きたあああああああああああああああああ！！！！！１１：マイクロブログを用いた教師なし叫喚フレーズ抽出

浅井 洋樹[†1]　秋岡 明香[†2]　山名 早人[†3]

†1 早稲田大学大学院基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1
早稲田大学メディアネットワークセンター 〒169-8050 東京都新宿区戸塚町 1-104
†2 早稲田大学 IT 研究機構 〒169-8555 東京都新宿区大久保 3-4-1
†3 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1
国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2
E-mail: †1†3 {asai, yamana}@yama.info.waseda.ac.jp, †2 akioka@muraoka.info.waseda.ac.jp

あらまし 短文での投稿に特化しているマイクロブログでは，着想から投稿までのタイムラグが短いため，突発的な感情を表現する投稿が頻繁に行われている．中でも叫ぶような強い感情を表現する際には，「○○きたあああああ」のような語尾の母音を繰り返す表現がしばし用いられる．本稿ではこの「きたあ」のような母音繰り返しが発生するフレーズを叫喚フレーズと定義し，叫喚フレーズを含むメッセージを調査することで，叫喚メッセージが発生する状況を明らかにする．さらに，辞書を必要としない統計的手法により叫喚フレーズを抽出する手法を提案し，マイクロブログ上での叫喚フレーズによる強い感情表現を抽出する．抽出評価として，メッセージ全体や TV 番組実況などのイベント限定した叫喚フレーズ抽出を行い，叫喚フレーズ抽出の性能について評価する．

キーワード マイクロブログ，叫喚フレーズ，不自然言語処理，盛り上がり検出

1. はじめに

近年のマイクロブログサービスの普及により，ユーザの短文メッセージがインターネット上に頻繁に投稿されている．代表的なマイクロブログサービスである Twitter での投稿数は 2012 年 6 月の時点で 1 日に 4 億件[1]にものぼり，これらの膨大なデータに対する研究が盛んに行われている[1].

この近年の研究対象となっているマイクロブログの大きな特徴として，一度に投稿できる文字数に厳しい制約を設けていることがあげられる．文字数の制約により投稿を思い立ってから投稿が完了するまでのタイムラグが短くなるため，ユーザの突発的な感情を表現するメッセージが数多く投稿されることがマイクロブログの一つの特徴となっている．このような突発的な感情表現が頻繁に行われる状況の代表例として TV 番組に関するメッセージを投稿する TV 番組実況という行為が挙げられる[2]．特に日本では TV 番組実況は盛んに行われており，2011 年 12 月 9 日に放映された番組においては放映時点で過去最大となる 1 秒あたり 2 万 5 千を超えるメッセージが投稿された[3].

マイクロブログ上で投稿される突発的な感情を表現したメッセージに含まれるフレーズは元の単語を何らかの形で変形した単語（例：わああい，やったーー）がしばし含まれる．このような崩れた表記の処理を行う研究が不自然言語処理に関する研究の一部として行われている[2][3][4].

日本語のメッセージにおいて，崩れた表記の中でも特に叫ぶような強い感情を表現する際には「○○きたあああああ」のような語尾の母音を繰り返す表現がしばし用いられることが指摘されている[5]．本稿ではこの「きたあ」のような語尾の母音繰り返しが発生する語を「叫喚フレーズ」と定義し，次の項目について述べる．

（１） 叫喚フレーズの調査
Twitter 上でのメッセージを分析し，叫喚フレーズが含まれるメッセージが発生する状況を調査・分類する．また，特定のイベントにおける叫喚フレーズの発生頻度の推移についても調査を行う．

（２） 教師なし叫喚フレーズ抽出手法の提案
語尾に母音の繰り返しが発生する叫喚フレーズを，辞書を必要としない統計的な方法によって抽出する手法を提案する．

（３） 叫喚フレーズ検出性能の評価
Twitter のすべてのメッセージや TV 番組実況などのイベントに関するメッセージに限定した条件で，叫喚フレーズ抽出を行い，検出性能の評価を実施する．

[1] Twitter Advertising (@TwitterAds), https://twitter.com/TwitterAds/statuses/210867782361948161.
[2] Variety Media, LLC, Social TV chatter grows 800% on Twitter over 2012, http://www.variety.com/article/VR1118063975/.
[3] Twitter 日本公式アカウント (@twj), https://twitter.com/twj/status/146751303584980992.

2. 叫喚フレーズの定義と調査

本節では本研究で述べる叫喚フレーズの定義について 2.1 節で述べた後に，叫喚フレーズが含まれるマイクロブログのメッセージに関して調査を行った結果を述べる．2.2 節では Twitter 上で出現する叫喚フレーズについて述べ，分類を行う．2.3 節では調査対象のメッセージを各イベントに関連したものに限定し，叫喚フレーズが出現するメッセージの出現頻度における時系列変化について調査した結果を述べる．

2.1. 叫喚フレーズの定義

はじめに本研究の対象とする叫喚フレーズの定義について述べる．ユーザが強い感動を表現する際に，語尾に母音を繰り返す表現をマイクロブログで用いることを著者らによる以前の調査[5]によって明らかにした．本研究ではこの調査結果を参考として叫喚フレーズを次のように定義する．

- 語尾の母音が 3 回以上繰り返して付加されている
  例）*うわあああああああ*，*ねむいいぃぃぃぃぃ*
- 母音は大文字，小文字を区別しない
- 母音はひらがな，カタカナの大小文字すべて

また，本稿では叫喚フレーズが現れているマイクロブログのメッセージを叫喚メッセージ，語尾の母音を繰り返す表現を叫喚表記と呼ぶことにする．

以上の叫喚フレーズの定義にもとづいて，本研究では叫喚メッセージを以下の正規表現によって抽出する．

あ{3,}|い{3,}|う{3,}|え{3,}|お{3,}|ぁ{3,}|ぃ{3,}|ぅ{3,}|ぇ{3,}|ぉ{3,}|ア{3,}|イ{3,}|ウ{3,}|エ{3,}|オ{3,}|ァ{3,}|ィ{3,}|ゥ{3,}|ェ{3,}|ォ{3,}

この正規表現にマッチしたメッセージを抽出することで，母音が 2 連続となる単語（例：かわいい）による誤検知を避けつつ叫喚メッセージを抽出可能であると考えられる．次節より本正規表現で抽出した叫喚メッセージについて調査した結果を述べる．

2.2. 叫喚メッセージの分類

まず叫喚フレーズが含まれるメッセージの内容について調査を行った．そこで本節では Twitter に投稿されている叫喚メッセージを抽出し，分類調査を行った．調査対象としたデータは次のとおりである．

- データ収集期間：
  2012 年 12 月 19 日～25 日
- 収集対象データ：
  サンプリングされた Twitter メッセージ
- 総収集メッセージ数：2,967,165
- 叫喚メッセージ数：59,707（全体の 2.01%）

なお Twitter メッセージの取得には，サンプリングされたメッセージが取得できる Twitter Streaming API (Sample)[4] を利用した．

収集した叫喚メッセージの中からランダムサンプリングにより 100 メッセージを抽出し，叫喚の要因別に筆者による分類を実施した．分類の項目とその実例を次に示す．

1. 外因による叫喚（外因）
   TV 番組などの他メディアに対する叫喚や，実世界のイベント・事象に対する叫喚
   ・あああああああああああああムギちゅわあああああああああああああああん #k_on
   ・年賀状印刷ミスったぁぁぁ！！
   ・AKB くるううううううう～！
2. 内因による叫喚（内因）
   自身の生理的な現象に対する叫喚や，奮起によって生じた叫喚
   ・うぁー、眠いいいいてか、雨降っとる！プレゼント濡れてないかなぁ？(￣▽￣;)
   ・ぬぅおおおおおおおがんばらなくちゃあああああああ
   ・帰りたいいいいいいいっ！！！
3. 会話より生じた叫喚（会話）
   マイクロブログ上でのユーザ同士の会話によって発生した叫喚
   ・@wahsing_7 わかるううううう！
   ・@aniota44　ああああ(´・ω・`)ちょっと凹みますね(´；ω；`)
4. 判断不可（不明）
   メッセージ単体からは判断できない叫喚
   ・ふおおおおお
   ・ひぃいぃぃぃいいぃいいいいぃぃいぃぃぃいいぃいぃぃいいぃいいぃ
5. その他・非叫喚（その他）
   分類 1 ～ 4 に分類されないメッセージ，または叫喚が確認できないメッセージ
   ・@egne_5nzer 紫原「ん…ふあぁぁぁ…おーはーよー…」　赤司「今日のおやつは…」　紫原「まいうぼおぉおおおうっ！！！！」　黒子「目が覚めたみたいですね」
   ・メリークルスニク！！ #メリークルスニクということでクルスニク一族クラスタさん集まれええええ
   ・┗(^o^)┛イタ電するぞぉおおおｗｗ(^o^)?┐もしもしｗｗｗｗｗｗｗｗ(^o^)Γ ?ﾁﾝｯ

[4] Twitter Streaming API (Sample), https://dev.twitter.com/docs/api/1.1/get/statuses/sample.

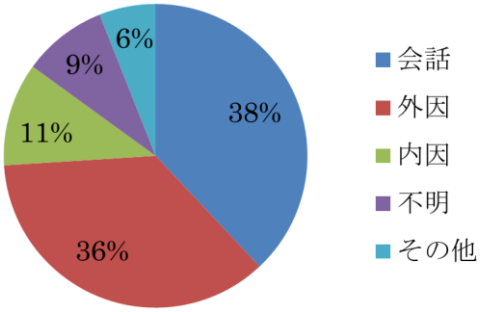

図 1 叫喚メッセージの分布

上記の項目に従って分類を行い，項目が占める割合を調査した．調査の結果を示した図 1 より，その他の割合が 6%となっていることが確認でき，2.1 節で示した叫喚フレーズの定義となる正規表現によってマイクロブログ上の叫喚を高い精度で抽出可能なことがわかる．また，本研究で定義した叫喚メッセージの内容としては，ユーザ間の会話や外的な要因から発生したもので 74%を占めていることが確認できる．

2.3. 共通要因によって同時発生する叫喚の調査

2.2 節の結果より，叫喚メッセージは外的な要因やユーザの会話から生じるものが多く占めるという結果が得られた．本節では，この中でも外因によって複数のユーザが同時に叫喚する状況について調査を行う．

複数のユーザが外的な共通する要因によって同時に叫喚が発生する代表例として TV 番組実況が挙げられる．TV 番組実況は放映中のテレビ番組を視聴中に，番組の内容に対する感想や感情をマイクロブログ等に投稿する行為であり，ソーシャルビューイングとも呼ばれている．TV 番組の内容に応じてメッセージ数の変動や，出現するキーワードの頻度が変化するため，実況メッセージを利用したメタデータ抽出[6]や重要シーン検出[7]といった研究が行われている．

本節ではこの TV 番組実況メッセージ数の時系列変化を調べ，他メディアと連動して発生する叫喚について調査する．調査対象としたデータを次に示す．

- 対象イベント：テレビ番組実況
  「エヴァンゲリヲン新劇場版:破 TV 版」
- イベント期間：
  2012 年 11 月 16 日 21 時～23 時
- データ収集期間：
  2012 年 11 月 16 日 20 時～24 時
- 収集対象ハッシュタグ：
  #エヴァ実況,#EVA,#エヴァ,#エヴァ,#エヴァンゲリオン,#エヴァンゲリヲン破,
  #エヴァ破,#破,#evangelion,#join_eva,#ntv,
  #金曜ロード SHOW
- 総収集メッセージ数：266,748
- 叫喚メッセージ数：14,495（全体の 5.43%）

なお番組に関連するハッシュタグが付加されたメッセージの取得には，Twitter Streaming API (Filter) [5]を利用した．

収集したデータに対して，1 分毎のメッセージ数の変化を図 2 に示す．番組に関するすべてのメッセージ数の推移に関しては，番組開始と同時に急激に増加し，番組放映中では時刻によって上下を繰り返している．また番組放送終了後はメッセージ数がゆるやかに減少していることが確認できる．一方，叫喚メッセージ数の推移に関しては番組放映前後ではほぼ出現せず，出現は番組放映時間帯に限られている．また，番組に関する全メッセージ数の変化と比較して，時間帯ごとにより顕著な変化が発生していることが確認できる．

次に叫喚メッセージが急激に増加した時間帯において出現するメッセージの調査を行う．時系列上でメッセージ数が急激に変化した時刻を検出する手法として Takeuchi らの変化点検出手法[8]を利用した．Takeuchi らの手法によって得られる急激にメッセージ数が変化した時刻を検出することが可能となる．検出された時刻とメッセージの一部を以下にまとめる．

1. 全実況メッセージのみ急激な変化
   A) キャラクターのセリフ（21:22）
      ・「はじめまして。お父さん。」 金曜ロード SHOW!『エヴァンゲリヲン新劇場版』2 週連続 ムーヴィシンクロナイザ #ntv #join_eva - http://t.co/Va0sY4BT
      ・カヲルくーん！！ #eva
   B) 番組終了（22:52）
      ・おわた #eva
      ・終わり！？？？？ #ntv #eva_ha
2. 叫喚実況メッセージのみ急激な変化
   A) 不幸な展開（22:12）
      ・うあああああああああああああああああああああ #eva
      ・それはらめえええええええ #eva
   B) キャラクターの登場（22:45）
      ・カヲル君きたあああああああああああああ！ #エヴァ
      ・ホモオオオオオオオオ #eva

[5] Twitter Streaming API (Filter), https://dev.twitter.com/docs/api/1.1/post/statuses/filter.

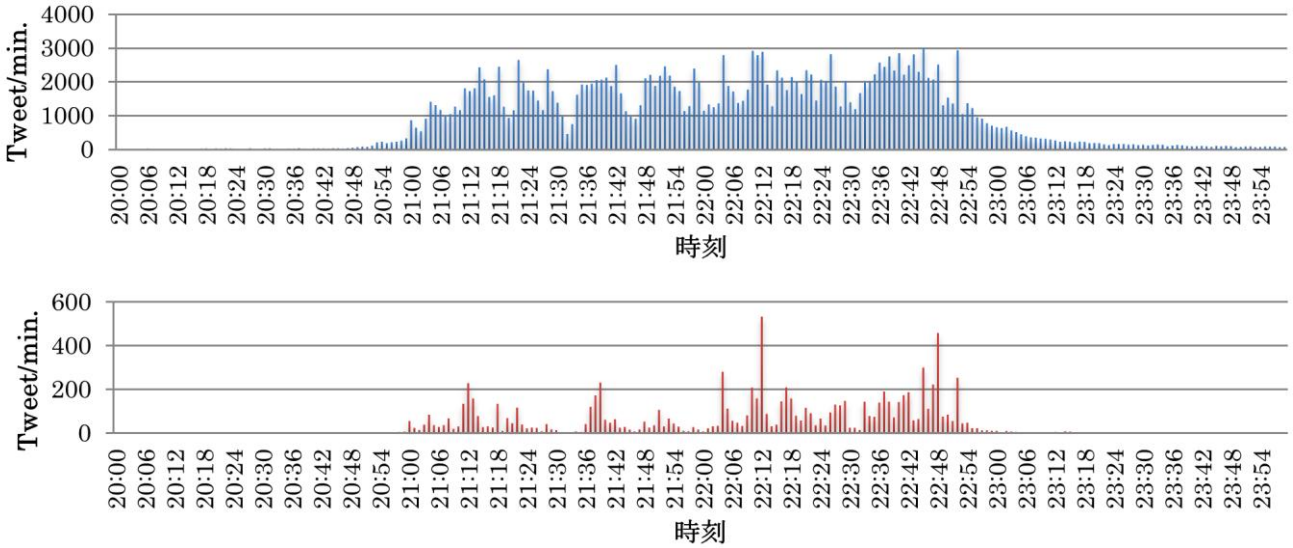

図 2 TV 番組実況（2012 年 11 月 16 日 21 時～23 時：エヴァンゲリヲン・破）の実況メッセージ数の変化．実況メッセージ全体（上），叫喚している実況メッセージ（下）

3. 全実況・叫喚共通の急激な変化
   A) クライマックスシーン（22:10）
      ・ｷﾀｰｰｰヽ(∀ﾟ)人(ﾟ∀ﾟ)人(ﾟ∀)ﾉｰｰｰ!! #eva
      ・そしてこの BGM #eva
   B) キャラクターの登場（22:48）
      ・あすかあああああああああああああ #エヴァ
      ・かわいいいいいいいいいいいい #eva

全実況メッセージのみが急激に変化した時刻においては，番組の開始や終了，印象の強いセリフに言及したメッセージが多数投稿されている．一方，実況に関する叫喚メッセージのみが急激に増加している時刻では，不幸な展開などの強烈な印象を与えるシーンや，一部のユーザに強い支持を得ているキャラクターに関するメッセージが多数確認された．全実況・叫喚メッセージ共通で急激な変化が確認された時刻においては，番組のクライマックスシーンや人気キャラクターの登場シーンといった大多数のユーザが盛り上がるシーンが検出される結果となった．

以上の 2 節の調査結果より，叫喚メッセージはユーザ同士の会話や外的な要因によって投稿されるメッセージが多くを占めることがわかった．また，TV 番組実況に関連する叫喚メッセージの時系列変化調査結果として，全実況メッセージと比較すると叫喚メッセージに絞って急激なメッセージ数の変化を検知することにより，次のようなシーンとキーワードが検出されることがわかった．

- ユーザの強い思い入れのある場面
- 強烈な印象を与える場面
- 一部のユーザ間で強い盛り上がりをみせる場面

3. 関連研究

2 節において叫喚メッセージの定義とその特徴について述べた．本節では叫喚表記に関連する研究として，英語圏での文字繰り返し強調表現や，日本語での崩れた表記に関する研究を関連する研究として述べる．

3.1. 英語の繰り返し表現に関する研究

本研究の対象としている日本語圏だけでなく，英語圏においても文字を重ねて繰り返す表現が確認されている．Grinter ら[9]の報告によると，SMS 等の文字数制限が厳しいような環境下において，文字を繰り返して語を強調する記述が現れることを確認している．またこのような文字を繰り返す記述を行なう場合は，筆者が重要な情報を伝える意図があると述べている．

Brody ら[4]は，マイクロブログ上で出現する英単語の文字を何度も繰り返して表記する表現（例：Cool→Coooooooollllllllll）を Lengthening と定義し，マイクロブログ上で用いられる語の感情極性辞書を生成した．Brody らの調査によると，Lengthening が行われる語は，主観性と感情と強い関わりがあることが述べられている．また，Lengthening の文字が繰り返されている部分を除去することで正規化し，元の語の抽出を行っている．これにより，一般的な辞書に登録されていないようなマイクロブログ特有の語を抽出している．

本研究の対象としている叫喚フレーズも，この英語の繰り返し表現に関する研究と同様に文字の繰り返しが行われている．繰り返し文字の正規化の部分において Brody らの手法を本研究に応用する．

3.2. 日本語の崩れた表記に関する研究

本節では，本研究の対象としている叫喚フレーズのような日本語の崩れた表記に関する研究について述べる．Web テキストにおいて，長音記号の挿入や置換，小書き文字の挿入など，様々な崩れた表記が出現することが笹野らによって報告されている[2]．この中においても本研究の対象としている叫喚フレーズの表記が「母音の挿入」として報告されている．また，著者らの以前の研究[5]において，語尾に母音を繰り返す表現は強い感情を表しているものが多いという結果が得られている．

一方，山本ら[10]は動画共有サービスであるニコニコ動画[6]の時刻同期コメントを用いて楽曲動画の印象を推定する研究を行っており，時刻同期コメントの表記のパターンにおいて本研究での叫喚表記と同様な母音繰り返しの表記が多く発生していることが報告されている．この母音の繰り返し表記に対して，山本らは繰り返し文字の正規化を行い，母音繰り返しの数によらず同一の素性として扱えるようや処理を実施している．

ニコニコ動画ではコメント文字数制限は 75 文字であり，Twitter の 140 文字と比較しても更に短いため，投稿されるテキストは「〇〇きたあああ」など 1 フレーズであることが多く，正規化を行ったあとのコメント全体を 1 つの素性として扱うことができる．しかし，本研究の対象としているマイクロブログでは文章の一部に叫喚フレーズが現れていることも少なくない．そこで本研究では文章中より一般的に行われる叫喚フレーズを抽出する方法を検討する．

4. 叫喚フレーズ抽出手法

本節では 2 節の叫喚フレーズに関する調査と，3 節の関連する研究を踏まえ，マイクロブログ上で一般的に行われている叫喚フレーズを抽出する手法を提案する．まず叫喚フレーズの特徴として，母音の回数が不定であることや，インターネットスラングのような一般的な辞書では対応できない語が頻繁に用いられることがあげられる．これを踏まえて本研究の叫喚フレーズ抽出のポイントは次のとおりである．

- 母音の繰り返し回数に依存せずに，テキスト中から叫喚フレーズの位置を発見可能
- 辞書を用いない統計的な手法により，未知語の抽出が可能

これらの特徴をもつ叫喚フレーズ抽出手法について，以下に詳細を述べる．

[6] ニコニコ動画, http://www.nicovideo.jp.

4.1. 叫喚メッセージの抽出と正規化

はじめに，全メッセージの中から叫喚メッセージを抽出，前処理を行ったあとに不定回数の母音繰り返し対応の正規化を行う．手順と例を以下に示す．

1. 前処理としてメンション（@username），ハッシュタグ，URL，日本語以外の文字，記号をメッセージから除去する
2. 叫喚メッセージを 2.1 節で定義した正規表現を用いて抽出する．
   例）年賀状印刷ミスったぁぁぁ
   うわあああ最悪だめだぁああ
3. 繰り返し母音を大文字化
   例）年賀状印刷ミスったあああ
   うわあああ最悪だめだあああ
4. すべての繰り返し母音部分に対して，母音一文字とそれ以前の文字列を抽出
   例）年賀状印刷ミスったあ
   うわあ
   うわあああ最悪だめだあ

以上のステップにより抽出し，正規化されたテキスト集合をもとに統計的に叫喚フレーズ抽出を行う．

4.2. 反転文字列順序木（RSO-Tree）の構築

次に 4.1 節で正規化されたテキスト集合を利用して叫喚フレーズを抽出する．抽出にあたっては，本研究で提案する反転文字列順序木(以下 RSO-Tree と呼ぶ)を利用する．RSO-Tree は反転した文字列の集合に対し，文字の順序の出現頻度を表す木である．構築例を図 3 に示す．

カッコ（[]）の中の数字は反転文字列の出現頻度を表している．木の構築にあたって，ある閾値以下の出現頻度となるノードは削除するという条件を設ける．これにより，出現頻度の高いフレーズのみが残るようになり，マイクロブログ上で多く用いられている叫喚フレーズが抽出可能となる．RSO-Tree の構築が終わった後に，リーフからルートまで辿ることで叫喚フレーズとその出現頻度を得ることができる．

正規化テキスト集合

神きたあ，優勝きたあ，うわあ，やったあ

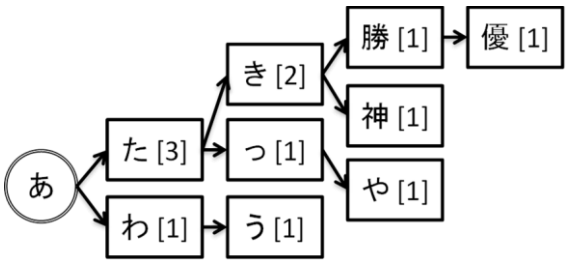

図 3　RSO-Tree の構築例

5. 叫喚フレーズ抽出性能評価

収集したマイクロブログのメッセージ全体や，特定イベントに関するメッセージ集合に対して，本研究の叫喚フレーズ抽出を行い，実際の抽出性能について評価する．実験に用いたデータは 2.3 節で述べたエヴァンゲリヲン TV 番組実況（TV 実況 A）に加えて，天空の城ラピュタ TV 番組実況（TV 実況 B）とサンプリングで取得したメッセージ（全体）である．TV 実況 B のデータの内容を次に示す．

- 対象番組：天空の城ラピュタ（アニメ映画）
- データ収集期間：
  2011 年 12 月 9 日 18 時～24 時
- 収集対象ハッシュタグ：#laputa
- 総収集メッセージ数：347,925
- 叫喚メッセージ数：7,655（全体の 2.20%）

また収集した全メッセージのデータ内容を次に示す．なおデータの収集には Twitter Streaming API (Sample)を利用し，リツイートは除外した．

- データ収集期間：
  2012 年 1 月 1 日～12 月 31 日
- 収集メッセージ総数：1,057,086,851
- 日本ユーザによるメッセージ数：122,469,244
- 叫喚フレーズ数：2,832,380
  （日本語ユーザによるメッセージの 2.31%）

まず TV 実況 A のデータを用いて RSO-Tree の構築を行った．出現頻度の閾値を 10 に設定した構築結果を図 4 に示す．「きたあ」や「うわあ」といった叫喚する表現や登場人物である「アスカ」などのイベント特有の語が出現していることが確認できる．次に木の構築結果から叫喚フレーズを抽出し，出現頻度が高い叫喚フレーズの順でソートを行った．全叫喚メッセージから得られた叫喚フレーズのうち，出現頻度が上位のものを表 1 に示す．同様に，TV 実況 A・B から得られた叫喚フレーズうち，出現頻度が上位のものを表 2，表 3 にそれぞれ示す．なお α (%)は RSO-Tree 構築の際に設定した出現頻度の閾値を決定するためのパラメータであり，閾値は全叫喚メッセージに対する割合 α によって決定している．

叫喚フレーズの抽出結果を参照すると，叫喚メッセージ全体から得られた叫喚フレーズは，「やったあ」「いいなあ」「かっけえ」などの本人の感想を表すフレーズや，「ふお」「いえ」「にゃあ」などの叫びを表すフレーズが上位に多く存在することが確認できる．一方，TV 実況 A や B といったイベントに限定して抽出を行った結果を参照すると，「うわあ」「やめろお」といったあるシーンの感想を表すフレーズが抽出できていることが確認できた．また，「アスカきたあ」「目が目があ」といったイベント特有のフレーズが抽出されていることも確認でき，イベントに対するユーザの感情を表すフレーズを抽出できていることが確認できる．

6. まとめと今後の課題

本稿では，マイクロブログ上で行われている語尾の母音を繰り返すことで叫ぶような強い感情を表現する叫喚フレーズに着目し，教師データを用いずにフレーズを抽出する手法を提案した．本手法によって叫喚フレーズを抽出することで，マイクロブログ上での強い感情表現に用いられる語や，イベント特有の叫喚フレーズを抽出することが確認できた．

フレーズ抽出の今後の課題としては，表記の揺れやフレーズの途中で適切に区切れない問題の解決が挙げられる．加えて．本抽出手法を応用することで，イベントにおける盛り上がりとその要因の検出や，マイクロブログでの形態素解析精度を向上させる辞書としての活用ができるのではないかと考えている．

謝　辞

本研究の一部は，JST「多メディアＷｅｂ解析基盤の構築及び社会分析ソフトウェアの開発」によるものである．

参　考　文　献

[1] 奥村学, ”マイクロブログマイニングの現在”, 信学技報, Vol.111, No.427, NLC2011-59, pp.19-24, 2012.

[2] 笹野遼平, 鍛治伸裕, “不自然言語処理～枠に収まらない「リアルな」言語処理～ 2. 新しい語・崩れた表記の処理”, 情報処理, Vol.53, No.3, pp.211-216, 2012.

[3] B. Han and T. Baldwin, “Lexical normalization of short text messages: Maken sens a #twitter”, Proc. of ACL2011, pp.368-378, 2011.

[4] S. Brody and N. Diakopoulos, “Cooooooooooooooollllllllllll!!!!!!!!!!!!! Using Word Lengthening to Detect Sentiment in Microblogs”, Proc. of EMNLP2011, pp.562-570, 2011.

[5] 髙橋雄太, 片岡義雅, 浅井洋樹, 山本祐輔, 秋岡明香, 山名早人, “繰り返し表現を用いた感情を含むツイートの抽出手法”, DEIM2012, C3-6, 2012.

[6] 宮森恒, 中村聡史, 田中克己, “番組実況チャットを利用したテレビ番組のメタデータ自動抽出方式”, 情報処理学会論文誌 データベース, Vol.46, No. SIG 18, pp.59-71, 2005.

[7] 中澤昌美, 帆足啓一郎, 小野智弘, “Twitter によるテレビ番組重要シーン検出及びラベル付与手法”, DEIM2011, F5-6, 2011.

[8] J. Takeuchi, and K. Yamanashi, “A Unifying Framework for Detecting Outliers and Change Points from Time Series”, IEEE Trans. on KDE, Vol.18, Issue. 4, pp.482-492, 2006.

[9] R.E. Grinter and M. Eldridge, “Wan2tlk?: Everyday Text Messaging”, Proc. of CHI2003, pp.441-448, 2003.

[10] 山本岳洋, 中村聡史, “視聴者の時刻同期コメントを用いた楽曲動画の印象推定”, WebDB Forum 2012, 3A-1, 2012.

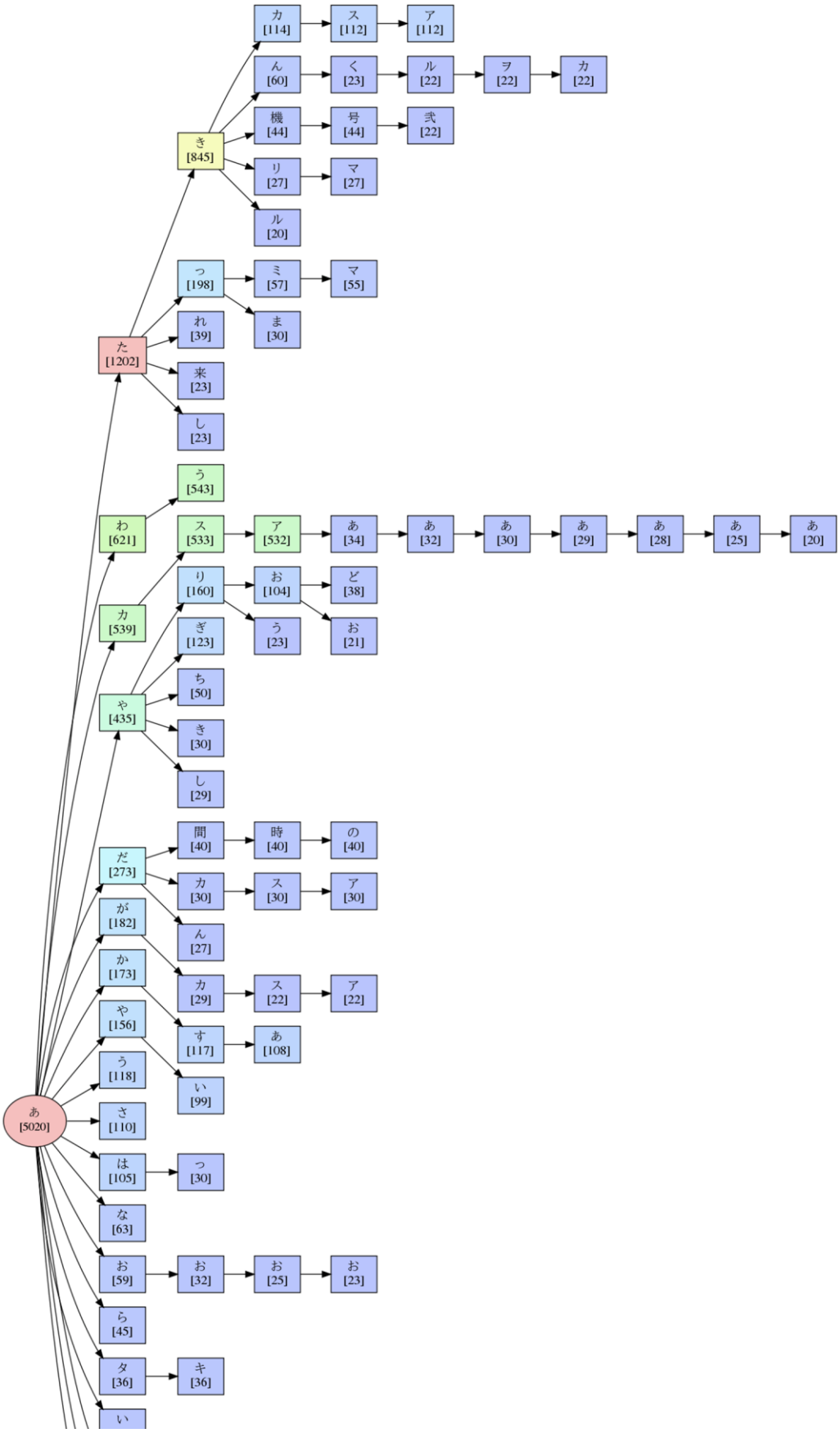

図 4　RSO-Tree の例（一部）：TV 実況「エヴァンゲリヲン・破」2012 年 11 月 16 日

表 1 全体から抽出した叫喚フレーズ (α=0.033)

[出現回数]	叫喚フレーズ
[19203]	ふお
[9714]	やったあ
[9433]	いいなあ
[9182]	もお
[9021]	やべえ
[8934]	すげえ
[8011]	いえ
[7258]	ふえ
[6644]	かっこい
[6396]	にゃあ
[6101]	してえ
[6071]	ウワア
[5970]	様あ
[5861]	ふあ
[5766]	ないよお
[5669]	かっけえ
[5652]	ウオア
[5649]	んあ
[5555]	ウオオオオオオア
[5135]	ただいまあ
[4885]	うめえ
[4855]	だろお
[4755]	ウオオオオア
[4693]	いてえ
[4501]	るぞお
[4464]	ひえ
[4454]	よっしゃあ
[4443]	よかったあ
[4392]	おはよお
[4381]	うぜえ
[4381]	おかえりい
[4358]	うお
[4289]	かなあ
[4235]	うおあ
[4209]	ぬお
[3957]	したい
[3942]	ウオオオオオオオア
[3933]	うわあ
[3912]	ぐあ
[3912]	やめてえ
[3900]	んなあ
[3839]	かよお
[3836]	くそお
[3792]	ぎい
[3786]	ってるう
[3757]	いお
[3750]	ありがとうございますう
[3699]	しいよお
[3692]	！うわあ
[3684]	ほしい

表 2　TV 番組実況「エヴァンゲリヲン・破」から抽出した叫喚フレーズ(α=0.10)

[出現回数]	叫喚フレーズ
[312]	うお
[244]	やめろお
[123]	ぎゃあ
[112]	アスカきたあ
[105]	綾波い
[99]	いやあ
[77]	やべえ
[67]	ウオ
[60]	キタア
[59]	カヲルくう
[55]	マミったあ
[54]	やめてえ
[52]	レイい
[50]	もういっちょお
[50]	ねえ
[47]	つえ
[45]	すげえ
[43]	あやなみい
[40]	の時間だあ
[38]	どおりゃあ

表 3　TV 番組実況「天空の城ラピュタ」から抽出した叫喚フレーズ(α=0.13)

[出現回数]	叫喚フレーズ
[264]	うわあ
[151]	あがれえ
[113]	うお
[99]	目が目があ
[97]	しい
[90]	ぱずう
[85]	こえ
[80]	シータあ
[77]	かっけえ
[74]	ムスカあ
[73]	すげえ
[68]	逝ったあ
[66]	しねえ
[65]	シータア
[65]	しーたあ
[59]	いやあ
[54]	つえ
[54]	上がれえ
[50]	オ
[49]	パズウ