，隠語として用いられる単語ではあるものの，取引対象がなくそれ単体では意味するところを充分に伝えることができないため，単独で使用される可能性が極めて少ないと思われる．一般的に取引には,「取引対象」,「取引対象を形容するもの（高品質等）」,「時間」,「場所」,「取引方法」,「取引量」,「金額」などの情報が必要となる．これらの情報を表すための語として，取引を行う者たちの間で共通認識が生じれば隠語が発生し得るが，まだ隠語として確立している単語がない状況では，お互いに誤解が生じないよう一般的な言葉が利用される．このうち取引対象については隠語としてやり取りされる傾向にあるが，それ以外の情報にあっては隠語にしてしまった場合に取引自体が不明確となるため，取引きが円滑に進まない懸念があることから隠語になりづらい傾向にあると思われる．そのため，隠語を用いて巧妙に意図を悟られずに取引しようとしても,「取引対象」,「場所（都内等）」,「金額」という少なくとも三つの情報が含まれている必要がある．さらには監視の目をかいくぐりながら迅速なやり取りを実現させるためには,「取引方法（手押し，郵送等）」や「取引対象を形容するもの（高品質等）」「取引対象の質（高品質，ハイグレード等）」なども文章内に含まれている必要がある．このような，そ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.

2.2.3 隠語と犯罪関連語の整理

本研究で対象とする単語について，具体的に，表 2.1 のとおり，大きく 3 つの観点で整理した.

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造語に該当する．たとえば，違法薬物売買関連の隠語としては，大麻の場合は「ハシシ」,「ポット」，覚醒剤には「シャブ」,「ガンコロ」といった隠語がある．一方，援助交際関連の隠語としては,「神待ち」といった隠語がある．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」などがこれに当たる．また，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」，覚醒剤を表す単語としては,「アイス」や「クリスタル」といった隠語がある．一方，援助交際関連の隠語としては，一万円の肖像画の人物名である「福沢諭吉」から一万円の単位を指す隠語として「諭吉」などが用いられている．

2. 犯罪行為または犯罪となる物を指すかどうか
   隠語とは，ある特定の社会・集団内でだけ通用する特殊な語であることから，犯罪行為でなくとも使用される．特筆すべき点としては，その単語単体では犯罪行為を指さない「犯罪関連語」が分類されることである．

3. 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

Hao らは black keyword について，その特徴として，フィラーワードと呼ばれる，いわゆる間を埋める表現やストップワードなどを表す単語とロングテールキーワードと呼ばれる複数の単語を組み合わせた検索されにくい出現頻度の低いキーワードを組み合わせたものであると表現している [13]．同様の考え方で隠語が世界中で生成され使用されておりそ

表 2.1: 隠語・犯罪関連語の整理

		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 (接着剤の名前)，カリフォルニアオレンジ (アーティスト名)
	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

の特性も似ていることから，日本語の複合語型隠語にあっても共通する部分があるとと考えられる．

なぜなら，複合語型隠語について収集した Tweet 文を元に分析したところ，2.2 節のうち，造語や認知度の低い対象の単語名については，複合語型隠語が多く出現する傾向にあったからである．まず複合語について，複合語とは，本来独立した単語が二つ以上結合して，新たに一つの単語としての意味・機能をもつようになったものと定義されている[1]．

複合語の隠語の例として，大麻の隠語としては，たとえば「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，たとえば，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」というように単単語の文節で区切られてしまう．

複合語型隠語が分かち書き時に文節で分断されてしまう対策として，まず分かち書きの

[1]デジタル大辞泉 (小学館) より引用

文節単位の調節機能の変更などが考えられるが，内部辞書に基づく分かち書きを行うため，内部辞書に存在しない単語の場合は，複合語の単位で分かち書きされない．

隠語として使われるものの中には，造語や認知度の低いものも数多く存在するため，分かち書き用の辞書に登録されていないことも多く，実際の環境における複合語型隠語は，登録すべき単語が不明であり，造語や認知度の低さの点から，自動的に辞書に追加される可能性が低いため認知した複合語型隠語を辞書に登録してとしても，変遷する最新の隠語をキャッチアップすることは非常に労力がかかる．

さらに，たとえば「レモンスカンク」のように，文節の中に「レモン」のような一般的な単語が含まれている場合，文節が区切られてしまいやすいことが原因として考えられる．

そこで，1.1 節でも述べたが，事前に隠語に限らず複合語を広く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，複合語型隠語の検出が期待できる．

2.4　Twitter の特徴について

犯罪に関連する投稿の詳細については，違法なやり取りを発覚しないようにするため，基本的には秘匿性の高いインスタントメッセージアプリケーションである「テレグラム」が使用される傾向がある．ただし，テレグラムに誘うため，一般的に使用されるツールでの投稿の中に，隠語を交えて投稿することで，対象とする取引について興味があったり，精通したりするものにだけ分かるようにして投稿する．そして，捨てアカウントと呼ばれる一時的にだけ用いるアカウントを用いて，短時間の投稿を繰り返し，発覚から逃れつつ目的を達しようとする．このようなことから，犯罪の端緒を捉えるため，違法な取引を防止したり，違法な投稿により犯罪に巻き込まれる被害者を減らすためにも，迅速にまた自動的に投稿を検知することは重要であり，そのためにも隠語を把握することは非常に重要なことである．

ただし，隠語の研究は，これまでウェブサイトでは検索する手法が研究されてきたが [19]，[30]，[31]，Twitter などのマイクロブログではそのまま適用することは難しいと思われる．

その理由は次のとおりである．まず，マイクロブログの特徴として，以下の特徴が述べられている [32]．

1. 短い文字数
   Twitter では，投稿できる一度の文字数は 140 字までの制限がある．

2. 文法が非公式で構造化されていない
   会話調で編集もされていないため，スラングや略語や誤字も多い．

また，Furkan らも Twitter については，短文でノイズが多いため，トピック分析も難しいと述べている [33].

Twitter における隠語に関連するツイートの分析を行ったところ，短文が多く出現するだけでなく，中でも犯罪取引に用いられるツイートは，犯罪の意図を隠そうとするため，さらに文章の体をなしてないことが多く見受けられた．そのため，文の係り受けなどを用いた分析や機械学習などについては難しいと考えられる．一方で，取引をできる限り短いやり取りで成立させるため，一つの投稿の中に取引対象や場所・金額・品質等の必要な情報について書き込む必要があるため，本研究では犯罪関連語と定義した隠語周辺に犯罪に関連した単語が出現しやすい傾向を発見した．そこで，取引に関連したツイートについては，隠語の周辺に犯罪に関連した単語が出現する傾向を利用し，単語分散表現を用いることで，効果的に犯罪取引に用いられる隠語や犯罪関連語の類似語を見つけることができると考えた．

なお，単語分散表現として本研究では Word2vec を用いたが，Word2vec とコサイン類似度を用いた研究については，近年でもいくつか報告されているが，未知の隠語の検出に使った事例はない [34],[35].

このようなことから，既知の隠語を手掛かりに，その類似する単語に着目し，未知の隠語の検出を目指す．

第3章　関連研究

本章では，関連研究について述べる．まず，3.1 節で，隠語検出に関連する関連研究について述べる．続いて，3.2 節では複合語に関する研究，そして 3.3 節では，単語分散表現について述べる．

3.1　隠語等の検出

3.1.1　ウェブサイト等による隠語検出

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，煽り用語及び特定のジャンルの専門用語等の検出に関する研究は，いくつか報告されている [30],[36],[31], [37],[19]．たとえば，専門用語の検出として北村らの研究があげられる [38]．北村らの研究は，文対応の付いた対訳コーパスから共起する単語列を対応付けることにより，対訳表現を自動的に抽出する方法を提案するというものである．そして，特定分野特有の専門用語等の翻訳について，高精度で適切な対訳表現を抽出したと報告されている．

しかしながら，北村らの手法は 2 言語間の専門用語の対応付けを行うものであるが，これを隠語に応用すると，隠語を使わずに表現された文と，まったく同じ意味を持つが隠語を使って表現された文のペアを多数用意する必要がある．たとえば,「都内　野菜　手押し」という文が「都内で大麻を手渡しで販売する」という文と同じ意味を持つというラベル付けを事前に行っておく必要がある．このようなペアが多数存在するとき,「野菜」という単語が「大麻」の隠語であると判断することが可能となる．北村らの元々のターゲットである翻訳のドメインでは，同じ意味を持つ日本語の文と英語の文のペアを多数取得することが可能である（たとえば，日本語と英語で書かれた特許文章）．しかし，隠語を含む文と含まない文で同じ意味の文のペアのデータベースは著者らの知る限り存在しない．したがって北村らの手法を未知の隠語発見に利用することはできない．さらに，文章単位で対応付

け（対訳と単語組）が必要なことについて，実運用を考えると，データベースの規模の拡大に対応できず，運用担当者の負荷が非常に高いため，継続的な運用は難しいと思われる．一方で我々の手法は，既知の隠語（リスト内）から未知の隠語（リスト外）を検出することがポイントであり，さらに再帰的手法を用いることで精度を上げて検出している．また，対応付けの文章も不要であるため，実運用を想定した場合でも有効であると言える．

隠語の検出として，橋本らの研究があげられるが，本研究は，文章中の単語の語彙の決定及び係り受け関係にある 2 文節間の深層格の決定を行う意味解析システムの開発の中で，特に隠語の有害語意と文脈に登場する他の語（周辺語）の語意との共起頻度を辞書化しこれを元に隠語の語意を決定することで，有害語意を検出するというものである [37]．橋本らの研究は，隠語を事前に把握している前提で，ある単語が隠語として使われているか普通の意味で使われているかを判定するものであり，未知の隠語を発見するものではない．一方で我々の手法は未知の隠語を検出することを目的としている．また橋本らの研究は係り受けができる前提であり，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と考えられる．そのほかとして，大西らは，アンダーグラウンド系掲示板において，投稿された単語及びその周辺語に着目し，隠語検出を試みている [19]．また，橋本らは，周辺語に着目し，周辺語を考慮することで，ダブルミーニングの隠語の検出率を大幅に改善できたことを報告している [37]．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，一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 [37]．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．

それ以外にも，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 [18]．その際，Word2vec[20] 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，そのうちの二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，Yuan らの研究は，ダークウェブ上の隠語が対象であるが，一般の若年層が多く使用している短文で文脈性のないマイクロブログを対象としていない．また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．

3.1.2 ウェブサイト等以外の隠語検出

また，Web サイトや掲示板以外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 [39],[13],[40]，[41]．中国語については，Zhao らは，中国におけるアンダーグラウンドマーケットにおけるサイバー犯罪に使われる隠語に着目し，教師なし学習を用いて隠語の検出を実施している [42]．本研究は，隠語を把握している前提であるため，未知の隠語を実験の想定としていない．

一方，日本語を対象としたものとしては，安彦らの ID 掲示板を対象としたものが報告されている [43],[44]．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．ただし，本手法は,「野菜」,「アイス」などダブルミーニングな隠語についてものへの対応が対象としておらず，また 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，4.2 節において調査した結果，一般的な投稿の方が圧倒的に多い (たとえば，野菜については 97.5%) ため，そういった点で ID 掲示板とは大きく異なる．

3.1.3 Twitter における隠語検出

本研究で対象とした Twitter については，犯罪の軽減を目的とした研究がなされている [45]，[46]．またその中でも，攻撃的な単語や不正な単語を検出する研究についても，行われている．[47]，[48]，[49],[50].

隠語に関する研究としては，住田らは，不正なツイートを対象とし，機械学習を用いて有害性を分類している [51]．この研究では，ツイート全体が有害かどうかを判断しているが，隠語自体を別途理解する必要があると考える．なぜなら，機械学習させるにあたり，学習データをアノテーションする必要があり，その際，隠語の知識がなくては正しくアノテー

ションできないおそれがあるからである．そこで新しい隠語の情報が増えることは，アノテーションの精度が上がり，その結果，有害性の判断の精度も上がることが期待できる．

住田らの研究のように，隠語等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 [43],[42]．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．

また青木らの研究についても，同様に相乗効果が期待できると考えられる．青木らの研究は，橋本ら [37] と同様，周辺語に着目しており，ある単語が一般的ではない使われ方がされていた場合，その周辺単語は一般的な用法として使われた場合のものと異なるという仮説に基づいて，着目単語とその周辺単語の単語ベクトルを利用し，注目している単語の周辺単語が均衡コーパスにおける一般的な用法の場合の周辺単語とどの程度異なっているかを評価することにより，一般的ではない用法の検出を行う手法である [52]．青木らの手法では，前提として，隠語を事前に認知している必要があり，その隠語が出現する文章を用意する必要がある．しかしながら，隠語は監視の目をかいくぐり，特定の人しか分からないようにする特徴があることから，新しい隠語を把握し続けることは非常に労力を要する．つまり，我々の手法との最大の違いは，青木らの研究が未知の隠語を新たに見つけるものではない一方で，我々の手法は，ある単語の類似語から似たような使われ方をする単語を検出することであり，それは我々ですら認識していない未知の隠語を発見できるところにある．また青木らの手法については，我々の手法で新たな隠語を発見し，それが隠語であると認知でき，その隠語を用いた文章を集めることができたあとで，一般的な使われ方をしているか，そうでないか効果を発揮できる補完的な関係である．つまり,「野菜」が「大麻」の隠語と検出できるのが，我々の手法であり,「野菜」という単語が文章内に出現した際に，隠語としての使われ方をしているのか，一般的な意味で使われているのか判断するのが青木らの手法との認識であり，青木らの手法を効果的に使うためには，我々の手法が必須と考えている．

3.2 複合語に関する研究

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等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，Wanzheng ら [53] 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，Wanzheng 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 のような誤字等も多い，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されてきている [54], [55], [56], [57], [58], [59], [60].

例えば，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 [61]．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 [62].

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，2.4 節でも述べたとおり，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，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．

3.3　単語分散表現学習

単語分散表現とは，単語の意味を的確に捉えたベクトル表現のことをいう．分散表現の獲得方法としては，例えば，シソーラスを用いた手法やカウントベースの手法，推論ベースがある．このうち，単語の類似語を求めることは，カウントベース手法でも可能ではあるが，共起回数を直接扱うのではコーパスにより強く依存すると思われるため，周辺に出現する単語を確率的に扱う推論ベースの手法を本研究では用いることにした．

推論ベースの中には，最も有名なものとして word2vec があるが，単語をベクトル表現にし，足し引きすることで別の単語をみつけることができ，例えば，式 3.1 のように王から男性を引いて女性を足すことで女王を推測することが報告されている [20]．

$$vec\,(\text{"}king\text{"}) - vec\,(\text{"}man\text{"}) + vec\,(\text{"}woman\text{"}) \approx vec\,(\text{"}queen\text{"}) \tag{3.1}$$

また，同論文の中でそれ以外にマドリードからスペインを引き，フランスを足すと他の単語ベクトルよりもパリが近くなるということも言われている（式 3.2）．

$$vec\,(\text{"}Madrid\text{"}) - vec\,(\text{"}Spain\text{"}) + vec\,(\text{"}France\text{"}) \approx vec\,(\text{"}Paris\text{"}) \tag{3.2}$$

推論ベースの単語分散手法について，word2vec 以外には Fasttext[63] や Glove[64], また近年では BERT[65] などが報告されている．

本研究ではその中でも実装が用意で，単語分散表現モデルの構築が早い word2vec を用いた．そこで Word2Vec を用いた実装の種類について説明する．

Word2Vec では，ベクトル化の際にニューラルネットワークを用いるが，以下の 2 種類がある

1. CBoW(Continuous bag-of words)
2. Skip-gram(Continuous Skip-Gram Model)

以下でそれぞれについて，簡単に説明する．

1. CBoW

   CBow とは，周辺の単語から中心の単語を推定する手法である．

2. Skip-gram

   Skip-gram モデルは，文または文書内の周辺の単語を予測することで，単語ベクトルを得る中心の単語から周辺の単語を推定する手法である.

   そして，本論文の中で，Skip-gram モデルの方が CBoW よりも精度が良いと報告されていることから，本実験では Skip-gram モデルを用いることとした.

3.4 大規模言語モデル

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 など様々な技術と共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 知ることはできない．また，ChatGPT は（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 (図 3.1, 図 3.2).（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.）

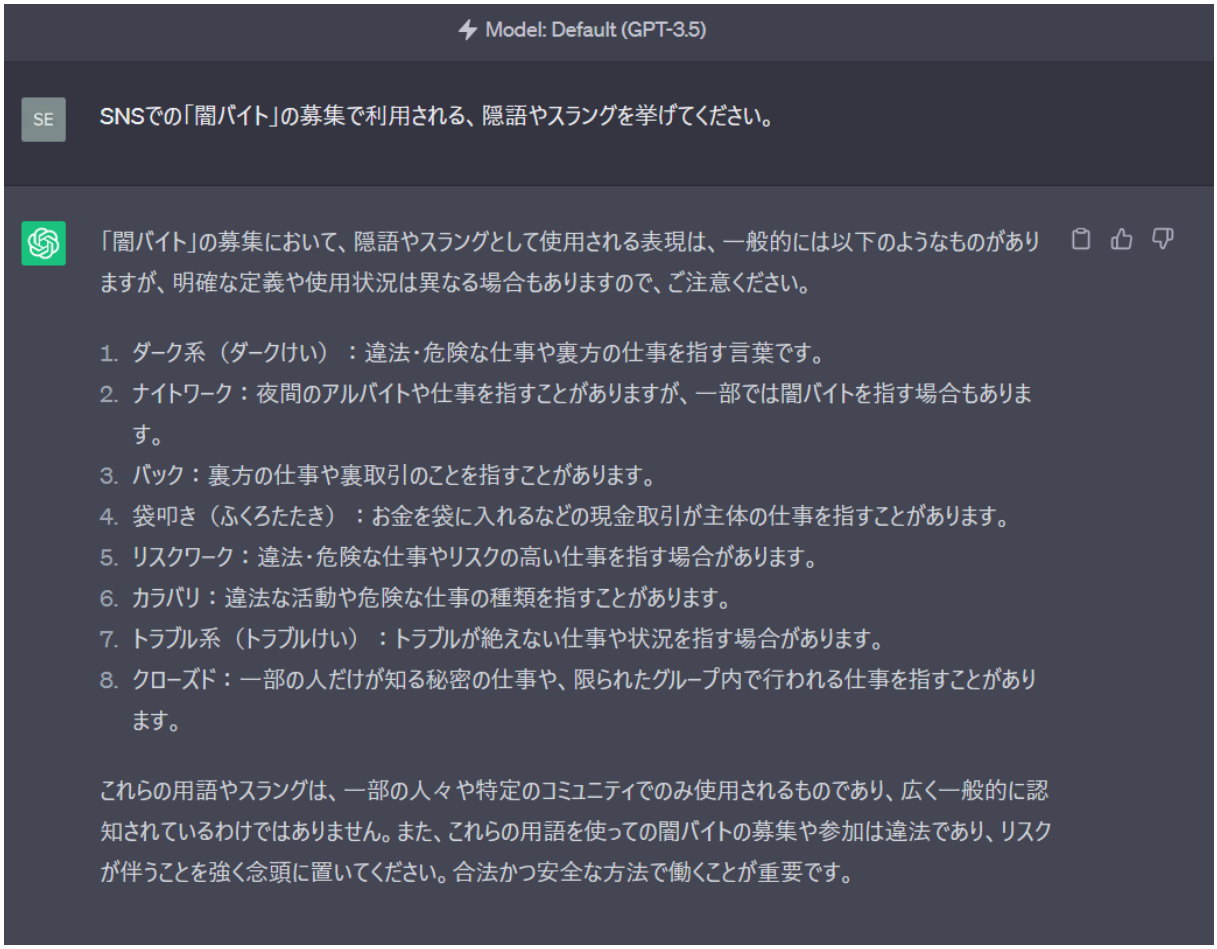


図 3.1: 闇バイトについて質問した回答 (GPT-3.5)

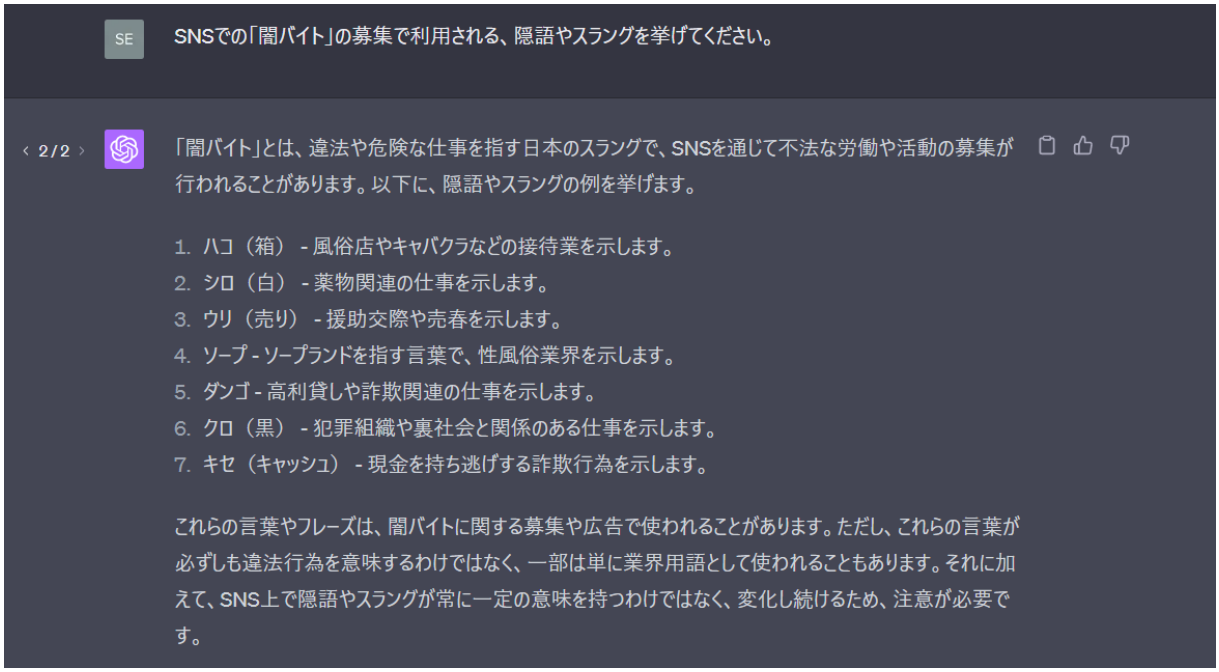


図 3.2: 闇バイトについて質問した回答 (GPT-4)

第4章　事前調査

本章では，隠語の特性について，隠語が変遷すると一般的に言われているところ，日本語においても近年に実際に発生しているのか調査した結果を説明する．まず，4.1 節で 2016 年と 2020 年の隠語の出現割合についての比較調査結果について説明する．続いて，4.2 節では一般的な単語に隠語としての意味が付与されるようになったものについての調査結果について説明する．

4.1　隠語の変遷について

1.1 章でも述べたとおり，言葉は時間と共に意味するものが少しずつ変化していく [16], [17] が，中でも犯罪の用途で用いられる隠語の特徴として，隠語の意味が一般的に認知されると新しい隠語が作られたり，今までは一般的な言葉の意味としてしか使われていなかった単語に，隠語の意味が付与されダブルミーニングで使用されたりする．[18]，[19].

また近年は SNS の発展に伴い，新しい隠語が生まれやすくなったと考えたことから，近年でも新しい隠語が発生したり，隠語が変遷したのかについて，実際に 2016 年と 2020 年のツイートデータを用意し，隠語の出現の割合について調査を行った．具体的には，2016 年 11 月から 2017 年 2 月と 2019 年 11 月から 2020 年 2 月までのそれぞれ同じ期間に TwitterAPI により収集したツイートデータのうち，以下の単語の出現割合について調査を行った．その結果，以下のとおりであった (表 4.1).

表 4.1 より，大麻を表す「ウィドー」,「クッシュ」，そして「パパ活」といった単語について，収集したツイートの範囲ではあるものの，2016 年のツイートデータからは隠語としてほぼ検出されなかった単語が 2020 年のツイートデータからは出現していることが見て取れる．

表 4.1: 期間内のツイートのうち，それぞれの単語が隠語として出現したツイート数

項番	単語	2016 年		2020 年	
		個数	全ツイートに対する割合	個数	全ツイートに対する割合
1	神待ち	204	0.0002%	845	0.0012%
2	パパ活	0	0%	10,100	0.014%
3	クッシュ	1	0.0000%	157	0.0002%
4	ウィドー	0	0%	70	0.0001%
5	野菜	0	0%	10,000	0.0144%
6	手押し	0	0%	1163	0.0017%
	取得したツイート数	111,408,818		69,301,877	

表 4.2: 期間の違いによる，一般的な単語が隠語として用いられた割合

	野菜		手押し	
	2016	2020	2016	2020
検出数	37,931	35,490	290	1,472
隠語として使用された数	0	894	0	1,163
隠語の割合	0%	2.5%	0%	81.8%

4.2　一般的な単語が隠語として用いられた割合

また野菜，手押しなど一般的に使われる単語について，2016 年と 2020 年のそれぞれにおいてどの程度隠語として用いられているのか調査を行った (表 4.2).

表 4.2 より，収集したツイートの範囲において，手押しも野菜も 2016 年時点では隠語としての意味を付与されて用いられた割合は 0%であったのに対し，2020 年には，手押しの場合は 81.8%，野菜については一般的な意味で出現する数が多いため，2.5%ではあるが，それまで全く隠語として用いられていなかった単語が，2020 年には隠語としての意味が付与されて使われるようになっていることが分かる.

このように，表 4.1，4.2 から，時間の経過とともに新しい隠語が使われるようになったり，一般的な単語の中に隠語の意味を付与させるような動きがあることが示唆される.

第5章　隠語検出手法の提案（提案手法１）

本章では，隠語検出を目的とした提案手法について説明する．まず 5.1 節にて，提案手法の前提を述べ，5.2 節にて，提案手法の概要について説明する．そして，5.3 節にて，隠語検出提案手法の中心アイデアについて説明する．続いて，5.4 節において，提案手法の流れとアルゴリズムについて説明する．なお，アルゴリズムについては Python で実装した．そして，5.5 節において，コアアイデアに加え，検出精度を向上させるために追加した機能について説明する．

5.1　提案手法の前提

本研究は，単語の類似語を元に，ツイートの中から隠語，さらには未知の隠語の検出を目指すものであるが，本研究の前提として，隠語を使って不正な取引を投稿するものも，その取引を検索しようとするものも共にいくつかの隠語や取引に使われるような不正な単語を知っており，それらの単語をキーに検索するという考えがある．

つまり，新しい隠語というものは突発的に出現するわけではない．なぜなら，あまりにも突拍子もなく隠語を出現させた場合，確かにサイバーパトロールからは逃れられるかもしれないが，一方で取引を検索しようとするものたちにも気づかれないため，取引が成立しないからである．そのため，これまで認知されていた隠語や不正な単語と共に出現することで連想等をさせて，最低限の認知はさせようとする．

11.3.1 節でヒアリングした中で，サイバーパトロールの実際の運用の場面では，対象のジャンルにおける不正な取引目的で使用される単語や隠語については，いくつかは認知しており，その単語を元にキーワード検索しているとのことであり，いくつかの単語を認識した上でその単語を手がかりに検索をするという意味では，実運用に即していると言える．

このような前提のもと，隠語の出現するツイートについては，以下の 4 種類があると想

定し，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（及び犯罪関連語）を基に，未知の隠語を検出することを目指す.

1. 既知の隠語（及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（及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（及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

5.2　提案手法の概要

本研究では，Twitterデータを対象に隠語を検出するため，隠語検出提案手法を提案し，システムを構築した.

本システムの主な入力データについては，以下のとおりである.

- 入力
  - Twitterデータ
  - 既知の隠語のうち,「野菜」,「氷」などダブルミーニングのものを除いた単語群（以下,「犯罪語リスト」という）
  - 既知の隠語の単語群（以下,「照合リスト」という.）
- 出力
  検出した隠語の一覧（以下,「検出隠語リスト」という.）

このうち，Twitterデータから犯罪語リストを元に不正な目的のみのツイート群のコーパス（以下,「Badコーパス」という.）を抽出し，そのコーパス内の単語，すなわち隠語の可能性がある単語群を抽出する（以下,「入力単語リスト」という.）.

この入力単語リストに基づき，一つ一つの単語に対し，Badコーパスと別途用意する一般的な用途で使用される単語群（以下,「Goodコーパス」という.）という二つのコーパス

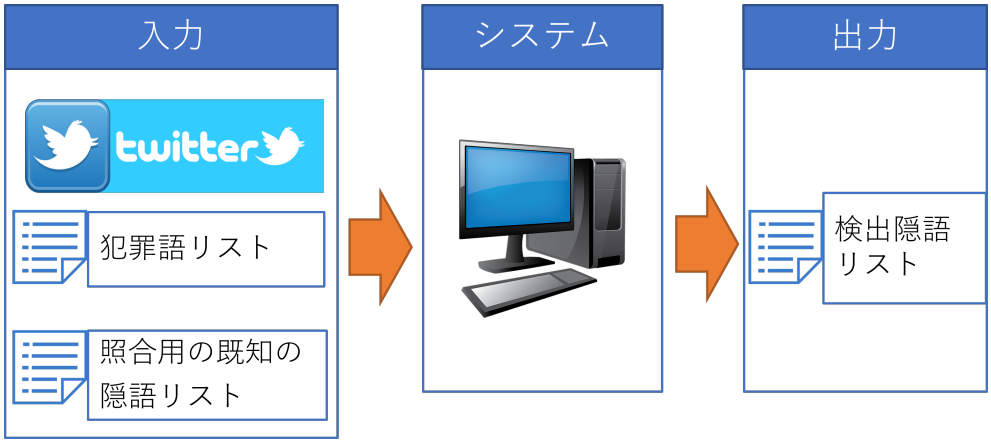


図 5.1: システムの入出力

のそれぞれで類似語上位 N 位まで求め，その各類似語に対し照合リストの隠語と照合・結果を比較し，隠語の一覧を出力する方法が提案手法となる（図 5.1）.

5.3 隠語検出手法の中心アイデア

犯罪を計画する者が隠語を用いて，いかに巧妙に犯罪の意図をカモフラージュしたとしても，前後のやり取りの文脈性の変化は少ないと考えられる．なぜなら，取引自体は迅速に成立させる必要があるからである．また隠語の周りには特徴的な単語が多く現れることから [37]，違法な取引に使用される単語は，その類似語も同様の意図で使われていると考えた．単純に，単語群の類似語から隠語を検出することは難しい．なぜなら，類似語を元に検出しようとしても，隠語には，そもそも出現頻度の低い珍しい単語や造語が使われていたり，前述した隠語例からもわかるとおり,「野菜」,「アイス」など日常的に使われる単語でカモフラージュされているものもあることから，隠語を検出するために,「野菜」,「アイス」などといった単語の類似語を検索したとしても，一般的な野菜に関連したツイートにまぎれてしまうからである．そこで，用途の異なるコーパス群を複数を用意し，それぞれで類似語を求めれば，同じ単語であっても異なる類似語が抽出でき，その単語の差異から隠語が検出できると考えた.

そこで，用意したツイートデータを以下の方法で二つのツイート群 (Bad コーパスと Good コーパス) に分類する.

1. Bad コーパス

   5.1 節の分類のうち，1 と 3 の分類の位置づけであり，不正な取引目的で使用されていた単語が含まれるツイート群と定義し分類した． なお，5.1 節で説明したとおり，不正な取引を投稿するものもその取引を検索しようとするものも共に，いくつかの隠語や取引に使われるような不正な単語を知っていることが前提であり， 今回は，犯罪語リストのうち，いずれかの単語が一つ以上出現するツイート群とした．犯罪語リストに用いた単語は，各実験で対象とするジャンルの単語（援助交際，違法薬物取引，闇バイト等）に関連するものを選定した．

2. Good コーパス

   5.1 節の分類のうち，4 の一般的な単語のツイート（既知の隠語も未知の隠語も利用されていないツイート）群の位置づけであり，不正な取引目的で使用されていた隠語の出現率が極めて低いものと定義し，分類した．作成方法として，次の二つの方法が考えられる．

(a) 良いツイートを抽出する方法

i. 作成方法

   全体のツイートから Bad コーパスへ分類したもの以外の全ツイートを Good コーパスに分類する．

ii. 利点

   単語分散表現モデルやパラメータを自分で自由に設定可能であり，簡単に，そして柔軟に作成できる．

iii. 欠点

   Bad コーパスで悪い意図のツイートを網羅できていない場合，Good コーパス内に悪い意図のツイートが紛れ込むおそれがあり，影響が少ないかもしれないが，Good コーパスの純度が下がるおそれがある．コーパス規模を広げた場合，単語分散表現による処理に膨大な時間を要する．実際に 3 か月規模のコーパス（約 10Gbyte）の作成を試み，一般的なデスクトップ端末の性能で 10 日以上の処理を要した．そのため，実運用を想定した際に，大きなリスク要因と捉えられる．

(b) 大規模なツイートコーパスを利用する方法

i. 作成方法

Twitter 上では不正な取引に関するツイートの数は全体から見ると無視できる数であると考え，既に作成されている SNS における一般的なコーパスモデルを使用する [66]．なお，当該コーパスを用いての類似語検出実験を行ったところ，検出結果は表 5.1 のとおりであり，少なくとも上位 10 件においては一般的な単語のみ出現していることを確認した．

ii. 利点

大規模コーパスを簡単に，また迅速に利用することができる．そのため，前述した「良いツイートのみを抽出する方法」により，自身で作成する際に実運用上，大きな問題と記載した時間について，本方法であればクリアできる．

iii. 欠点

作成しているモデルに依存する．単語分散表現が用意されていなかったり，パラメータが設定が自由に変更できないおそれがある．また，提供元が少ない．

この二つの方法について，上記のとおり整理したところ，方法 (a) については，コーパス作成に膨大な時間を要することから，実運用上，大きな問題と考えたが，方法 (b) でこの問題が解決できることから，方法 (b) を採用することとした．

そして，その後は，Word2vec[20] を用いて，それぞれのコーパスの単語分散表現モデルを生成した．そして，二つのコーパス間では同じ単語であっても，単語の用いられ方が異なることから，類似語が異なるとの予測のもと，Python のライブラリである gensim[67] を用いて，それぞれのコーパスから生成した単語分散表現モデルを元に，それぞれの単語分散表現モデル内に存在する同じ単語に対して，コサイン類似度に基づく，類似度上位の単語を調査した．たとえば，覚醒剤の一種である「LSD」の隠語である「紙」という単語のそれぞれのコーパスにおける類似語を調べたところ，表 5.2 のとおりとなった．ここにおける Good コーパスの作成方法は,「良いツイートのみを抽出する方法」とした．

表 5.2 から，同じ単語であっても二つのコーパスで全く異なる単語が検出されること，さらには Bad コーパスで構成されたモデルの類似語からは隠語が多く検出された (表 5.2 中の太字は隠語と判断した単語) ということがわかった．

表 5.1: ホットリンク社の大規模 SNS コーパスによる類似語結果

順位	手押し		アイス		氷		野菜		冷たい	
1	手押	0.8652	アイスバー	0.8193	氷雪	0.7952	キャベツ	0.8972	冷たく	0.8406
2	手押し車	0.8407	ロックアイス	0.8155	氷ら	0.7810	果物	0.8969	冷たき	0.8079
3	手押しポンプ	0.7695	アイスター	0.7914	凍っ	0.7754	豆類	0.8669	温かい	0.8052
4	軌道自転車	0.7355	アイスタ	0.7854	氷塊	0.7676	トマト	0.8666	暖かい	0.7928
5	トロッコ	0.7268	アイスティー	0.7723	凍ら	0.7594	キュウリ	0.8659	冷た	0.7820
6	起重機	0.7263	アイススター	0.7690	海氷	0.7582	ニンジン	0.8628	熱い	0.7677
7	車輪	0.7166	アイスノン	0.7591	氷筍	0.7564	ジャガイモ	0.8618	生暖かい	0.7675
8	荷車	0.7046	アイスバイン	0.7574	浮氷	0.7550	軟弱野菜	0.8537	冷やさ	0.7464
9	人力	0.7001	レジアイス	0.7572	氷山	0.7537	根菜	0.8492	冷や	0.7371
10	馬車	0.6972	アイススケートリンク	0.7560	融け	0.7455	タマネギ	0.8480	吹き出す	0.7301

これより，二つのコーパス間で同じ単語にも関わらず，検索される類似語が大きく異なるという点と，Bad コーパスで隠語の類似語を検索した場合，似たような隠語や関連する不正な取引に使われる単語が出現するのではないかという二つの点に着目し，未知の隠語の発見を目指す.

5.4　隠語検出手法の流れ

隠語検出手法の詳細な流れは以下のとおりである.

1. TwitterAPI を利用し，ツイートデータを収集し，適切な前処理（詳しくは 6.2 を参照）を行う.
2. 犯罪語リストを元に，Bad コーパスを作成する.
3. Bad コーパスの単語分散表現モデルを生成する．なお，Good コーパスについては，既に生成された単語分散表現モデルを使用する.
4. 単語分散モデルを基に，Bad コーパス内の単語を抽出し，入力単語リストを作成する.
5. コアアルゴリズムの説明 (Algorithms 1，2)

表 5.2: 「紙」における各コーパスの類似単語 (上位 10 位)

Good コーパス		Bad コーパス	
1	字詰め	1	業販
2	試筆	2	市内
3	便箋	3	営業中
4	裏紙	4	メニュー
5	ハードカバー	5	**スカンク**
6	アルシュ	6	**リキッド**
7	用紙	7	**ノーザン**
8	断裁	8	**グミ**
9	模造紙	9	**ハイレギュラー**
10	方眼	10	**ヘイズ**

(a) 入力単語リストの各単語について，二つのコーパスのそれぞれで照合リストと *Hit* した類似語数（スコア）を計算し (Function ***SIMILAR***)，同じ単語の二つのコーパス間のスコアを比較し一定以上のスコアの差があり，かつ Bad コーパスにおいて一定以上のスコアであれば，当該単語を隠語と判定する (***Main***)（図 5.2a）.

(b) 各単語は，構築した単語分散表現モデル (Good_Corpus，Bad_Corpus) を使ってコサイン類似度上位 N 位までの類似語を検索する (*Get similar words*).

(c) N 個の類似語について，一つずつ照合リスト (Codeword_List) 内の単語と照合させる.

(d) もし照合リスト内のいずれかの単語と合致した場合，スコアを加点する (X=X+1). つまり今回の実験では最大で N 点となる.

(e) 照合リストのどの単語とも合致しなかった場合，照合リストにはまだ登録されていない未知の隠語である可能性を考慮し，その単語を元にコサイン類似度上位 $N/2$ 位までの単語を検索し，再帰的に ***Main*** を実施し，スコアを求め隠語かどうか判定する（図 5.2b）.

(f) その類似語のうち，照合リストに合致しなかった単語についても，さらに $N/4$ 個の類似語を検索し，スコアを求め隠語かどうか判定する.

Algorithm 1 ***Main***

Input: Word_List,N,Good_Corpus,Bad_Corpus
Output: Codewords
  for all $Word$ in $Word\_List$ **do**
    $Cnt\_Bad \Leftarrow SIMILAR(Word, N, Bad\_Copus, 1)$
    $Cnt\_Good \Leftarrow SIMILAR(Word, N, Good\_Copus, 1)$
    $Diff \leftarrow abs(Cnt\_Bad - Cnt\_Good)$
    if $(Cnt\_Bad/N >= \alpha)$ or $((Diff/N >= \beta)$ and $(Cnt\_Bad/N >= \gamma))$ **then**
      $Codeword\_List.append(WORD)$
    end if
  end for
  $return(Codeword\_List)$

5.5　精度向上方策及び汎用性についての検討

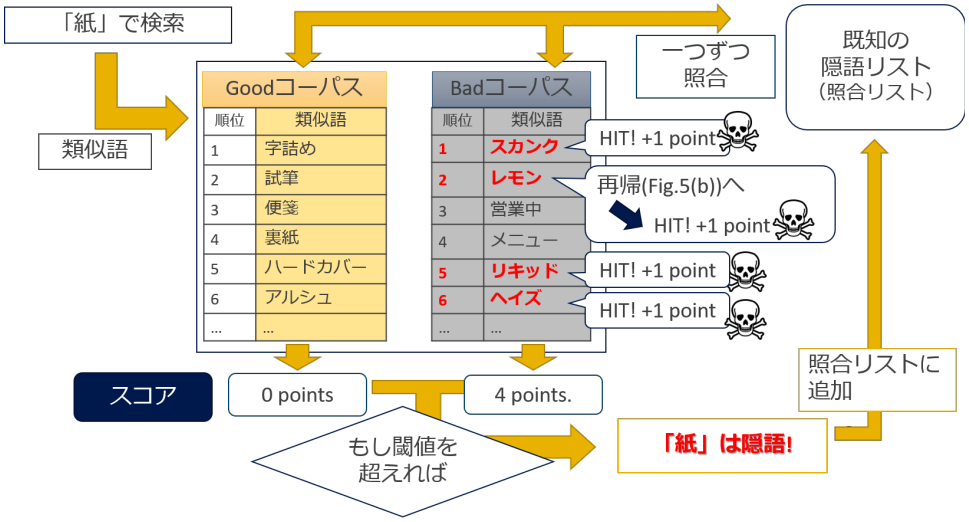
コアアイデアに加え，さらに隠語検出の精度を向上させるため，いくつかの機能を検討及び検証し，以下の 2 つの機能を追加する．

- 犯罪関連語の検出
- 品詞分類によるフィルタ

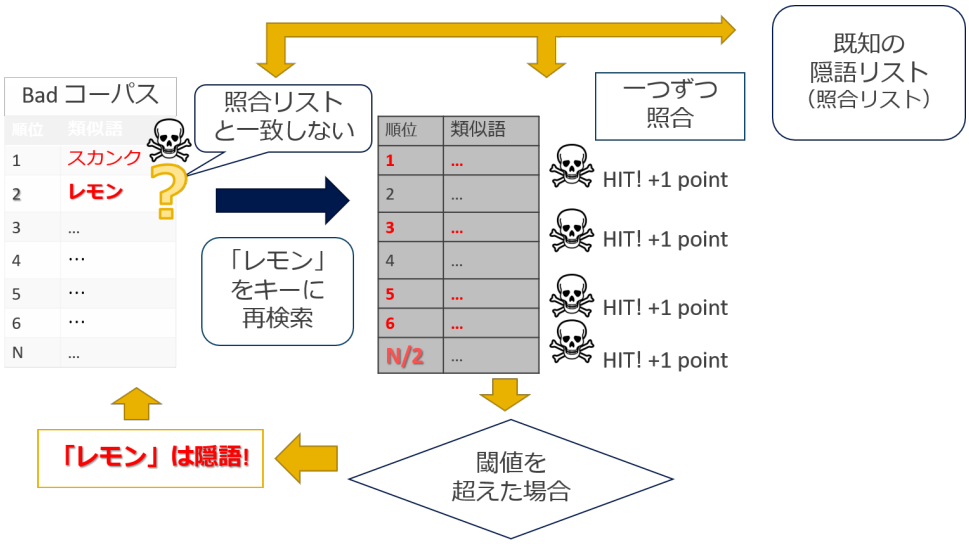
5.5.1　犯罪関連語の検出

犯罪関連語については，2.2 節でも述べているが，隠語を用いて巧妙に意図を悟られずに迅速に違法な取引を成立させるためには，最低限，取引に必要な情報を含めなければならないため，隠語だけでなく,「取引対象」,「取引対象を形容するもの（高品質等）」,「時間」,「場所」,「取引方法」,「取引量」,「金額」などに当たるその単語自体だけでは隠語として成立しないが，隠語と一緒に出現する傾向が高い犯罪関連語も隠語と同じ文章内に含まれている必要があると考えられる．

隠語について，たとえば，覚せい剤の隠語である「アイス」について，Bad コーパスから作成した単語分散表現モデルにおける類似語の確認を行ったところ，表 5.3 のとおりとなった．ここで，単語については「隠語」,「犯罪関連語」，それ以外の「無関係」の 3 つの区分に分類した．



(a) 隠語検出アルゴリズム



(b) 隠語検出アルゴリズム (照合リストと一致しなかった場合)

図 5.2: 隠語検出アルゴリズム

表 5.3 より,「アイス」の類似語の中には,「隠語」は上位 10 位までのうち，2 個と少ないものの，犯罪関連語である「郵送」や「営業中」といった単語が上位 10 位までのうち，7 個確認された.

そのため，犯罪関連語も隠語検出において，対象とする仕組みを導入することで，検出精度を向上させることが期待できると考え，追加実装を行った．なお，照合リストのうち，1 つの隠語が HIT した際，1point 加算するところ，犯罪関連語については，犯罪関連語用

Algorithm 2 Function ***SIMILAR***

Input: Word,N,Corpus,Loop_count
Output: Number of matches with codewords

```
X ⇐ 0
Sim_words ⇐ Corpus.Get_similar_words(Word, N)
for all Sim_word in Sim_words do
  if Sim_word in Codeword_List then
    X ⇐ X + 1
  else if Loop_count <= 2 then
    Y ⇐ SIMILAR(Sim_word, N/2, Corpus, Loop_count + 1)
    if Y/N >= δ then
      X ⇐ X + 1
    end if
  end if
end for
return(X)
```

照合リスト (以後,「グレーリスト」という．) を用意し，1/2point の HIT とした (5.4)．また追加実装にあたっては，5.5.2 節の機能と合わせて組み込んだ．

5.5.2　品詞分類によるフィルタ

コアアイデアだけでは，隠語と判定した単語については，スノーボール方式で自動的にすべからく照合リストに追記していく．そのため隠語か犯罪関連語に自動的に分類することで，犯罪関連語と隠語とを区別して自動的に追加可能となると考えた．

そこで，隠語と犯罪関連語を分類するための特徴を把握するため，形態素解析器である Sudachi[68] を用いて，ツイートデータのうち 191,079 語（Good コーパス（184,212 語）＋ Bad コーパス（6,867 語））を対象に，品詞分類を行い分析した．

単語の属性について分析した結果，隠語として識別した単語が主に名詞句であったことから，名詞に絞ることで精度が向上すると考え，さらに分類を行った．分類方法として，網羅的に単語を確認した結果，名詞の中で原則として全て隠語として照合リストに追記することとすることとした．ただし，項目によっては隠語対象からの除外もしくは隠語ではなく犯罪関連語への分類とした．各品詞の分類及び考察については表 5.5 のとおりである．

表 5.3: 「アイス」の類似語（上位 10 個）

順位	単語	分類
1	市内	犯罪関連語
2	郵送	犯罪関連語
3	営業中	犯罪関連語
4	野菜	隠語
5	極上	犯罪関連語
6	業販	犯罪関連語
7	ブラック	隠語
8	おはようございます	無関係
9	メニュー	犯罪関連語
10	テレ	犯罪関連語

そして，表 5.5 を元に，隠語候補として検出された隠語と犯罪関連語を品詞分類し，それぞれを自動的に追加登録可能にした．これにより，Precision を向上させることが期待できると考え，追加実装を行った．

表 5.4: グレーリストを導入した隠語検出アルゴリズム

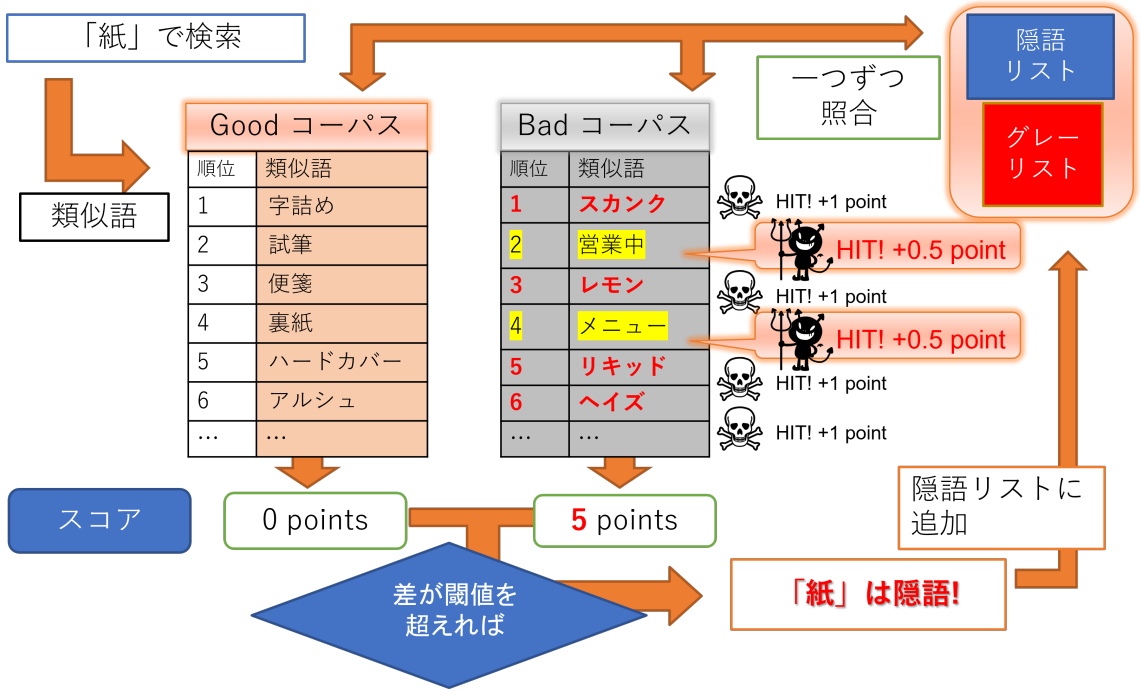


表 5.5: 品詞分類による分類表

大項目	中項目	小項目	例	分類	理由
名詞	固有名詞	地名	カルフォルニア，名古屋	犯罪関連語	犯罪行為に関わる取引きでは，お互いを信用することが難しいため，対面での取引が行われることが多い．また居住地から遠い場所での取引はコストと時間がかかるため，事前に場所の指定がなされることが多い．地名は場所の指定に用いられることになるため，もしこれが隠語として別の意味を持たせる場合，その別の意味のものを指すのか，場所の指定を指すのか混乱が生じる．したがって隠語にはなりづらい一方，隠語とともに出現することは多いと考えられる．
		一般	バイアグラ	犯罪関連語	
		人名	ホフマン，ジャック	隠語	
	普通名詞	一般	氷，クリスタル，葉っぱ	隠語	
		サ変可能	絶賛，宅配，味見	犯罪関連語	本分類の単語は取引対象自体を指すものではなく，取引対象の質や取引に関する情報を指すものが多い傾向にあった．そして，取引対象を形容するために使われる単語が取引対象を指す意味を持ってしまうと，取引対象を形容するために使われているのか，取引対象自体のどちらを指すのか混乱するため，取引対象を指す隠語にはなりづらい．
		サ変形状詞可能	安心，満足，直接	犯罪関連語	
		形状詞可能	安全，好評	犯罪関連語	
		助数詞可能	キロ，ドル，袋	無関係	これらの単語は取引においても使われるので，幅広にやり取りを行う Twitter では一意に特定の取引対象を比喩することは難しことから，隠語にも犯罪関連語にもなりづらい．
		副詞可能	ただいま，朝方，来週	無関係	
	数詞		二，十	無関係	

第6章　実験1(隠語検出実験)

本章では，隠語検出プログラムを実装し，実験を行った結果を示す．まず6.1節では，実験の概要を，6.2節では，実験のプロセスについて説明する．続いて6.3節では，評価で用いる評価指標を説明する．そして，6.4節では，比較手法について説明し，そして6.5節で実験の結果を示す．最後に，6.6節で評価結果をまとめ，考察する．

6.1　実験の概要

入力単語リストとして，Badコーパスに含まれる単語のうち，出現頻度が20回以上の単語1,892 語を用いて，隠語検出の実験を行った．なお，これらの単語については事前にアノテーション済みである（詳細は6.2.6節を参照）．

具体的には，1,892語に含まれる45語の隠語のうち，10語を既知の隠語つまり「照合リスト」とする．そして，提案手法を用いることで残りの35語の検出を目指す．実験の工程については，次節で記載する．なお，前処理の内容は6.2.2節に記載している．

6.2　実験のプロセス

実験において図6.1の流れで処理を実施した．

6.2.1　データ収集

TwitterAPIを利用し，ツイートデータを47日間分収集した（5.4GByte）し，本文データのみを使用した．

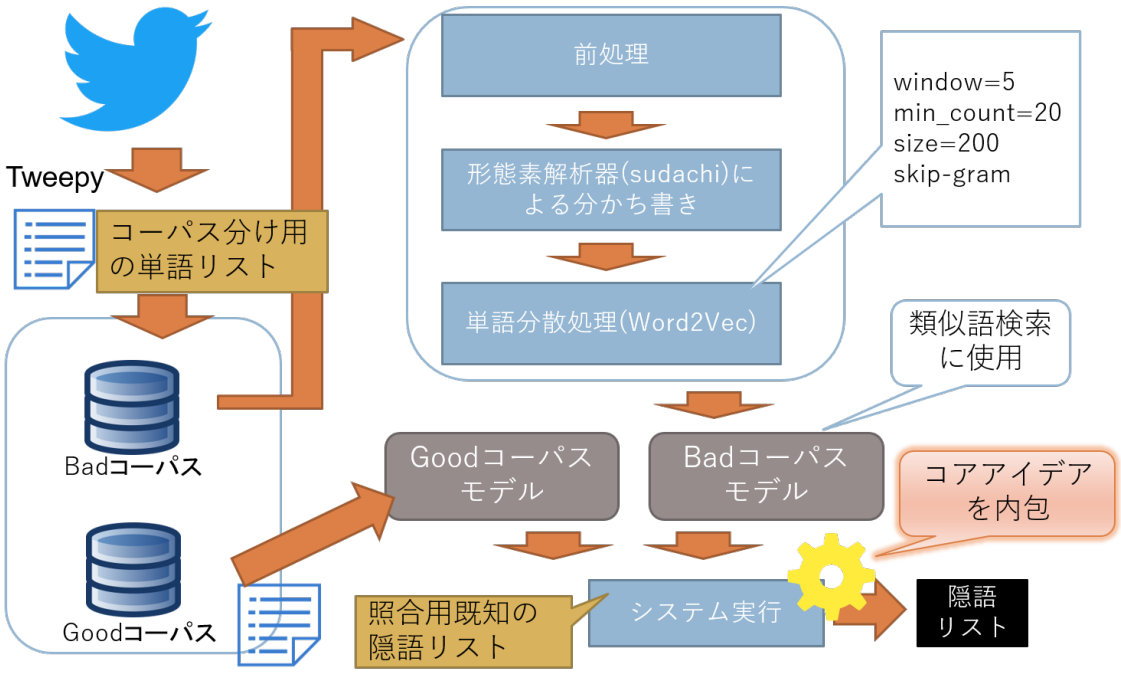


図 6.1: 実験プロセス

6.2.2　前処理の実施

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

1. 半角英数字
2. URL
3. 全角記号
4. 改行文字
5. Twitter に定型でよく現れる単語
   「RT」「まとめ」「お気に入り」

6.2.3　コーパス作成

用意したコーパスについては，以下の二つである．

1. Bad コーパス（8MByte）
   本実験では，前処理が完了したツイートデータ群を一文ずつ二つのジャンル（違法薬物売買，援助交際）における不正な取引目的で使用されていたと判断した 10 個の単語と照合させ，いずれかの単語が一つ以上出現するツイートを収集して Bad コーパスを作成した．なお，一般的なツイートについては，除外している．

2. Good コーパス
   5.3 節で説明した定義について，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 [66].

6.2.4 形態素解析

日本語は特有の文章構造を保有しており，英語等と異なりスペース等で区切られないため，単語分散処理を行う前に，形態素解析処理及び分かち書きが必須である．分かち書きとは，事前に単語の辞書を内部で持ち，それに従い，文章を適切に単語単位に分割するものである．これによって，単語単位で文章を分けることができる．ここで問題となるのが，いかに適切に文章を分けることができるかということである．なぜなら，今回，マイクロブログの中でも Twitter を対象としたが，その特徴として，短文であり，新語やスラングが多く，意図的に文章を切っているものも多くみられる等の特徴のため，正しく分かち書きされないおそれもあるからである．また，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，正しく分かち書きがされる必要がある．

これらのことから，以下の二つの理由から形態素解析器として Sudachi[68] を採用した．

1. 内部辞書が定期的に更新されており，新語にできる限り対応している保守の観点

2. 新語を想定し，単語の分割単位を選択できるという観点

6.2.5 単語分散表現処理

形態素解析処理の実施後，Word2vec[20] を用いて，単語分散表現処理を実施した．
Word2vec のパラメータは表 6.1 のとおり設定した．

表 6.1: Word2vec のパラメータ

パラメータ	設定値
size	200
min-count	20
window	5
Skip-Gram or CBow	Skip-Gram[69]

表 6.2: 本実験における Algorithms 1， 2 での可変的なパラメータ設定値

パラメータ	設定値
N	20
α	0.2
β	0.15
γ	0.1
δ	0.2

6.2.6　入力単語リストの作成

二つのコーパスから作成された単語のうち，二つのコーパスで共通して出現した単語（1,872 単語）と Bad コーパスのみに出現した単語（10 単語）を抽出した．

それらの単語を，評価を行うため，隠語に関する知識を有していない 3 名により，対象の単語が出現するツイート本文を確認した上で，2.2 節で説明した 3 種類（隠語・犯罪関連語・一般的な単語（それ以外））へ分類した．

そして，作成した入力単語リストを元に実装したシステムを実行した．

6.2.7　提案システムの実行

Algorithms 1， 2 における可変的なパラメータについて，本実験では，表 6.2 のとおり，設定した．なお，閾値については，フレキシブルに変更可能に設計してあり，本実験では検出精度が一番良い値を閾値として決定した．

6.3 評価指標

評価について，以下の 4 つの指標を用いて評価を実施した．なお，式中には，真陽性 (TP)，偽陽性 (FP)，偽陰性 (FN)，真陰性 (TN) で表す．

1. Precision
   適合率と呼ばれるもので，正と予測したデータのうち，実際に正であるものの割合で求める．計算式は，数式 (1) のとおりである．

$$Precision = \frac{TP}{TP + FP} \tag{6.1}$$

2. Recall
   再現率と呼ばれるもので，実際に正であるもののうち，正であると予測されたものの割合で求める．計算式は，数式 (2) のとおりである．

$$Recall = \frac{TP}{TP + FN} \tag{6.2}$$

3. Accuracy
   正解率（精度）と呼ばれ，正や負と予測したデータのうち，実際にそうであるものの割合計算式は，数式 (3) のとおりである．

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6.3}$$

4. F-measure
   F 値と呼ばれ，Precison（適合率）と Recall（再現率）の加重調和平均として定義される．計算式は，数式 (4) のとおりである．

$$F - measure = 2\frac{Precision * Recall}{Precision + Recall} \tag{6.4}$$

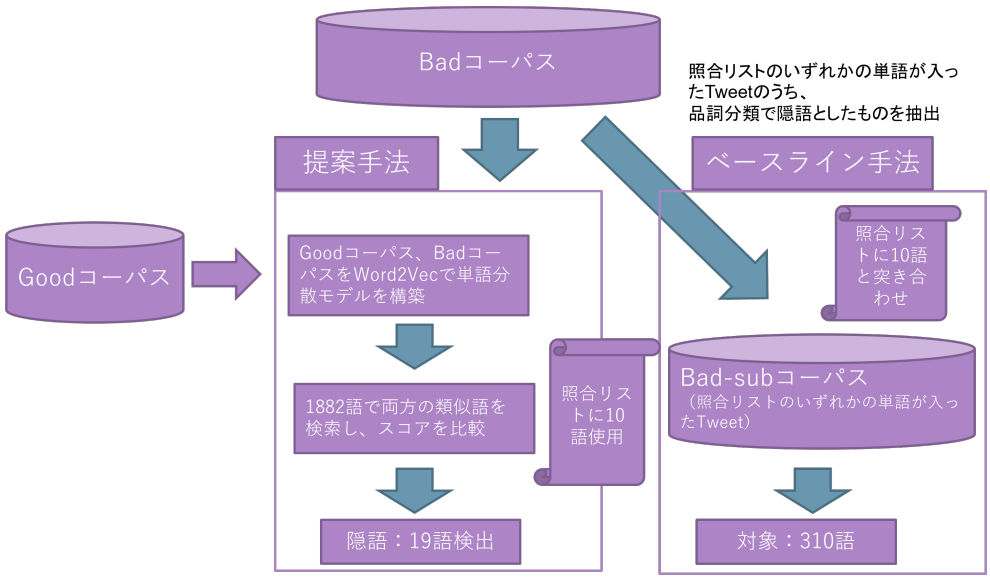


図 6.2: 提案手法とベースライン手法との関係図

6.4　比較手法

提案手法による効果を検証するため，比較手法を用意した（以下,「ベースライン手法」という）．本研究における提案手法は，不正な取引に使用される単語の周りには似たような目的に使用される単語が現れるとの仮説の元，類似語に着目している．そこで，ベースライン手法では，不正な取引に使用される単語が出現したツイートのうち，提案手法に使用した表 5.5 の品詞分類に基づき，隠語を抽出した．提案手法とベースライン手法の関係は図 6.2 のとおりである．

ベースライン手法による，隠語の検出方法は以下のとおりである．

1. Bad コーパスに対し，提案手法で用いたものと同じ照合リストを照合させる．
2. 照合リストのうち，いずれかの単語が含まれた文章を全て抜き出し，Bad-Sub コーパスを作成する．
3. Bad-Sub コーパスの中で，提案手法の品詞分類の抽出手法を用いて，表 5.5 に基づき，隠語を抽出する．

表 6.3: 評価結果

分類	全単語		提案手法		ベースライン手法	
	個数	割合	個数	割合	個数	割合
隠語	35	1.9%	19	55.9%	23	7.4%
隠語以外	1,847	98.1%	15	44.1%	287	92.6%
合計	1,882		34		310	

表 6.4: 結果の詳細

評価手法	提案手法	ベースライン手法
Precision	0.559	0.074
Recall	0.543	0.657
Accuracy	0.984	0.682
F-measure	0.551	0.133

6.5 実験結果

アノテーションの結果，隠語は 45 語あり，そのうち，10 単語を照合リストとして用意した．

その照合リストとして用いた 10 単語を除いた 1,882 語を入力単語リストとして，システムを実行した．その結果，隠語として，34 語検出され，そのうち 19 語の隠語が含まれていた．

提案手法とベースライン手法の結果は表 6.3，6.4 のとおりであった．

表 6.3，6.4 より，表 Precision（適合率)，Accuracy（正解率），F-measure（F 値）において，提案手法はベースライン手法と比べ，より優れた結果を得ることができたことがわかった．

なお，提案手法において，検出できた隠語の例としては「ディーゼル」,「ジョイント」などがあった．また，通常は一般的に用いられる「レモン」,「アイス」,「スカンク」,「グミ」等の隠語も検出できていた．

検出された単語の詳細については，付録付録. B 節のとおり．

6.6　考察

6.6.1　精度向上に向けて

今回，本提案手法を用いることにより，隠語検出についてベースライン手法に比べ，Precision において 0.485 高い結果を出すことができた．一方で，出現頻度が 20 回以上を対象にアノテーションしていたため，20 回未満の隠語（たとえば,「バブルガム」等）を検出できていなかった．

この理由として，単語分散表現を獲得する際の Word2vec[20] のパラメータのうち，n 回未満登場する単語を破棄する「min_count」オプションの値を 20 回としていたため，これらの単語は出現頻度が設定値を下回っていたことから，分散表現モデル生成時に破棄されたことが原因と考えられる．

今回の実験では，ノイズを減らすため，設定値を 20 で分散表現モデルを生成したが，本研究では隠語と対象としており，隠語の出現割合は表 4.2 でも言及したとおり，たとえば，野菜については 2.5%と低かったことから，元々の出現数は少ないことが考えられる．そのため，出現頻度の閾値については，検証の結果，4 回未満とし，今後の実験ではこの値を採用することとした．

6.6.2　犯罪語リストのジャンルについて

隠語の出現するツイートを確認する中で，共起する隠語や犯罪関連語がジャンルによって異なることがわかった．大麻などの薬物取引関連の単語においては，たとえば「野菜」と一緒に「手押し」や「高純度」などが共起することが多いが，援助交際関連の単語として，たとえば「神待ち」では「手押し」や「高純度」といった単語とは共起せず,「諭吉」,「苺」,「パパ」,「JK」などといった単語が共起する頻度が高いことがわかった．提案手法は，犯罪語リストを元に Tweet を抽出し，それを元に単語分散表現モデルを構築し，モデル内の単語の関連性を元に隠語リストを元に隠語を検出する．そのため，犯罪語リストや隠語リストを細かく調整することは可能であり，また実験 1 では，二つのジャンル（違法薬物売買，援助交際）における不正な取引目的で使用されていたと判断した 5 個ずつの単語を用いたが，この二つのジャンルは共に隠語が共起しあう関係ではないことから，ジャ

表 6.5: 機能の有無による差異

評価手法	条件 [1]	条件 [2]	条件 [3]
Precision	0.559	0.613	0.112
Recall	0.543	0.543	0.600
Accuracy	0.984	0.994	0.907
F-measure	0.552	0.576	0.189

[1] 今回の提案手法
[2] 品詞分類によるフィルタ機能のみ（犯罪関連語検出機能を含まない）
[3] 品詞分類によるフィルタ機能，犯罪関連語検出機能の両方の機能を含まない

ンル（薬物取引や援助交際等）を分けて，コーパスを作成し，隠語リストを用意したほうがより精度が上がると思料される．

6.6.3 追加機能の効果の検証

今回の実験では，コアアイデアに加え，5.5 に示したとおり，隠語検出精度が向上する機能を検討及び検証を行い，効果が確認された二つの機能（5.5.2 節で示した品詞分類によるフィルタ機能，5.5.1 節で示した犯罪関連語検出機能）についても共に実装した．そこで追加した機能の実験における効果について，検証を行った．今回は以下の条件で比較した．

1. 今回の提案手法
   品詞分類によるフィルタ機能及び犯罪関連語検出機能の二つが含まれている．

2. 品詞分類によるフィルタ機能のみ (犯罪関連語検出機能を含まない)

3. 品詞分類によるフィルタ機能，犯罪関連語検出機能の両方の機能を含まない

機能の有無による結果は表 6.5 のとおりである．

これより，追加機能が効果的に精度を向上させていたことがわかった．ただし，提案手法の方が，品詞分類によるフィルタ機能，犯罪関連語検出機能の両方の機能を含まない条

件である [3] より Recall が低かった．これは，品詞分類機能を追加しないことで，検出された単語がすべからく照合用の隠語リストに追加されることで，照合用の隠語リストが増えることとなり，HIT しやすくなったことが考えられる．その結果，機能無し条件の方が，隠語が提案手法に比べ隠語が多く検出された．ただし，F スコアでも大きく差もあることから，追加機能が効果的に作用していることがわかった．

また犯罪関連語検出機能については，今回の実験については，本機能を搭載した提案手法より，搭載しない条件の方が若干ではあるものの良い結果となった．これについて結果を確認したところ，提案手法では，検出した単語の中で正解の隠語の数は同数であったものの，グレーリストにより搭載しない条件よりも多く検出できた単語が隠語ではなかったため，精度が落ちたことが原因であった．ただし，それらの単語についても全く無関係の単語ではなく，犯罪関連語ではあった．そのため，犯罪関連語も含めた結果では提案手法の方が良い結果となっていた．このようなことから，犯罪関連語の検出機能は有効であると考える．

また，品詞分類については，設定した分類次第で見落としてしまう隠語があるおそれもあるので慎重に設定していく必要がある．また別の事前実験では犯罪関連語検出機能も大きく効果を出していたことも確認していることから，共起性をさらに考慮することで，より効果的に機能することが期待できる．

6.6.4　複合語型隠語の検出について

前処理として分かち書き作業を実施した中で，複合語が分割されるという課題があった．その分かち書きを元に単語分散表現を構築し，構築した単語分散表現に内包する単語を元に隠語を検出するシステムであることから，分かち書きの時点で複合語が分割されていると複合語が検出できないこととなる．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．これらの単語については，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」というように文節で切れてしまい,「ゴリラ」が検出されることはあっても「ゴリラグルー」として検出ができていなかった．

複合語型隠語が分割される対策として，形態素解析器における分かち書きの文節単位を

調節することが考えられるところ，設定次第で一般的に認知されている複合語については，複合語の単位で適切に文節が区切られることが可能ではあるが，一般的に認知されていない造語も含まれる隠語の場合は，文節が区切られてしまう．特に複合語型隠語が一般的な単語を含むものであれば，なお文節は区切られてしまう傾向にある．たとえば，大麻として使われていた「レモンスカンク」についても,「レモン・スカンク」のように文節が区切られてしまった．

本手法では，構築された単語分散表現モデル内に内包された単語をベースに隠語を判定するシステムであり，分かち書きの結果,「レモンスカンク」として区切られていなければ，単語分散表現の中に存在しないことになるため，隠語として判定されることがない．

一方で，形態素解析器内部で用いられる辞書に登録されていれば分かち書きは可能だが，本研究で検出対象とする隠語は当然ながら一般的な辞書には登録されておらず，さらに文節の中に「レモン」のような一般的な語が含まれている場合，文節が区切られやすいことが原因として考えられる．そのために，事前に複合語を辞書登録し，その単位で文節を区切るようにすれば，課題が解決でき，複合語型隠語を検出できるようになると考えられる．しかしながら，実際の環境における複合語型隠語は，登録すべき単語が不明であると思われる．また辞書が更新されたとしても，複合語型隠語の場合，造語や認知度の低さの点から，辞書に追加される可能性が低いと思われる．これまでに複合語隠語の検出を課題とした研究はないことから，複合語型隠語発見の観点からも複合語を自動的に検出できることが望ましいため，課題となる．

第7章　複合語型隠語検出手法の提案（提案手法２）

本章では，6.6.4 節でも述べたとおり，6 章の隠語検出実験では検出できなかった複合語型隠語の検出を目的とした手法について説明する．まず 7.1 節にて，複合語型隠語検出手法の中心アイデアについて提案する．続いて，7.2 節において，複合語型隠語検出手法の流れとアルゴリズムについて説明する．なお，アルゴリズムについては Python で実装した．最後に，7.3 節において，検出精度を向上させるために追加した機能について説明する．

7.1　複合語型隠語検出手法の中心アイデア

6.6.4 節でも考察したとおり，他の多くの隠語検出研究と同様に実験 1 の手法においても，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確認された．

そこで，まず複合語の検出が課題となるが，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」の間で区切られるが，単語分散表現モデルを構築すると，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状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，つまり，単語同士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同士の類似語として，互いの単語が上位に出現することが推測された．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」の類似語一位は「クラック」であり，一方の「クラック」の類似語一位は「グリーン」が検出となり，双方の類似語一位同士という結果が得られた．なお，この際の Word2Vec における設定パラメータで

Algorithm 3 ***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)
```

ある Windows size は 2 としている．このことから，共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．

7.2　複合語型隠語検出手法のアルゴリズム

まず複合語検出について，以下の方法を提案する (Algorithms 3, 4).

Bad コーパスで構築した単語分散表現モデル内の単語群単語群 $\mathcal{A}$ 内の単語 α の上位 M 位までの類似語を検索し（単語群 $\mathcal{S}(\alpha)$ とする），$\mathcal{S}(\alpha)$ 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する (*Get similar words*).

もし $S(\alpha)_i$ に α が含まれていれば,「$\alpha{\cdot}S(\alpha)_i$」と,「$S(\alpha)_i{\cdot}\alpha$」という複合語候補を作成する（*Concat*）.「$\alpha{\cdot}S(\alpha)_i$」と「$S(\alpha)_i{\cdot}\alpha$」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見なすこととした (*Count*).

ここで，複合語について前後の単語を総当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，本論文では，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．

具体的には，以下の流れで複合語の検出を行った．

Algorithm 4 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**
    if Sim_word = Word **then**
      Y ⇐ Y ∪ Concat{Sim_word, Word}
      Y ⇐ Y ∪ Concat{Word, Sim_word}
    end if
  end for
  return(*Y*)

1. Bad コーパスを分かち書きをし，Word2Vec[20] 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，単語群 $\mathcal{A}$ （$\mathcal{A}$= $\{\alpha_1,\ldots,\alpha_n\}$($\alpha_i$ は i 番目の単語）を作成する.(CreateList 関数)

2. α の上位 M 位までの類似語を検索する ($\mathcal{S}(\alpha)_1 \ldots \mathcal{S}(\alpha)_M$)．($\mathcal{S}(\alpha)_j$ は j 番目の単語)

3. 次に，$\mathcal{S}(\alpha)_i$ のそれぞれについて，類似語検索を行う．($\mathcal{S}(\mathcal{S}(\alpha)_i)_1 \ldots \mathcal{S}(\mathcal{S}(\alpha)_i)_M$)

4. もし $\alpha \subset \mathcal{S}(\mathcal{S}(\alpha)_i)_j$ となるものがあれば，“$\alpha \cdot \mathcal{S}(\alpha)_i$” と “$\mathcal{S}(\alpha)_i \cdot \alpha$” 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，全て対象とした.

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.

7.3 精度向上についての検討

7.2 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及び検証し，機能追加を行った.

7.3.1 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，二つのコーパス間 (Good コーパス，Bad コーパス) 間で同じ単語 α の類似語を検索し，同

じ単語が出現した単語 α は，隠語としての文脈で出現していないと考えた．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，複合語の処理を回避することとした．

7.3.2　形態素解析によるフィルタ

複合語型隠語は，Tweet を分析する中で，主に名詞同士から構成されているものが多いと判断したことから，複合語を結合させる前に，品詞分類を実施するようにし，名詞句だけを抽出することとした．

7.3.3　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当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，候補数を絞ることとした．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．

7.3.4　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」といった単単語や「警視庁」，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．

複合語の辞書的な意味に従うと，辞書に載っているものにも複合語がある．たとえば「援助交際」．しかし正しく分かち書きされるので，本論文の対象となる語ではない．本研究で対象とするのは,「分かち書きで分かれてしまうが，本来分けるべきではない語」であり，簡単な説明としては,「辞書に載っていないが，語を分割するとそれぞれは辞書に載っているもの」となる．ただし，辞書に載っていなくても，正しく分かち書きされる可能性はある．

第8章　実験2(複合語検出実験)

本章では，7 章で提案した複合語型隠語検出手法を実装し，そのうち，複合語が実際に検出されていたのか確認実験を行った結果を示す．まず 8.1 節で，実験概要を説明する．続いて 8.2 節では，実験のプロセスについて説明する．そして 8.3 節で評価を行った他の実験条件を説明し，8.4 では実験の結果を示す．最後に，8.5 節で評価結果について，考察する．なお，提案手法は Python で実装した．

8.1　実験の概要

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 9 章で説明する複合語型隠語検出実験を行なった．

8.2　実験のプロセス (複合語検出)

実験において図 8.1 の流れで処理を実施した．

8.2.1　アカウント単位でのデータ収集

5.2 節で示した「犯罪語リスト」を元に，収集したツイートについて，発信アカウント単位，すなわち犯罪の意図を含めたツイートを行うユーザー単位で再度ツイートを集め直した（図 8.2）．

これにより，作成したコーパスにおける,「犯罪語リスト」の単語の影響を軽減させるだけでなく，5.1 節で述べた「隠語の出現するツイートについての 4 分類」のうち，2 の「未知の隠語だけが利用されたツイート」を含むことも期待できる．また,「犯罪語リスト」の単語の中には,「野菜」,「氷」等のダブルミーニングの隠語は正しい意味で使われているツ

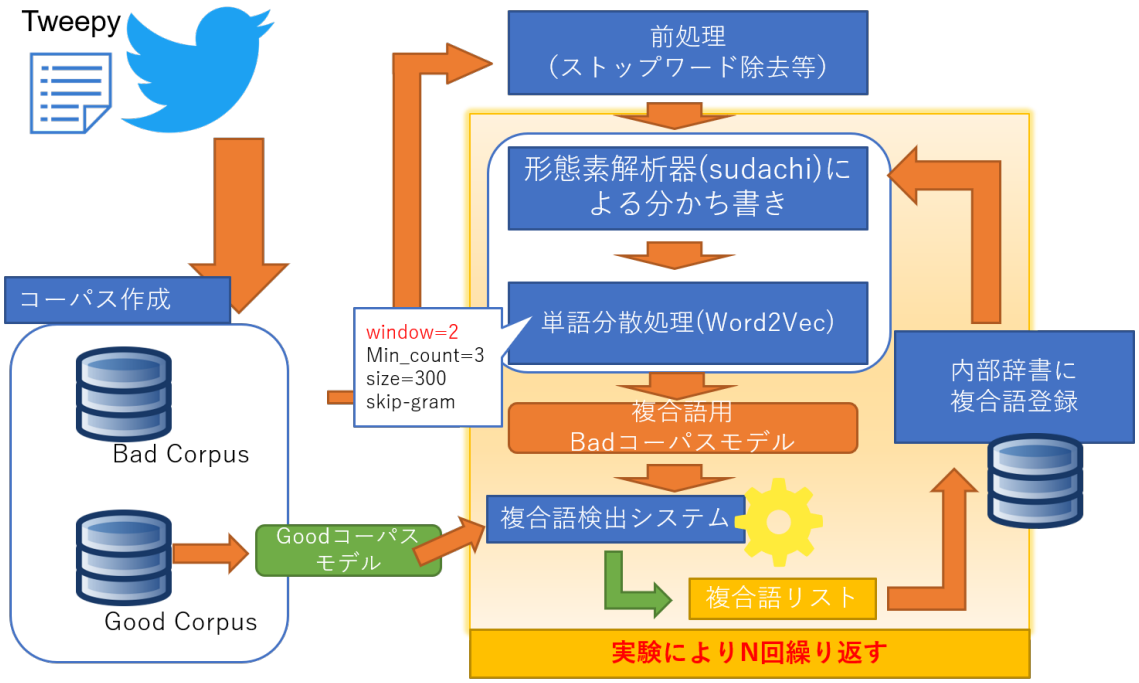


図 8.1: 実験 2 の実験プロセス

イート入らないようにするために含んでいないが，ユーザー単位でツイートを集めることでこのようなダブルミーニングの隠語の文章も含まれることが期待できると考えた．

具体的には 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，本文データについて，6.6.2 の考察を元に，薬物の取引関連という 1 ジャンルに絞り，いくつかのキーワードで「犯罪語リスト」を作成し，ツイートを抽出した後，その投稿アカウントを着目し，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する．その後，そのリストをキーに再度同じ期間における Tweet を収集し，Bad コーパスを作成する．

8.2.2　前処理

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

1. URL

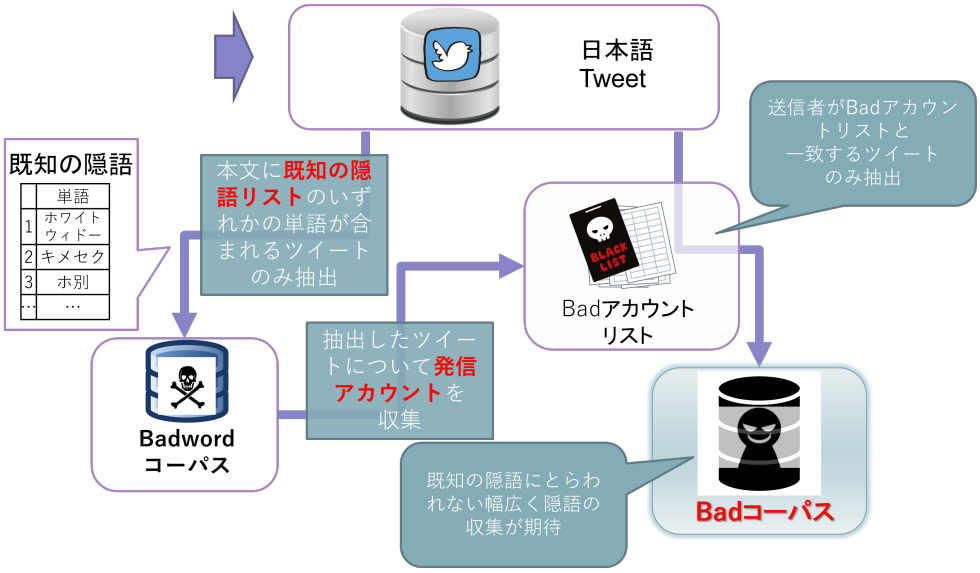


図 8.2: コーパスの作成方法

2. 改行文字

3. Twitter に定型でよく現れる単語
   「RT」「まとめ」「お気に入り」

8.2.3 コーパス作成

用意したコーパスについては，以下の二つである．

1. Bad コーパス
   8.2.1 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 [66]．なお，Good コーパスは本実験では，7.3.1 節で説明した提案手法 2 の機能追加において活用している．

8.2.4　形態素解析

日本語は特有の文章構造を保有しており，英語等と異なりスペース等で区切られないため，単語分散処理を行う前に，形態素解析処理及び分かち書きが必須である．分かち書きとは，事前に単語の辞書を内部で持ち，それに従い，文章を適切に単語単位に分割するものである．

また，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，正しく分かち書きがされる必要がある．さらには，複合語型隠語は一般的な単語の組み合わせも多く，複合語型隠語より短い文節で分かち書きされてしまう．そのため，分かち書き用の内部辞書に複合語を登録することで，複合語型隠語がより短い文節で分かち書きされないようにした．

なお，本研究では形態素解析器として Sudachi[68] を用いている．

8.2.5　単語分散表現モデルの構築

Word2Vec を用いて実行する．複合語検出用の Word2Vec のパラメータは表 8.1 のとおり．

表 8.1: 複合語検出用の Word2Vec のパラメータ

パラメータ項目	設定値
Size	300
Min-Count	3
Window Size	2
Negative	20
手法	Skip-Gram[69]

8.2.6　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，7.2 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，本実験ではアルゴリズムの上位 M 位について，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，$\mathcal{X} = 2$ 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 連複合語だけでなく，3 連複合語などへの対応を想定し，8.2.4 から 8.2.6 までの処理を 10 回繰り返した．

8.3 実験条件

提案手法に加え，以下の三つの手法で比較評価を実施した．

8.3.1 ベースライン手法 A(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，全てのバイグラムを取得したものをベースライン手法 A 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 語が作成された．そして，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 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った．そのため，ベースラインの精度がもっとも高くなるように調整を行った結果, 出現頻度の閾値を 2 回から 14 回以上と緩和し，対象単語数を 14,218 語と選定した．

8.3.2　ベースライン手法 B(N-gram 生成)

Small らによるフレーズ生成方法を参考に，ベースライン手法 B として用意した [61].
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 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，主フレーズとみなす.

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．なお，複合語としてい成立しているかの評価については，著者１名で実施した.
その結果，707 語が候補となり，131 が成立しており，18.53%であった.

8.3.3　機能追加無し条件

提案手法から，7.3 節に記載のとおり，4 つの機能の搭載を除外し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.

8.4　実験結果（複合語検出実験）

三つの手法で複合語候補を検出し，複合語であるか否かの二つに分類した．また複合語として判定された単語については，さらに隠語もしくは犯罪関連語か分類した.
複合語検出の結果は表 8.2 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 (7.2 節 (3)) については，提案手法は $\mathcal{X}$=2，ベースラインは 8.3.1 でも説明したとおり，$\mathcal{X}$=14 とした.
表 8.2 より，提案手法はベースライン手法 A,B 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.

表 8.2: 複合語候補の分類

手法	複合語候補		複合語	Precison[b]
	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
提案手法（機能追加無し）[a]	101,573	521	308	0.591
ベースライン手法 A	686,561	14,218	388	0.027
ベースライン手法 B	-	707	64	0.091

[a] 8.2 節のプロセスの 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）

8.5 考察 (複合語検出実験)

8.5.1 提案手法の有効性について

表 8.2 より，複合語の検出については，ベースライン手法 A,B 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言える．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言える．

また，単語の関連性に着目した本手法が複合語検出について有効な手法なのかについて，固有名詞という観点から調査した．

表 8.3: 固有名詞の出現割合

手法	複合語	固有名詞	割合
提案手法	264	128	48.5%
提案手法（機能追加無し）	308	112	36.4%

これより，類似語に基づいたフィルタ有の手法により，固有名詞が 48.5%と半数近く検出していた (表 8.3)．氏名などの固有名詞は連続して出現する頻度が高いため，類似語に着目した手法により，固有名詞が高い頻度で出現していることは複合語を類似語に基づき検出する手法として，効果がある手法であったと言える．また複合語型隠語が提案手法の

フィルタ機能を追加ものが一番高い検出率であったことからも，複合語型隠語を含めた複合語の検出する手法として，提案手法は有効であると言える．

なお，Word2Vec のパラメータのうち，現在の単語と予測される単語の最大距離を設定する「Window Size」については 1 から 4 まで事前検証をしたところ，複合語検出として Window Size=2 が一番良い結果が得られたため，今回のパラメータとしている．

8.5.2　試行回数について

今回，提案手法の 8.2 節について 10 回試行した結果について考察する．図 8.4 より，どちらの手法であっても，回数を重ねるに従い，出現回数が減っていくのが分かる．回数を重ねることで，関係性のある単語が複合語として登録されることで，関係性の強い単語が減少したものと考えられる．

表 8.4: 10 回の試行における登録した複合語の数の比較

回数	提案手法	提案手法 w/o 追加機能
1 回目	159	277
2 回目	40	76
3 回目	18	37
4 回目	17	32
5 回目	12	23
6 回目	11	21
7 回目	11	12
8 回目	9	16
9 回目	6	9
10 回目	12	18
合計	264	521

また，作成された単語のうち，成立していた単語の中には，大麻の隠語として用いられる「レモンスカンク」,「ホワイトウィドー」,「OG クッシュ」,「ゴリラグルー」,「ブルードリーム」,「グリーンクラック」といった複合語も作成されていた．一方で，固有名詞も多く出現しており，一般的な固有名詞としては，ゲーム名である「フォートナイト」やゲームやアニメのキャラクター名である「左馬刻」,「理鶯」,「寂雷」といった単語も検出され

ていた．さらには出現した単語について確認することで，隠語と気づいていなかった単語（「ハッシュプラント」,「チキータバナナ」）を発見することができた．

第9章　実験3(複合語型隠語検出実験)

本章では，5 章で提案した提案手法 1 だけでは検出できなかった複合語型隠語を検出できるか実験を行った結果を示す.

まず 8.1 節で，実験概要を説明する．続いて 8.2 節では，実験のプロセスについて説明する．そして 8.3 節で評価を行った他の実験条件を説明し，8.4 節では実験の結果を示す．最後に，8.5 節で評価結果について，考察する．なお，提案手法は Python で実装した.

9.1　実験の概要

用意した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，コーパス内にある単語群 21,210 語に含まれるうち，照合用の単語リストとして 18 語を用意し隠語の検出を目指す.

9.2　実験のプロセス (複合語型隠語検出実験)

実験において図 9.1 の流れで処理を実施する.

9.2.1　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) のうち収集したうち，本文データのみを使用する．続いて，薬物の取引関連のいくつかのキーワードを元にツイートを抽出した後，その投稿アカウントを着目し，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する．その後，そのリストをキーに再度同じ期間における Tweet を収集し，Bad コーパスを作成する.

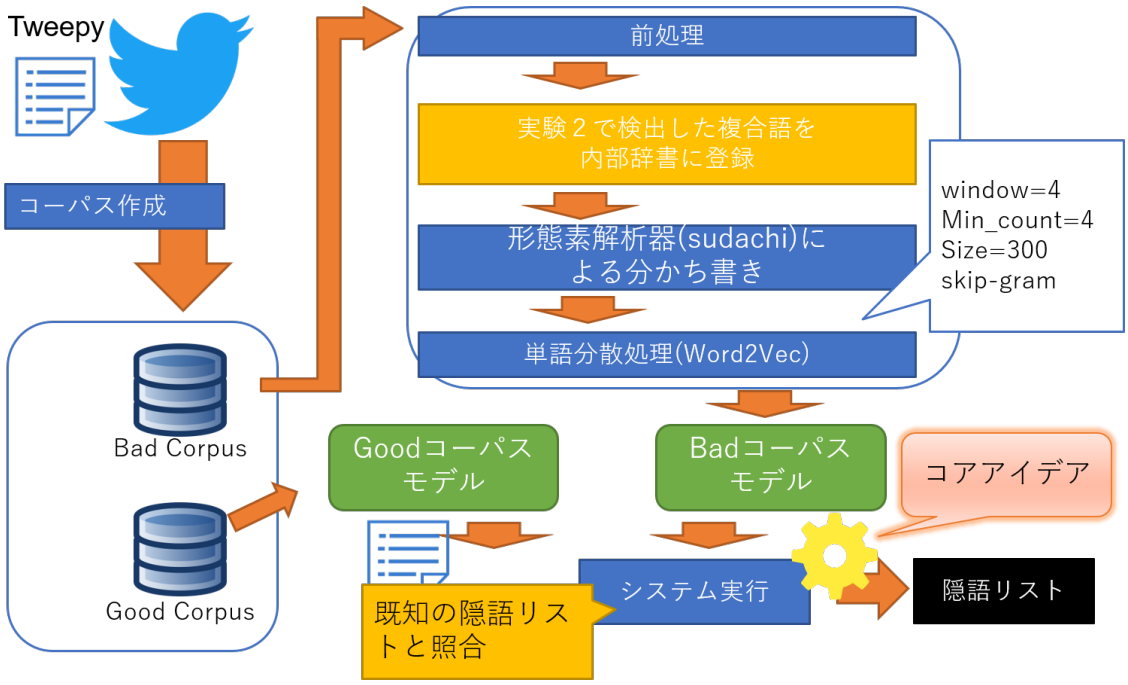


図 9.1: 実験 3 の実験プロセス

9.2.2　前処理

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

1. URL
2. 改行文字
3. Twitter に定型でよく現れる単語
   「RT」「まとめ」「お気に入り」

9.2.3　コーパス作成

用意したコーパスについては，以下の二つである．

1. Bad コーパス
   8.2.1 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 コーパスを利用する [66].

9.2.4 複合語の検出と辞書登録

提案手法のみ，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，8.2 節のプロセスを 10 回繰り返し，10 回の試行で検出された全ての検出語の内部辞書への追加を行う．本実験では，実験 2 の 8.4 節で検出された 264 語を登録する．

9.2.5 形態素解析

分かち書き用の内部辞書に複合語を追加し，分かち書きを行う．これにより，実験 1 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される．なお，本研究ではこれまで同様，形態素解析器として Sudachi[68] を用いる．

9.2.6 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 9.1 のとおり設定した．

表 9.1: Word2Vec のパラメータ

パラメータ項目	設定値
Size	300
Min-Count	3
Window Size	4
手法	Skip-Gram[69]

9.2.7 提案システムの実行

システムへのインプットとして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及び Bad コーパスのみに出現する単語を抽出して単語リストを作成する．なお，両

表 9.2: 精度比較

評価手法	提案手法 2		提案手法 1	
	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	

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 語であった．

また，今回の実験では検索する類似語上位 N の値は 20 とした．

9.3　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 9.2 のとおりであった．表 9.2 の提案手法 1 の結果については，複合語辞書を使用しない以外は，全て実験 3 と同じ実験条件 (コーパス, 既知の隠語リスト) で実験を行った．

5 章の提案手法 1 ではこれまで検出できなかった複合語を，今回の提案手法 2 と組み合わせることにより複合語型隠語も確認できた．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．検出された単語の詳細については，付録. D 節のとおり．

また Precision についても提案手法 1 に比べ， 7.1%ポイント精度の良い結果が得られた．またに提案手法 1 に比べ，精度だけでなく検出した隠語の数も 20 個多く検出できた．

9.4　実験結果 (闇バイト)

違法薬物とは別に，闇バイト関連の単語について，闇バイト関連の単語を用いて 2019 年 7 月から 2020 年 7 月の期間の Tweet を元に Bad コーパスを作成し，提案プログラムを実行した．実験の結果は表 9.3 のとおりであった．

表 9.3: 分類結果

分類	個数	割合
隠語	5	11.4%
犯罪関連語	27	61.4%
無関係	12	27.3%

提案手法により，隠語及び犯罪関連語を合わせて 72.8%検出できた．中には，隠語として，受け子出し子を意味する「ud」や強盗を意味する「叩き」，違法物品を運搬する意味の「運び」などを検出された．闇バイトのジャンルにおいても提案手法を用いることで隠語や犯罪関連語を検出できることが示されたと言える．

一方で，出現した単語については，隠語よりも犯罪関連語が多く検出された．これについては，作成したコーパス内のツイートを確認したところ，闇バイトに関する隠語が少なかったためであると考えられる．この背景として違法薬物売買や児童買春については，以前から悪質性は一般的に認知され，取り締まりの対象であったことから，取引に応じる側も違法なやり取りを自覚しているものが多いため，隠語でのやり取りがされていた．一方で，闇バイトについては，2023 年に発生した強盗殺人事件でフォーカスされるまでは，違法性を自覚していないものも誘因することを意図している場合もあり,「短時間高収入」,「即日即金」などと言った一般的に理解できる言葉を使用した，隠語を必要としない取引が違法薬物売買と比べて多かったためと考えられる．このような隠語が使われる機会が少ないと考えられるジャンルについては，犯罪関連語も他のジャンルよりも，より重要な情報となり得ることも考えられることから，犯罪関連語と判定された単語も含めて幅広く確認することでより効率的な運用につなげることが期待できる.「闇バイト」ジャンルについて，今後は省庁横断的な犯罪防止対策が強化されることから，取引を隠語でやり取りすることが増加することと予想されるため，より対策を検討する必要があると思われる．

第10章　考察

本章では，本研究の実験を行った結果の考察について述べる．

10.1　隠語検出実験について

10.1.1　提案手法のハイパーパラメータについて

提案手法では，閾値をどう設定するかによって精度（Precision 及び Recall）は変化させることが可能である．Precision と Recall はトレードオフの関係にあるため，状況に応じて，Recall を重視したり，Precision を重視したり，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させるべきものであり，その点を踏まえて設計している．具体的には，たとえば，事前に用意した隠語のうち，5 個隠し，その 5 個のうち，何個検出できるか見つけられるかを事前に実施し，目標となる Recall を 60%としたとき，5 個中 3 個見つけた時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．本論文の実験においては，隠語及び犯罪関連語検出機能の検出における閾値を設定できるようにしたところ，実験 1 開始前にそれぞれの項目を調整し，一番良い結果の値を 6.2 のとおり，設定値として決定し，その後の実験 2,3 についても同じパラメータ値を用いた．

10.1.2　精度と誤検出した単語について

検出精度については，一定以上の精度は重要であると考えるところ，隠語検出に関しての実運用を想定し，警察庁におけるサイバーパトロールに関する担当者へ今回の実験結果についてヒアリングを実施した結果，実験 1 の実験結果である Precision が約 0.6 という精度は実用に耐えうる精度であるとのコメントを得た．またヒアリングの中で，隠語に限ら

ず，隠語と関連した犯罪関連語の検出も効果的であるとの意見を得られた．犯罪関連語は隠語そのものではないものの，隠語を用いた取引に利用されることが多く，犯罪捜査においては犯罪関連語を把握することも重要である．そこで，本研究において隠語であると検出した 34 語について確認したところ，6.5 節のとおり，19 語は隠語であり，誤検出した 15 語について確認したところ，そのうち犯罪関連語が 11 語であった．犯罪関連語も検出対象として正解に含めた結果は，Precision が 0.882 と非常に高く，検出した単語の中には，隠語もしくは犯罪関連語を多く検出していることがわかる．ただし，犯罪関連語については品詞分類を用いて自動的に追加登録するようにしているところ，隠語として検出されているものもあることからより隠語，犯罪関連語の検出精度を向上させる方法について検討する必要がある．

10.2　複合語型隠語検出実験について

10.2.1　検出された複合語型隠語について

隠語検出手法だけでは検出できていなかった複合語隠語について，複合語検出手法により複合語辞書を作成することで，隠語検出手法のみでは検出できなかった隠語を検出することができるようになった．具体的には，ジャックヘラー，レモンスカンク，グリーンクラック，サワーディーゼル，ホワイトウィドー，ゴリラグルー，ホワイトウィドウ，パイナップルチャンク，ブルードリームなどが検出できており，どれも既存手法では検出できない隠語であった．

分析を通じて，単語自体が隠語であるもの，野菜，アイスなどの一般的な意味が幅広く使われている単語については本来別の意味を指すにも関わらず，隠語としての意味が連想できるものやそれ以外に，単単語の中には，隠語としての使われ方がされているものの周囲の文脈から周囲の単語と同様に隠語と判断できるものもあることが判明した．単独ではいろいろな意味で解釈できる幅が広いため，単独で用いられても意味が通じにくいもの「ホワイトあります」,「パープルあります」なども隠語だけでなく，周囲に，本来のマンゴーとは異なる文脈の単語（たとえば，野菜やアイス，手押しなど）があることによって，この単語が隠語であり，ホワイトウィドーであったり，パープルヘイズを意味すると示唆で

きるのである．例えば，ホワイト，マンゴーやブルーベリー，蜂蜜などがそれに当たる．

10.2.2 複合語としては検出されなかった単単語について

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9.2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．たとえば,「ビック」(「ビックバッツ」,「ビックバット」等を確認）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」等を確認）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」等を確認) など，複合語型隠語の一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．

10.2.3 3連続以上の複合語について

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，さらに単語同士が結合した，いわゆる3連複合語が検出されていたこともわかった．10回の試行の結果は表10.1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．

10.2.4 未知の隠語の検出について

本提案手法により，実際に未知の隠語が検出されていたのか，現場の警察官にヒアリングを実施し，確認を行った．実験3で検出した73語について，警察庁刑事局組織犯罪対策

表 10.1: 3 連複合語の出現数の結果

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%

第一課所属の現役警察官 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．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 10.2 のとおり．

表 10.2: 未知の隠語かどうかのヒアリングの結果

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	

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言える．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わかった．

また私自身も，本文とインターネットで調べることで隠語と理解できた単語は数多くあった．今回の単語にあっては，当初の知識では 73 語中 6 語しか認知していなかった．本実験においても検出した隠語については，元のツイートを確認していった結果，隠語と認知していなかったものが確認された．具体的には,「ソマンゴ」,「モーリー」,「AK」,「チーズ」,「パイナップルチャンク」などといった単語であった．

10.3 提案手法の限界について

提案手法は，単語分散表現モデルを構築し，類似語に対し，照会リストと照合させ，隠語を検出する手法である．提案手法で隠語を検出できないリスクについて検討する．

隠語を検出できない状況は大きく二つあると考える．

- 照合リストと HIT しない
- 単語分散表現モデル内の単語に出現しない

それぞれの状況について考察する．

照合リストと HIT しない

提案手法は類似語と照合リストとの突合させ，隠語を検出することから，照合リストの単語の選択次第では類似語と HIT せず，隠語として検出されないおそれがある．ただし，この点については，照合リストとして既知の隠語を充実させることで対応可能と考える．

単語分散表現モデル内の単語に出現しない

一方で，提案手法は，単語分散表現モデル内の単語に対し類似語に基づき隠語を検出することから，照合リストを充実させたとしても，隠語の出現頻度が低いことで単語分散表

現モデル構築時の最低出現回数の閾値以下の単語については，そもそもの隠語候補としてシステムにインプットする単語として存在しなければ隠語かどうかの判定ができない．

これが発生する状況としては，認知度があまりに低く隠語として使用される頻度が低いことが考えらえ，あまりにも流通していない違法薬物等の名前や隠語として使用され始めてから日が浅いものなどが検出できないおそれがある．対応方法としては，単語分散表現モデル構築時の単語の出現頻度の閾値を下げることであるが，一方で下げすぎるとノイズが増えシステム全体の精度も低下するおそれもあることから，認知度の低い隠語の把握の必要性・緊急性とのバランスで調整すべきと考える．

第11章　おわりに

本章では，11.1 節で本研究の課題と提案手法について，11.2 節で評価結果についてまとめる．そして最後に，11.3 節では，実用化に向けて隠語検出における今後の課題と検出した隠語の活用方法案について述べる．

11.1　本研究の課題と提案手法

本論文では，サイバーパトロールにおいて犯罪の端緒を迅速に把握することを支援するため，犯罪の用途に用いられる隠語や犯罪関連語を検出することを目的として，コーパス間の単語の類似語の差異に着目する手法を提案した．

一般に隠語は，認知度が低い単語や「野菜」,「氷」などの一般的な単語にカモフラージュさせ使用されることから，大規模言語モデルを作成したとしても隠語を検出することが難しいという課題があった．そこで，犯罪に関係する単語の類似語もまた，犯罪を意図するものであると考え，同じ単語であっても一般的なコーパスと不正な目的のコーパスでは類似語が異なる点に着目し，用意した二つのコーパス間の同じ単語の類似語を比較し，隠語を検出する新たな手法を提案した．

また，多くの隠語検出研究において，分かち書きされた単語に基づき隠語を検出するため，二つ以上の単語を結合させた，すなわち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ところ，複合語は分割された単語同士の出現頻度が共に高いため，単語の関連性が高く現れることが判明したことから，単語の関連性を利用し，まずは複合語を検出し，形態素解析器の内部辞書に検出した複合語を登録することで，前述した隠語検出手法を実施することで，複合語型隠語を検出する手法を提案した．

11.2　評価の内容と結果

提案手法を用いて隠語検出実験を実施した結果，照合用に用いた単語以外の隠語，すなわち未知の隠語を検出することができた．また比較用に用意したベースライン手法と比べても，高い精度を得ることができた．

続いて，複合語型隠語を検出するため，複合語の検出を実験により確認すると共に，その後の複合語型隠語の検出実験により，隠語と複合語型隠語の検出を目指した．実験の結果，複合語型隠語の検出を確認でき，Precision もベースライン手法だけでなく，実験 1 に比べ精度が向上し，また実験 1 では検出できていなかった複合語型隠語も 10 語検出することができた．

これらのことから，本提案手法を拡張することで，変遷していく隠語を自動的に検出でき，サイバーパトロール等への貢献が期待できる．

11.3　実用化に向けて

本節では，提案手法について今後の解決するべき課題を述べ，実用化に向けて検討する．

11.3.1　ヒアリング調査の実施

2019 年 11 月に，A 県警察による Twitter におけるパパ活防止の取り組み実施についての報道を受けて，実際の愛知県警察に趣き，担当の警察官にサイバーパトロールの実施方法について，ヒアリングを実施し，現場の声とニーズの把握を目指した．

本取り組みは，Twitter 上でパパ活に関連する投稿を発見し，それらの投稿に対し，A 県警の定型文をリプライし，犯罪であること，サイバーパトロールで監視していることを伝えることで犯罪を踏みとどまらせ，被害を巻き込まれたり加担することを防止させることを目的としたものであった．

本取り組みの運用方法としてはパパ活に関連した違法な投稿を検出する方法として，基本的にはパパ活に関係しそうなキーワードによる検索やアカウントの調査などを手動で実施しているとのことであった．

またキーワードの選定方法や複数のキーワードを指定して検索を行なう「アンド検索」などの工夫についても，担当者が自分で考えて実施しているとのことで，取り締まりの精度が担当者の知識に依存しているとのことであり，つまりは属人的な対応になっているといえる．一方で，キーワード検索を手動で行い，不正な投稿を探そうとしても，不正な取引のための投稿を行なう者たちは，サイバーパトロールによる検索逃れたり，自身の投稿を削除されないようにするため，不正なやり取りを直接的な表す単語は避け，隠語を用いてやり取りする．

隠語は，その単語を認識しているもの同士しか分からないことから，手動でキーワード検索で監視しようとするものにとっては，まずは隠語自体を把握していないと，それらの単語を用いてやり取りする者たちを見落としてしまうことになる．また属人的に対応している場合，部署異動等により，初めて業務に携わるものにとっては，そもそも検索するための単語を把握しておらず，隠語も時と共に変遷することから，継続的に把握していないと最新の隠語が把握できなくなってしまう．その場合，不正な投稿を見つけるノウハウを蓄積する間にも多くの不正な取引を見逃すおそれがある．

このようなことから，サイバーパトロールを実施し犯罪を早期に検知するためにも，まずは隠語，さらには時と共に変遷する隠語のうち，最新の隠語を把握することが重要であると思われる．なお，担当者にヒアリングした際も隠語だけでも認知できることは非常に有効であるとの意見があった．

このように，ヒアリングで現場の声からも隠語の検出のニーズが確認できた．

隠語の検出のニーズを改めて確認できた中で，実用化に向けて隠語検出における今後の課題と検出した隠語の活用方法案について述べる．

11.3.2 隠語検出において課題となる点

提案手法について，複合語型隠語の検出精度の向上が課題として挙げられる．具体的には，今回の実験を通じて照合リスト外の，いわゆる未知の隠語も含めた隠語，さらには複合語型隠語の検出が確認できたが，10.2.2 節でも考察したとおり，今回の実験では文節が切れて複合語型隠語の一部のみ（「ビック」（「ビックバッツ」,「ビックバット」等を確認）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」等を確認）等）検

出されたものが確認された．これらの一般的な単語と結びつきが強いと思われる複合語型隠語の検出は今後の課題と考える．

11.3.3　コーパスや隠語リストの運用について

実運用を想定した際，日々の運用においてのコーパスの更新や隠語リストの更新についてであるが，隠語は変遷していくものの，実際のツイートを複数年確認したところ，短期間では変わらないこと，そして仮に変化したとしても，周辺には関連する単語は出ることもわかった．一方で本提案手法は，ダイレクトに隠語と一致させ新たな隠語を検出するのではなく，類似語から新たな隠語を検出する手法である．そのため，一度隠語リストを作成した場合，仮に隠語が陳腐化したとしても，すぐに使えなくなるものではなく，長期間有効に使用できると考える．また過去の隠語リストに加え，1 年ごとなど一定のスパンでコーパスを作成しなおすことで新たな隠語を検出することが期待できる．

11.3.4　検出した隠語の活用方法について

本手法により隠語を認識したうえで次の課題となるのは，隠語が用いられている不正なツイートを検出することである.「野菜」が隠語と認識した上で,「野菜」をキーワードにツイートを検索したとしても，大量及び不正な取引とは関係のない一般的な「野菜」に関するツイートが数多く検索されることは，4.1 節のとおりである．隠語の出現するツイートを確認する中で，共起する隠語や犯罪関連語がジャンルによって異なることがわかった．大麻などの薬物取引関連の単語においては，たとえば「野菜」と一緒に「手押し」や「高純度」などが共起することが多いが，援助交際関連の単語として，たとえば「神待ち」では「手押し」や「高純度」といった単語とは共起せず,「諭吉」,「苺」,「パパ」,「JK」などといった単語が共起する頻度が高いことがわかった．さらに同じジャンルにおいても，隠語によって共起する単語が異なることもあった．一方で,「野菜」の場合,「生活」（飲料水の製品名と思料）や「トマト」などと共起する場合，その野菜が出現するツイートは，一般的な用途でのみ使われている可能性が高い傾向がみられた．またそのため，ジャンル（薬物取引や援助交際等）や単語単位で共起する単語を細かく調整可能とすることで，隠語・

犯罪関連語との共起による不正なツイートの検出，一方で一般的な単語との共起による無関係なツイートの除外をより効果的に実現させることが期待できる．

謝辞

お世話になった方々にこの場をお借りしてお礼申し上げます.

本研究にあたり，ご多忙中のところ，丁寧にご指導くださった大須賀 昭彦 教授，清 雄一　教授，田原 康之 准教授に感謝いたします．また，共にゼミに参加し，様々な知見をくださった大須賀・田原・清研究室の皆様に感謝の意を表します．ゼミの開催時間についても，私の業務の都合を考慮いただき，毎週夜に開催いただきまして，誠にありがとうございました.

また私の審査をお忙しい中，快く引き受けてくださいました大学院 情報理工学研究科の柏原 昭博 教授，内海 彰 教授，稲葉 通将 准教授に感謝申し上げます．先生方のご指摘のおかげで，新たな観点を提供いただき，より研究と博士論文をより良いものとすることができました.

特に清 雄一 教授につきましては，日々の研究だけでなく，投稿する論文や国際発表，その他様々な場面で頼りにさせていただきました．厚く御礼申し上げます.

また，私は警察庁刑事局組織犯罪対策部組織犯罪対策第一課に在籍し，日々の業務の傍ら，社会人博士過程へ就学していたところ，業務の調整やお心遣い・ご協力くださった，前田 和彦 警視，同僚の高見 修平 警部を始め，上司・同僚の方々に心から感謝申し上げます.

最後に，これまで博士前期課程から含めると 5 年間に渡り，自身も働きながら家事や育児等で多忙にも関わらず，私を支え研究についても相談に乗ってくれた妻 英理美，いつも元気を与えてくれ，そしてパパに代わって進んで綸汰朗のお世話をしてくれた，長女 陽菜理，次女 詩緒理，そして元気にスクスクと育ち，いつもニコニコして心を和ませてくれる次男 綸汰朗に心から感謝します．そして，長男 律志朗，いつもお空から見守ってくれてありがとう.

参考文献

[1] インターネット・ホットラインセンター. インターネット・ホットラインセンター　年間統計（2022年）. https://www.internethotline.jp/pdf/statistics/2022.pdf, 2023.

[2] インターネット・ホットラインセンター. インターネット・ホットラインセンター統計情報. https://www.internethotline.jp/pages/statistics/index, 2023.

[3] 法務省. 令和３年版 犯罪白書. https://hakusyo1.moj.go.jp/jp/68/nfm/n68_2_4_2_2_3.html#h4-2-2-3, 2021.

[4] 第五次薬物乱用防止五か年戦略フォローアップ 令和５年８月８日取りまとめ. https://www.mhlw.go.jp/content/11120000/000956682.pdf.

[5] 警察庁.　令和４年における少年非行及び子供の性被害の状況（更新版）（令和５年９月 15 日）.　https://www.npa.go.jp/bureau/safetylife/syonen/pdf-r4-syonenhikoujyokyo-kakutei.pdf, 2023.

[6] 令和元年における少年非行, 児童虐待及び子供の性被害の状況. https://www.npa.go.jp/safetylife/syonen/hikou_gyakutai_sakusyu/R1.pdf.

[7] SNS等に起因する被害児童の現状と対策. https://www8.cao.go.jp/youth/kankyou/internet_torikumi/kentokai/40/pdf/s4.pdf.

[8] 情報通信白書令和 3 年版.　https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2021/2021-index.html.

[9] 平成30年におけるSNSに起因する被害児童の現状. https://www8.cao.go.jp/youth/kankyou/internet_torikumi/kentokai/41/pdf/s4-b.pdf.

[10] 警察庁. 特殊詐欺認知・検挙状況等 (令和４年・確定値) について. https://www.npa.go.jp/bureau/criminal/souni/tokusyusagi/tokushusagi_toukei2022.xlsx, 2023.

[11] Asia-Pacific drug trade thrives amid the COVID-19 pandemic. https://www.reuters.com/article/us-asia-crime-drugs/asia-pacific-drug-trade-thrives-amid-the-covid-19-pandemic-idUSKBN22R0E0.

[12] Rolf van Wegberg, Fieke Miedema, Ugur Akyazi, Arman Noroozian, Bram Klievink, and Michel van Eeten. Go see a specialist? predicting cybercrime sales on online anonymous markets from vendor and product characteristics. In *Proceedings of The Web Conference 2020*, WWW '20, p. 816?826, New York, NY, USA, 2020. Association for Computing Machinery.

[13] Hao Yang, Xiulin Ma, Kun Du, Zhou Li, Haixin Duan, Xiaodong Su, Guang Liu, Zhifeng Geng, and Jianping Wu. How to learn klingon without a dictionary: Detection and measurement of black keywords used by the underground economy. pp. 751–769, 05 2017.

[14] 京都新聞社. ツイッターで覚醒剤販売呼びかけ，容疑の男再逮捕（2022 年 9 月 14 日 ）. https://www.kyoto-np.co.jp/articles/-/880074, 2022.

[15] 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 ）. https://www.kantei.go.jp/jp/singi/hanzai/kettei/230317/honbun-1.pdf, 2023.

[16] Rada Mihalcea and Vivi Nastase. Word epoch disambiguation: Finding how words change over time. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 259–263, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[17] Derry Wijaya and Reyyan Yeniterzi. Understanding semantic change of words over centuries. 10 2011.

[18] Kan Yuan, Haoran Lu, Xiaojing Liao, and XiaoFeng Wang. Reading thieves' cant: Automatically identifying and understanding dark jargons from cybercrime marketplaces. In *27th USENIX Security Symposium (USENIX Security 18)*, pp. 1027–1041, Baltimore, MD, August 2018. USENIX Association.

[19] 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, August 2013.

[20] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013.

[21] 第五次薬物乱用防止五か年戦略. https://www.mhlw.go.jp/content/11120000/000339984.pdf.

[22] 第六次薬物乱用防止五か年戦略. https://www.mhlw.go.jp/content/11120000/000339984.pdf.

[23] In the know zone. http://www.intheknowzone.com/substance-abuse-topics/marijuana/street-names.html.

[24] 警察庁 HP. 薬物乱用のない社会を. https://www.npa.go.jp/bureau/sosikihanzai/yakubutujyuki/yakubutu/nodrug.pdf.

[25] 愛知県警察 HP. 乱用されている薬物. https://www.pref.aichi.jp/police/anzen/soshiki/yakuju/ranyou.html.

[26] Laura Miller. Those naughty teenage girls: Japanese kogals, slang, and media assessments. 2004.

[27] Metropolitan Police Department. Metropolitan Police Department Website. https://www.keishicho.metro.tokyo.lg.jp/kurashi/tokushu/furikome/furikome.html, 2023.

[28] 時事通信社. 「闇バイト」投稿，警告１．５倍　甘言添えた勧誘急増ー困窮の若者標的か・警視庁（2023 年 01 月 28 日）. https://www.jiji.com/jc/article?k=2023012800151&g=soc, 2023.

[29] Jawaid Mangnejo, Arif Khuhawar, Muneer Kartio, and Saima Soomro. Inherent flaws in login systems of facebook and twitter with mobile numbers. *Annals of Emerging Technologies in Computing*, Vol. 2, pp. 53–61, 10 2018.

[30] Wonhee Lee, Samuel Sangkon Lee, Seungjong Chung, and Dongun An. Harmful contents classification using the harmful word filtering and svm. In Yong Shi, Geert Dick van Albada, Jack Dongarra, and Peter M. A. Sloot, editors, *Computational Science – ICCS 2007*, pp. 18–25, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[31] 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, March 2013.

[32] Kevin Dela Rosa and Jeffrey Ellen. Text classification methodologies applied to micro-text in military chat. pp. 710–714, 12 2009.

[33] Furkan Sahinuç and Cagri Toraman. Tweet length matters: A comparative analysis on topic detection in microblogs. In Djoerd Hiemstra, Marie-Francine Moens, Josiane Mothe, Raffaele Perego, Martin Potthast, and Fabrizio Sebastiani, editors, *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part II*, Vol. 12657 of *Lecture Notes in Computer Science*, pp. 471–478. Springer, 2021.

[34] Kaiming Yao, Haiyan Wang, Yuliang Li, Joel J. P. C. Rodrigues, and Victor Hugo C. de Albuquerque. A group discovery method based on collaborative filtering and knowledge graph for iot scenarios. *IEEE Transactions on Computational Social Systems*, pp. 1–12, 2021.

[35] Lu Huang, Fangyan Liu, and Yi Zhang. Overlapping community discovery for identifying key research themes. *IEEE Transactions on Engineering Management*, Vol. 68, No. 5, pp. 1321–1333, 2021.

[36] Tracie Farrell, Oscar Araque, Miriam Fernandez, and Harith Alani. On the use of jargon and word embeddings to explore subculture within the reddit ' s manosphere. In *Proceedings of the 12th ACM Conference on Web Science*, WebSci '20, p. 221?230, New York, NY, USA, 2020. Association for Computing Machinery.

[37] 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, may 2010.

[38] 美穂子北村, 裕治松本. 対訳コーパスを利用した対訳表現の自動抽出. 情報処理学会論文誌, Vol. 38, No. 4, pp. 727–736, apr 1997.

[39] Dominic Seyler, Wei Liu, XiaoFeng Wang, and ChengXiang Zhai. Towards dark jargon interpretation in underground forums. In Djoerd Hiemstra, Marie-Francine Moens, Josiane Mothe, Raffaele Perego, Martin Potthast, and Fabrizio Sebastiani, editors, *Advances in Information Retrieval*, pp. 393–400, Cham, 2021. Springer International Publishing.

[40] Liang Ke, Xinyu Chen, and Haizhou Wang. An unsupervised detection framework for chinese jargons in the darknet. pp. 458–466, 02 2022.

[41] Hailin Wang, Yiwei Hou, and Haizhou Wang. A novel framework of identifying chinese jargons for telegram underground markets. pp. 1–9, 07 2021.

[42] K. Zhao, Y. Zhang, C. Xing, W. Li, and H. Chen. Chinese underground market jargon analysis based on unsupervised learning. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pp. 97–102, 2016.

[43] 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.

[44] 智史安彦, 諒加藤, 悦司北川. 機械学習を用いた薬物売買におけるサイバーパトロールシステムの開発. 情報処理学会論文誌, Vol. 61, No. 3, pp. 535–543, mar 2020.

[45] Daniel O'Day and Ricardo Calix. Text message corpus: Applying natural language processing to mobile device forensics. pp. 1–6, 07 2013.

[46] C. Kansara, R. Gupta, S. D. Joshi, and S. Patil. Crime mitigation at twitter using big data analytics and risk modelling. In *2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pp. 1–5, Dec 2016.

[47] Guang Xiang, Bin Fan, Ling Wang, Jason Hong, and Carolyn Rose. Detecting offensive tweets via topical feature discovery over a large scale twitter corpus. pp. 1980–1984, 10 2012.

[48] Gregor Wiedemann, Eugen Ruppert, Raghav Jindal, and Chris Biemann. Transfer learning from LDA to bilstm-cnn for offensive language detection in twitter. *CoRR*, Vol. abs/1811.02906, , 2018.

[49] Ali Hakimi Parizi, Milton King, and Paul Cook. UNBNLP at SemEval-2019 task 5 and 6: Using language models to detect hate speech and offensive language. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pp. 514–518, Minneapolis, Minnesota, USA, June 2019. Association for Computational Linguistics.

[50] 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, 3 2021.

[51] 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, mar 2018.

[52] 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.

[53] Wanzheng Zhu and Suma Bhat. Euphemistic phrase detection by masked language model, 2021.

[54] Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R Voss, and Jiawei Han. Automated phrase mining from massive text corpora, 2017.

[55] Ellie Small, Javier Cabrera, John B. Kostis, and William Kostis. Abstract mining, 2018.

[56] Kazi Saidul Hasan and Vincent Ng. Automatic keyphrase extraction: A survey of the state of the art. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1262–1273, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[57] Zakariae Alami Merrouni, Bouchra Frikh, and Brahim Ouhbi. Automatic keyphrase extraction: a survey and trends. *Journal of Intelligent Information Systems*, Vol. 54, pp. 391 – 424, 2019.

[58] Kamil Bennani-Smires, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl, and Martin Jaggi. Simple unsupervised keyphrase extraction using sentence embeddings. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pp. 221–229, Brussels, Belgium, October 2018. Association for Computational Linguistics.

[59] Debanjan Mahata, John Kuriakose, Rajiv Ratn Shah, and Roger Zimmermann. Key2Vec: Automatic ranked keyphrase extraction from scientific articles using phrase embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 634–639, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[60] Xinnian Liang, Shuangzhi Wu, Mu Li, and Zhoujun Li. Unsupervised keyphrase extraction by jointly modeling local and global context. *CoRR*, Vol. abs/2109.07293, , 2021.

[61] Ellie Small and Javier Cabrera. Principal phrase mining, 2022.

[62] 優介木村, 和馬楠, 優香寺本, 賢治波多野. 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, aug 2020.

[63] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomás Mikolov. Enriching word vectors with subword information. *CoRR*, Vol. abs/1607.04606, , 2016.

[64] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics.

[65] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, Vol. abs/1810.04805, , 2018.

[66] Sakae Mizuki Shogo Matsuno and Takeshi Sakaki. 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.

[67] Radim Řehůřek and Petr Sojka. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50, Valletta, Malta, May 2010. ELRA.

[68] Kazuma Takaoka, Sorami Hisamoto, Noriko Kawahara, Miho Sakamoto, Yoshitaka Uchida, and Yuji Matsumoto. Sudachi: a Japanese tokenizer for business. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).

[69] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. *CoRR*, Vol. abs/1310.4546, , 2013.

付録

付録. A　使用した単語リスト（実験１）

A-1　犯罪語リスト

表 A.1: コーパス分けに利用した犯罪語リスト

	単語	意味
1	ホワイトウィドー	大麻
2	キメセク	ドラッグをキメながらの性行為
3	ガンコロ	覚醒剤
4	テレグラム	テレグラム
5	クッシュ	大麻
6	ホ別	ホテル代別（援助交際を暗示）
7	穂別	ホテル代別（援助交際を暗示）
8	わりきり	援助交際
9	円光	援助交際
10	援交	援助交際

A-2　既知の隠語リスト

表 A.2: 実験に使用した既知の隠語リスト

	単語	意味
1	手押し	手渡し
2	ハイグレ	マリファナの品質（ハイグレード（上質））
3	ウィドー	大麻の品種名
4	クラック	コカイン
5	シャブ	覚醒剤
6	援交	援助交際
7	円光	援助交際
8	キメセク	ドラッグをキメながらの性行為
9	エログル	援助交際用のグループトーク
10	オフパコ	援助交際

付録. B　検出した隠語一覧 (実験 1)

表 B.3: 検出した隠語一覧（実験１）

	単語	意味
1	グミ	大麻や覚醒剤の形状やその意味から大麻や覚醒剤そのものを隠喩する場合もある
2	ヘイズ	大麻の品種名
3	スカンク	大麻の品種名（スカンク No.1）や全ての高 THC 大麻品種の総称
4	パープルクッシュ	大麻の品種名（パープルクッシュ）
5	ブラック	大麻の品種名（ドバイブラック，ブラックダイヤモンド）
6	レモン	大麻の品種名（レモンヘイズ，レモンクッシュ，レモンスカンク等）
7	ハイレギュラー	大麻の品質（上質）
8	ホワイトクッシュ	大麻の品種名
9	アイス	覚醒剤
10	リキッド	大麻リキッド
11	ディーゼル	大麻の種や大麻の品種名（サワーディーゼル，レッドディーゼル ブラックディーゼル等）
12	ノーザン	大麻の品種名（ノーザンライト，ノーザンライツ等）
13	紙	LSD
14	グリーン	大麻そのものや大麻の品種名（グリーンクラック等）
15	ホワイト	大麻の品種名（ホワイトウィドー）
16	オレンジ	大麻の品種名（オレンジバッズ）
17	クッシュ	大麻の品種名
18	スーパー	大麻名の品種名（スーパースカンク，スーパーレモンヘイズ，スーパーシルバーヘイズ，スーパーレモンスカンク等）
19	レッド	大麻（レッドシャーク，レッドディーゼル等）

付録. C　使用した単語リスト (実験 3)

C-1　実験に使用した隠語リスト (実験 3)

表 C.4: 既知の隠語リスト

	単語	意味
1	手押	手渡し
2	ハイグレ	マリファナの品質 (ハイグレード (上質))
3	ウィドー	大麻の品種名
4	クラック	コカイン
5	シャブ	覚醒剤
6	リキッド	大麻の液体リキッド
7	グミ	大麻のグミ状
8	レモン	大麻の種や大麻の品種名
9	ホフマン	大麻の種や大麻の品種名
10	罰	覚せい剤
11	アイス	覚せい剤
12	紙	大麻
13	クッシュ	大麻の種や大麻の品種名
14	レモンクッシュ	大麻の種や大麻の品種名
15	レモンスカンク	大麻の種や大麻の品種名
16	マンゴークッシュ	大麻の種や大麻の品種名
17	キャンディークッシュ	大麻の種や大麻の品種名
18	サワーディーゼル	大麻の種や大麻の品種名

付録. D　検出した隠語一覧表 (実験 3)

表 D.5: 複合語型隠語検出実験（実験３）単体で確認できたもののみ 1/3

	単語	意味
1	チャリ	覚醒剤
2	オレンジバッズ	大麻の品種名もしくは製品名
3	氷	覚醒剤
4	ペン	覚醒剤
5	ブルーチーズ	大麻
6	タンジェリン	大麻の品種名もしくは製品名 (タンジェリンドリーム)
7	キャンディー	大麻の品種名もしくは製品名 (キャンディークッシュ, キャンディージョイント，ブレインキャンディー, サワーキャンディー)
8	アムネシア	大麻の品種名もしくは製品名 (アムネシアヘイズ)
9	バッツ	大麻の品種名もしくは製品名 (オレンジバッツ, トップバッツ，オレンジトップバッツ)
10	チャリンコ	覚醒剤
11	ストロベリークッシュ	大麻の品種名
12	クッキーズ	大麻の品種名もしくは製品名 (ストロベリークッキーズ)
13	ハワイアン	大麻の品種名もしくは製品名 (ハワイアンクッシュ, ハワイアンヘイズ，ハワイアンスノー)
14	バブルガム	大麻の品種名もしくは製品名 (バブルガムクッシュ)
15	ブラント	大麻の品種名もしくは製品名 (パワープラント)
16	ブルドッグ	大麻の品種名もしくは製品名 (ブルドッグヘイズ)
17	ハイブリッド	大麻の品種名
18	ガールスカウト	大麻の品種名もしくは製品名 (ガールスカウトクッキーズ)
19	ノーザン	大麻の品種名もしくは製品名 (ノーザンライト, ノーザンライトヘイズ)
20	鼻	覚醒剤
21	ブルドック	大麻の種や大麻の品種名 (ブルドックヘイズ)
22	パイナップル	大麻の品種名もしくは製品名 (パイナップルエクスプレス, パイナップルチャンク)
23	ディーゼル	大麻の品種名もしくは製品名 (サワーディーゼル, ホワイトディーゼル)
24	リーフ	大麻の品種名もしくは製品名 (レモンリーフ，シュガーリーフ)
25	マンゴー	大麻の品種名もしくは製品名 (マンゴークッシュ, マンゴーラッシュ，マンゴーヘイズ)

表 D.6: 複合語型隠語検出実験（実験３）単体で確認できたもののみ 2/3

	単語	意味
26	ハイレギュ	大麻の品種 (ハイレギュラー)
27	チーズケーキ	大麻の品種名もしくは製品名 (ストロベリーチーズケーキ)
28	クッキー	大麻の製品名 (エディブルクッキー, ガールスカウトクッキー，パープルクッキー)
29	アップル	大麻の品種名もしくは製品名 (グリーンアップル, ファンタアップル)
30	ストロベリー	大麻の品種名もしくは製品名 (ストロベリータホ, ストロベリーコフ)
31	蜂蜜	大麻の種や大麻の品種名
32	ブルーベリー	大麻の品種名もしくは製品名
33	OG	大麻の品種名もしくは製品名 (OG クッシュ, ノトーリアス OG)
34	サワー	大麻そのものや大麻の品種名 (サワーディーゼル, サワークッシュ)
35	シルバー	大麻そのものや大麻の品種名 (シルバーヘイズ, シルバーバック，シルバーフェイズ)
36	スカンク	大麻そのものや大麻の品種名 (レモンスカンク, ダイアモンドスカンク)
37	ホワイト	大麻そのものや大麻の品種名 (ホワイトウィドウ, ホワイトクッシュ，ホワイトライノジョイント)
38	オレンジ	大麻そのものや大麻の品種名 (オレンジバッド, オレンジバッツ，オレンジトップバッツ，オレンジヘイズ カリフォルニアオレンジ，オレンジクッキー)
39	グリーン	大麻そのものや大麻の品種名 (グリーンクラック グリーンヘイズ)
40	AK	大麻の製品名 (AK47)
41	ヘイズ	大麻そのものや大麻の品種 (グリーンヘイズ, ロイヤルヘイズ，シルバーヘイズ, スーパーレモンヘイズ等)
42	チーズ	大麻そのものや大麻の品種 (UK チーズ, ロイヤルチーズ，レモンチーズ, ブルーチーズ，ストロベリーチーズケーキ等)

表 D.7: 複合語型隠語検出実験（実験３）単体で確認できたもののみ 3/3

43	ピンク	大麻そのものや大麻の品種 (ピンクパンサー, ピンクスターバースト，ピンクランツ, ピンクヘイズ，ピンクレモネード等)
44	パープル	大麻そのものや大麻の品種 (パープルヘイズ, パープルクッキー，パープルクッシュ パープル OG クッシュ等)
45	ゴリラ	大麻そのものや大麻の品種 (ゴリラグルー, ロイヤルゴリラ，ゴリラ OG，ゴリラクッキー等)
46	野菜	大麻
47	ホワイトウィド	大麻の品種名
48	ジャックヘラ	大麻の品種名
49	キメセク	ドラッグをキメながらの性行為
50	アフガンクッシュ	大麻の品種名
51	クリティカルクッシュ	大麻の品種名
52	ストロベリーコフ	大麻そのものや大麻の品種名
53	サワークッシュ	大麻そのものや大麻の品種名
54	ハワイアンクッシュ	大麻そのものや大麻の品種名
55	パープルクッシュ	大麻そのものや大麻の品種名
56	バナナクッシュ	大麻そのものや大麻の品種名
57	優勢ハイブリッド	大麻の品質
58	エディブル	大麻そのものや大麻の品種 (エディブルクッキー等)
59	ババクッシュ	大麻そのものや大麻の品種名
60	キングクッシュ	大麻そのものや大麻の品種名
61	コットンキャンディクッシュ	大麻そのものや大麻の品種名
62	ソマンゴ	大麻そのものや大麻の品種名
63	モーリー	大麻そのものや大麻の品種名 (モーリーピュア)
64	ミルキーウェイ	大麻そのものや大麻の品種名
65	ジャックヘラー	大麻そのものや大麻の品種名
66	レモンスカンク	大麻そのものや大麻の品種名
67	グリーンクラック	大麻そのものや大麻の品種名
68	サワーディーゼル	大麻そのものや大麻の品種名
69	ホワイトウィドー	大麻そのものや大麻の品種名
70	ゴリラグルー	大麻そのものや大麻の品種名
71	ホワイトウィドウ	大麻そのものや大麻の品種名
72	パイナップルチャンク	大麻そのものや大麻の品種名
73	ブルードリーム	大麻そのものや大麻の品種名

研究業績

学術雑誌

1. **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， Vol.142， No.2， pp.177-189，2022

2. **Takuro Hada**, Yuichi Sei, Yasuyuki Tahara and Akihiko Ohsuga: Codeword Detection, Focusing on Differences in Similar Words Between Two Corpora of Microblogs, Annals of Emerging Technologies in Computing, Vol.5, No.2, pp.90-102, 2021

国際会議

3. **Takuro Hada**, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga Detection of Compound-type Dark Jargons Using Similar Words,
15th International Conference
on Agents and Artificial Intelligence (ICAART), Vol.1, pp.427-437,2023.

4. **Takuro Hada**, Yuichi Sei, Yasuyuki Tahara and Akihiko Ohsuga: Codewords Detection in Microblogs
Focusing on Differences in How Words are Used Between Two Corpora,
3rd IEEE International Conference on Computing, Electronics & Communications Engineering (iCCECE), pp.103-108,2020.

国内大会・研究会

5. **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, SMASH22 Summer Symposium, 信学技報, Vol.122, No.186, AI2022-28, pp.58-63, 2022

6. **羽田拓朗**, 清雄一, 田原康之, 大須賀昭彦, コーパス間での単語の類似語の差異を利用した複合語型隠語の検出, 電子情報通信学会 人工知能と知識処理研究会, 信学技報, Vol.120, No.362, AI2020-32, pp.50-55, 2021

7. **羽田拓朗**, 清雄一, 田原康之, 大須賀昭彦, コーパス間で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, SMASH20 Summer Symposium, 情報処理学会研究報告, Vol.2020-ICS-200, No.2, pp.1-8, 2020

8. **羽田拓朗**, 清雄一, 田原康之, 大須賀昭彦: コーパス間の単語の用途の差異に着目したマイクロブログにおける隠語検出, 電子情報通信学会総合大会 情報・システム講演論文集１, p.66, 2020

受賞

9. **羽田拓朗**, JP 生きがい振興財団警察研究論文奨励賞「情報通信研究の部」最優秀賞, 2023.

10. **羽田拓朗**, 清雄一, 田原康之, 大須賀昭彦, SMASH2021 Summer Symposium, 奨励賞, 2022.

11. **羽田拓朗**, JP 生きがい振興財団警察研究論文奨励賞「情報通信研究の部」優秀賞, 2021

12. **羽田拓朗**, 令和 3 年度 電気通信大学学生表彰

13. **羽田拓朗**, 令和 5 年度 電気通信大学学生表彰

著者略歴

羽田　拓朗（はだ たくろう）

1980 年 1 月 30 日	三重県亀山市に生まれる
1998 年 3 月	私立高田中・高等学校 卒業
1998 年 4 月	国立大学法人静岡大学 情報学部 情報科学科 入学
2002 年 3 月	国立大学法人静岡大学 情報学部 情報科学科 卒業
2002 年 4 月	国立大学法人静岡大学 大学院 情報学研究科
	情報学専攻 博士前期課程 入学
2004 年 3 月	国立大学法人静岡大学 大学院 情報学研究科
	情報学専攻 博士前期課程 修了
2004 年 4 月	NDS 株式会社 入社
2008 年 3 月	NDS 株式会社 退社
2008 年 4 月	中部管区警察局 入庁
	現在まで，警察庁刑事局組織犯罪対策部
	組織犯罪対策第一課暴排係　在籍
2019 年 4 月	国立大学法人 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 入学
2021 年 3 月	国立大学法人 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
2021 年 4 月	国立大学法人 電気通信大学 大学院 情報理工学研究科
	情報学専攻 博士後期課程 入学
2024 年 3 月	国立大学法人 電気通信大学 大学院 情報理工学研究科
	情報学専攻 博士後期課程 修了予定