原著論文

人名の言語的特徴の分析に基づくキラキラネーム判定

山西 良典, 大泉 順平, 西原 陽子, 福本 淳一

立命館大学

Detection of *KIRAKIRA* Names Based on Linguistic Features of Person Names

Ryosuke YAMANISHI, Junpei OIZUMI, Yoko NISHIHARA and Junichi FUKUMOTO

Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

Abstract : This paper describes linguistic features of generally unreadable person names, which are defined as “*KIRAKIRA* names,” and proposes a method to detect *KIRAKIRA* names based on the features. Through the discussions, the following eight features are founded as the linguistic features of *KIRAKIRA* names: 1) Too many Kanji characters, 2) Too many syllables, 3) Multiple usage of a common Kanji character, 4) Kanji variants are used, 5) The pronunciation of Kanji is generally unknown, 6) Too many stroke count for Kanji, 7) Mismatching of gender between a person and the name, and 8) The pronunciation of name equals an imported word. Based on the features, *KIRAKIRA* names are automatically detected by using Support Vector Machine. The experiments to detect *KIRAKIRA* names were conducted for 10,000 names. The results of the experiments showed 81.79% accuracy, 76.89% precision, and 91.84% recall.

Keywords : *Person name, KIRAKIRA name, Natural language processing*

1. は じ め に

名前は、親が子どもに贈る最初の贈り物である。日本の戸籍法第五十条には「子の名には、常用平易な文字を用いなければならない」とある。日本国に出生届を提出する場合に、名前として使える文字には制限があり、漢字、ひらがな、カタカナと定められている。漢字は常用漢字と人名用漢字の2種類があり、いずれも法務省のWebページにて公開されている[1]。親は、法律上の制約を満たした上で個性や印象、子どもの将来にける想いを込めて、子どもの名前を考える。

名前に用いる文字には制限がある一方で、その読み方や名前自体の文字数には明確な制限は設けられていない。そのため、近年では一目見ただけでは読み方が不明な名前や常識的には考えられないような名前が増加している。このような名前は「キラキラネーム」と呼ばれている。キラキラネームは、2000年前後からテレビ番組や新聞などのメディアで取り上げられるようになった[2]。読めない名前を理由にしたいじめの問題[3]や病院での患者の取り違えの問題[4]などが報告されており、「変わった名前」だけでは済まされない社会的な問題となっている。常識的な感性を考慮せずに名付けられると、名付けられた本人は上記のような不利益を被る可能性が大きくなると予想される。

本稿では、子どもにつけられた名前の中で、常識的には考えられないような名前（つまり、キラキラネーム）の言語的特徴を調査し、人名に抱く感性について「一般的／奇妙」という観点で分析する。そして、分析を通して得られた人名に

抱く感性に影響する言語的特徴に基づいたキラキラネームの自動判定手法を提案する。提案手法により、親は子どもの名前を考える際に自身の感性によって考えた名前の候補を本システムに入力することで、キラキラネームをスクリーニングすることが可能になると考える。

2. 関 連 研 究

名前に関する研究は、主に自然言語処理や教育学の分野で議論されてきている。しかしながら、近年に注目されるようになったキラキラネームに関連した研究は多くはない。

自然言語処理の分野では、人名を扱った研究として人名の曖昧性解消が研究されている[5]。人名の曖昧性解消とは、同姓同名の人物についての情報を分類・整理するための課題である。特にWeb検索において任意の人物について検索をしようとした場合に、名前は同一であっても全く異なる人物の情報が検索結果に無秩序に表示されてしまう問題を解決することが大きな目的となっている。人名の曖昧性解消においては、職業関連情報の有効性の検討[6]や類似文章をマージする半教師有りクラスタリング手法[7]などが報告されている。これらは人名と実在の人物とを正しく対応させることを目的としており、人名の一般性や奇妙さについては扱われていない。

名前の音と感性に関連した研究として、形状と名前の関連性を報告した「プーバ/キキ」効果が知られている[8]。図形の形状に応じて付与される名前には一般性が存在し、母国語によらないことが報告されており、音象徴と形状の関連性が議

論されている。言語情報処理研究としては音象徴に着目したキャラクターへの自動的な名前付けが報告されている [9]。この研究では、「強さ」という抽象的な感性を音象徴によって表現し、機械学習によって名前から「強そう」という感性を抱くキャラクターの名前を自動生成している。この研究ではキャラクターの名前を対象としているため、カタカナのみを文字種として扱っているが、日本人の人名では一般的に漢字が使われることが多い。特に、キラキラネームは漢字に対して奇抜な読みがなが割り当てられていることが多い。

名前や名付けに関する教育学分野での従来研究では、子供の名前に使われる漢字の変化とその誘因を探ったもの [10] や、子供の名付けに見られる文化的な特徴、ジェンダー面での特徴を考察したものがある [11]。これらの研究の中でも、読めない人名について言及がなされている。一方で、「読めない人名」の言語的特徴を調査した研究は少ない。人名の読みに関した言語的特徴の究明は、名付け時に一般的に読めるか否かを評価する際の一助になると考えられる。

3. キラキラネームの言語的特徴の究明

人間は、人名を目にしたとき、直感的にキラキラネームであるかどうかを判断可能である場合が多い。これは、人名の言語的特徴を無意識に評価しているためであると考えた。本節では、キラキラネームの判定に寄与すると考えられる言語的特徴を明らかにするための予備調査の調査条件および調査結果を述べる。キラキラネームは、日本のみならず海外でも報告されている。海外では、その発音が他の単語と合致することでキラキラネームとなる場合が多い。一方で、日本のキラキラネームでは、発音のみならず、漢字の意味や漢字の画数などの多様な要因が考えられる。つまり、日本人の人名に関するキラキラネームの判定は海外の人名のキラキラネーム判定に比べ、より複雑な問題と考えることができる。

3.1 調査条件

調査は、以下の手順に従って行った。

- (1) 調査者が手作業でキラキラネームと思われる人名（キラキラネーム候補）と、一般的であると思われる人名（一般名候補）をそれぞれ200件ずつ用意する。
- (2) 被験者は、調査者が呈示した人名について、キラキラネームであるか一般的な人名であるかという基準で分類する。
- (3) 被験者は、キラキラネームと分類した人名について、キラキラネームであると判断した理由を示す。
- (4) 調査者は、キラキラネームと一般的であると思われる人名の分類結果をまとめる。キラキラネームであると判断した理由をまとめる。

調査手順の(1)では、著者のうち一人が調査者としてキラキラネーム候補をWebサイト「DQNネーム [12]」に掲載されていた人名一覧から200件抽出した。一般名候補は、明治安田生命のWebサイトの「名前ランキング2013」 [13]に掲載されていた人名から200件抽出した。

調査手順の(2)では、被験者はいずれも情報理工学部に所属する大学4年生12名（男性11名、女性1名）であった。各被験者はキラキラネーム候補を100件、一般名候補を100件の合計200件を見て、分類した。1件の人名は合計で6名の被験者により評価された。

調査手順の(3)では、被験者にキラキラネームに分類した名前に対して、分類した理由を記述させた。記述の際、文字数・語数の制限は設けなかった。

調査手順の(4)では、6名中3名以上の被験者がキラキラネームに分類した人名をキラキラネームの正例とし、それ以外の人名を一般的な人名（即ち、負例）とした。キラキラネームに付与された21種類の理由を著者のうち1名の言語処理研究者の言語学の知見を基にまとめ直し、言語的な特徴を表す文言に著者のうち1人が整形した。そして、著者4名全員によって言語的特徴としてモデル化した。キラキラネームに対する感性はファジイなものであるが、本研究ではキラキラネームのスクリーニングという目的を見据え、過半数以上の被験者がキラキラネームであると評価した人名をキラキラネームであると定義した。

3.2 調査結果

表1と表2に、それぞれキラキラネームの正例と負例を示す。キラキラネームに分類された理由に付与された理由としては、「名前の中に多くの漢字が使用されている」「常用しない

表1 キラキラネームの正例の一部

括弧内には読みと性別を示す。

南偉斗(ないと:男),	騎矢(ないと:男),	新萌(にも:男),
日克(にこ:男),	奈依(ない:女),	武雅(ぬが:男),
笑笑(にこにこ:女),	虹恋(にこ:女),	愛葉(にいな:女),
温温(ぬくぬく:女),	生水土(ふゆは:女),	暖(ぬくい:男),
恋桜(れお:男),	勇斗(ふあいと:男),	風生(ふう:男),
恋人(れんじ:女),	玲海(れいあ:女),	晴望(れの:女),
郎蘭(ろーら:女),	舞桜(わるつ:女),	勉次(べんつ:男),
八空雲(やくも:女),	烈楠(れつくす:男),	露音(ろん:男),
藍蒼碧(ななみ:女),	路追人(ろおど:男),	楽(わく:男),
路美於(ろみお:男),	陽音(りずむ:女),	初夏(わか:男),
依真理(よりこ:女),	美広心(よおこ:女),	靖野(やや:女),
都海(みやび:女),	桜夢(みゆ:女)	

表2 キラキラネームの負例（一般的な人名）の一部

括弧内には読みと性別を示す。

隆太(りゅうた:男),	志保(しほ:女),	昇(のぼる:男),
重喜(しげき:男),	麻衣(まい:女),	博(ひろし:男),
達夫(たつお:男),	智也(ともや:男),	豊(ゆたか:男),
博美(ひろみ:女),	成美(なるみ:女),	剛(つよし:男),
友美(ともみ:女),	冬華(ふうか:女),	隆(たかし:男),
章吾(しょうご:男),	奈美(なみ:女),	周(しゅう:男),
順平(じゅんぺい:男),	千佳(ちか:女),	蓮(れん:男),
昌一(しょういち:男)		

人名の言語的特徴の分析に基づくキラキラネーム判定

漢字が使われている」「漢字の読みとして存在しない読みが当てられている」「複雑な漢字が使われている」「男なのに女の子の様な名前」「外人かぶれで恥ずかしい」「読みに伸ばし棒が使われている」「カタカナの当て字」の他、「あだ名みたいで名前らしくない」「名前として不適切」「いじめられそう」「常識的に考えてありえない」などがあった。

3.3 キラキラネームの言語的特徴

3.2節に示した調査結果から、著者のうち1名の言語処理研究者の言語学の知見を基にキラキラネームの分類理由を集約すると、キラキラネームに見られる言語的特徴は表3のようにまとめられる。以下、それぞれの特徴について詳細を説明する。

(1) 漢字の個数

名前の漢字の個数は、キラキラネームに用いられている漢字の個数は、一般的な人名に用いられている漢字の個数よりも多いことで高い値を持つ特徴である。例としては、「華美音(けびん：男)」「牙琥翔(がくと：男)」があった。キラキラネームでは漢字1文字に対して、1音、または1音以下のみの読み方を充てるものが多いことが、この特徴が得られた理由として考えられる。

(2) 読みの発音数

読みの発音数は、キラキラネームに付与された読みの発音数は、一般的な人名よりも多いことによって高い値を持つ特徴である。例としては、「大賀寿(たいがーす：男)」「大勇明育(おさむのすけ：男)」があった。この特徴をもつ人名がキラキラネームとなる理由は、特徴(1)と同様で、キラキラネームでは漢字1文字に対して、読みを1音、または1音以下充てるものが多いことと考えられる。

(3) 同一漢字の複数回使用

同一の漢字の複数回使用とは、キラキラネームに同一の漢字が複数回用いられていることが多いことを示す特徴である。例としては、「笑笑(にこにこ：女)」「七七七(ななみ：女)」があった。

(4) 異体漢字の使用

キラキラネームには異体漢字が含まれることが多い。異体字とは、常用漢字と同様に使用可能であり読むことは可能であるが、形が異なっている漢字を指す。例としては、

「愛來(あいら：女)」の「來」、「愛凜(あめり：女)」の「凜」が異体漢字にあたる。

(5) 漢字の音訓読みでない読み方の使用

キラキラネームの読みは漢字の音訓読みに一致しないものが多い。例としては、「弓(あむろ：女)」の「弓」1文字に対しての「あむろ」や「希音(ねおん：女)」の「希」に「ね」という音を充てたものなどがあった。キラキラネームにおいて通常の漢字の音訓読みでない読み方が使用される理由として、漢字の読みではなく漢字から連想される音を読みとして与えることが多いと考えられる。

(6) 漢字の総画数

キラキラネームは、一般的な人名よりも漢字の総画数が多い傾向にある。キラキラネームでは、上述の特徴(1)で示したように漢字の個数が多かったり、特徴(4)に示したように異体字の使用が多かったりすることによって、漢字の総画数も多くなる傾向があると考えられる。

(7) 漢字と性別の不一致

キラキラネームの中には男性に与えた人名であるにも関わらず、女性の名前として多用される漢字が使われていることがある。これを、漢字と性別の不一致として特徴量化する。例えば、「真里菜(まりな：男)」や「麟平(りんぺい：女)」などの人名がこの特徴を有する人名として挙げられる。

(8) 読み方がカタカナ単語として存在

キラキラネームの読み方のなかには、それ自体がカタカナの単語として存在することが多い。例えば、「夢泡(むーあ：女)」は英単語「moor」に相当し、「寿江琉(じゅえる：女)」は英単語「jewel」に相当する。

本稿では、調査結果から得られたこれら8つの特徴をキラキラネームの言語的特徴と定義する。そして、これらの特徴を用いることで、キラキラネームであるか一般的な人名であるかを判別する手法を提案する。

4. キラキラネーム判定手法

本稿では、キラキラネームの判定において、以下の3つ組の情報を人名情報として用意するものとする。

- (1) 人名の漢字
- (2) 人名の読み方
- (3) 名前が付与される人物の性別

これら3つ組の情報から、表3に示した8つ組の特徴量を要素にもつ特徴ベクトルを生成する。そして、それらの特徴ベクトルを基に、Support Vector Machine (以下、SVM) [14]によって、入力された人名がキラキラネームであるかを判別するパターン認識モデルを学習し、新たに入力された人名がキラキラネームであるか否かを判別する。

4.1 人名の言語的特徴ベクトル

まず、人名 n についての人名情報について、(1) 人名の漢字を c 、(2) 人名の読み方を p 、(3) 名前が付与される人物の性別を s とする。表3に示した8つ組の特徴量を要素にもつ、

表3 キラキラネームの言語的特徴

特徴の説明
(1) 漢字の個数
(2) 読みの発音数
(3) 同一漢字の複数回使用
(4) 異体漢字の使用
(5) 漢字の音訓読みでない読み方の使用
(6) 漢字の総画数
(7) 漢字と性別の不一致
(8) 読み方がカタカナ単語として存在

人名の言語的特徴ベクトルは、下式で表現される。

$$n(c, p, s) = (CNumber_c, PNumber_p, Same_c, Variant_c, \\ Pronounce_p, Stroke_c, Unmatch_{c,s}, Imported_p) \quad (1)$$

右辺の各要素の説明とその算出方法は、以下で詳細に述べる。

$CNumber_c$ は、人名に含まれる漢字の個数を表す要素であり、人名に含まれる漢字の個数を値に持つ。例えば、名前 n の漢字 c が「太郎」であれば、 $CNumber_c = 2$ となる。

$PNumber_p$ は、人名の読み方の発音数を表す要素であり、人名の発音数を値に持つ。例えば、名前 n の読み p が「たろう」であれば、 $PNumber_p = 3$ となる。

$Same_c$ は、人名の漢字中で同一漢字が複数回用いられているかどうかを表す要素であり、複数回用いられている場合には1、用いられていない場合には0を値に持つ。例えば、名前 n の漢字 c が「笑笑」であれば、 $Same_c = 1$ となり、「太郎」であれば、 $Same_c = 0$ となる。

$Variant_c$ は、人名に異体漢字が使用されているか否かを示す要素である。異体字が用いられていれば1、用いられていなければ0を値に持つ。例えば、名前 n の漢字 c が「亜美」であれば、「亜」が異体字であるため $Variant_c = 1$ となり、「太郎」であれば、異体字は使用されていないため $Variant_c = 0$ となる。このとき、異体字の判定には、異体字を収集したデータベースを用いる。このデータベースは著者のうち1名がWebページ¹を参照して作成した。データベースには1,332個の異体字を収録している。

$Pronounce_p$ は、漢字の音訓読みでない読み方の使用を示す要素である。漢字に対して一般的に充てられている音訓読みと、人名に当てられた読みが一致しなければ1、一致すれば0を値に持つ。音訓読みのありなしの判定には、漢字の音訓読みを集めたデータベースを用いる。法務省が子の名に使用できる漢字として公開している常用漢字表と人名用漢字表を元にしてテキストを作成した[1]。データベース内には、13,389個の漢字と、40,715個の音訓読みがある。1つの漢字に対して平均3.04個の音訓読みが存在する。漢字の音訓読みと人名の読み方の一致では、完全一致と部分一致のいずれでも一致しているものとした。部分一致の判定では、まず名前 n の漢字 c を、 $c_0, c_1, \dots, c_{CNumber_c}$ と漢字の個数だけに分ける。続いて、名前 n の読み p を $p_0, p_1, \dots, p_{PNumber_p}$ と読み方の発音数に分割する。その後、漢字の音訓読みのテキストを参照し、図1に示したアルゴリズムを用いて部分一致を判定する。

$Stroke_c$ は、人名に使用された漢字の総画数を表す要素であり、漢字の総画数を値に持つ。漢字の総画数の値を算出するためには、総画数を集めたデータベースを用いる。データベース内には13,389個の漢字についての画数が収録されている。画数の収集には漢字の画数をまとめたWebページ²を参照した。

1. $i=0$ とする。 $j=0$ とする。 $start=0$ とする。
2. 漢字 c_i の読み方に部分一致する最大の系列 $(p_{start}, \dots, p_j)$ を探すまで $j=j+1$ とする。
3. $i == CNumber_c$ かつ $j == PNumber_p$ であれば、部分一致と判定して、終了する。 $i == CNumber_c$ または $j == PNumber_p$ であれば、部分一致しないと判定して、終了する。それ以外の場合は、手順4.へ行く。
4. $start=j+1$ とする。 $i=i+1$ とする。 $j=j+1$ とする。2.へ戻る。

図1 漢字の音訓読みと人名の読み方の部分一致判定アルゴリズム

$Unmatch_{c,s}$ は、人名の漢字と性別の不一致を示す要素であり、人名の漢字 c が頻繁に用いられる性別と、名前が付与される人物の性別 s が一致していなければ1、一致していれば0を値に持つ。人名の漢字と性別の不一致を判定する際には、性別ごとに頻繁に使用される漢字を集めたデータベースを用いる。明治安田生命の「名前ランキング2013」に掲載されていた男性の名前の漢字から、男性の人名中でのみ2回以上使われていた漢字を男性の人名として頻繁に用いられる漢字とし、データベースに収録した。女性の名前として頻繁に使用される漢字も同様の手順で用意し、データベースに収録した。データベース内には、男性の名前として頻繁に使用される漢字が73個、女性の名前として頻繁に使用される漢字が62個ある。

$Imported_p$ は、人名の読み方がカタカナ単語として存在するか否かを表す要素であり、読み方がカタカナ単語として存在すれば1、存在しなければ0を値に持つ。人名の読み方がカタカナ単語として存在するかの判定には、カタカナ単語を集めたデータベースを用いた。このデータベースでは、IPA辞書の中でカタカナとして登録されている名詞を抽出し、その原型を収録した。カタカナ単語は9,920個ある。

4.2 人名の言語的特徴ベクトルの学習

4.1節で述べた人名の言語的特徴ベクトルを学習するためにはSVMを用いる。このとき、SVMの学習には、SVM-Light [15]を用いた。学習時には、キラキラネームを正例、それ以外の一般的な人名と判断された人名をキラキラネームの負例として学習する。

本稿で用意した人名の言語的特徴の中の一部の特徴間には、従属関係があると考えられる。このため、一般的な多変量解析では、多重共線性による過学習が発生するおそれがある。本稿では、カーネル法に基づいたSVMを用いることで、従属関係をもつ特徴量群が引き起こす可能性のある多重共線性を回避したパターン学習を実現する。

5. キラキラネーム判定実験

人名の言語的特徴ベクトルを用いたキラキラネーム判定手法について、その精度を評価するための実験を行った。

¹ http://www.wap.hi.u-tokyo.ac.jp/ships/itaiji_list.jsp など

² <http://rtk.art.coocan.jp/cjk/stks/> など

人名の言語的特徴の分析に基づくキラキラネーム判定

5.1 実験条件

実験に用いた人名には、子どもにつけられた名前が掲載されたWebページ[12, 13]から取得したものをを用いた。取得した人名を被験者の主観評価実験によってキラキラネームと一般的な名前に分類した。

被験者は16名で、いずれも情報理工学部に所属する大学4年生であり、男性15名、女性1名であった。1件の人名について、2名が評価するように名前を割り当てた。評価者2名のうち、2名共がキラキラネームと評価したものを、キラキラネームの正例、それ以外を一般的な人名、すなわちキラキラネームの負例とした。この主観評価実験の結果、男性と女性のキラキラネームそれぞれについて2,500件ずつ、男性と女性の一般的な人名がそれぞれ2,500件ずつ集まり、合計で10,000件の名前を評価実験のために用意した。

用意した10,000件の名前を10個に分割し、10分割交差検定を行った。正例を正例と分類する精度、および名前を分類する適合率と再現率を算出し、提案手法の評価に用いた。

5.2 実験結果と考察

表4に、10分割交差検定の結果得られた精度、適合率、再現率を示す。ここで、精度はキラキラネームと一般的な人名をそれぞれ正しく判定していた割合を示す。適合率は提案手法がキラキラネームと判定した人名のうち正しくキラキラネームであった割合を示し、再現率は全てのキラキラネームのうち提案手法によって正しくキラキラネームと判定した割合を示す。同表より、提案手法によるキラキラネーム判定では、平均精度81.8%、平均適合率76.9%、平均再現率91.8%が確認された。

正しく判定することができた人名の例を表5に示す。性別に関係なく、キラキラネームと一般的な人名とがうまく判定されている様子が見て取れる。特に、再現率は90%を越えており、おおよそのキラキラネームは本稿で提案した8つの言語的特徴量によって判別可能であると考えられる。

次に、提案手法によってキラキラネームであるにも関わらず一般的な人名と誤判定した例を表6に示す。これらの名前

をキラキラネームと判定するためには、本稿で用意した8種類の特徴量以外の特徴を捉える必要があると考えられる。例えば、「食べ物の名前」として存在するか否かを評価する特徴量を追加することで、「燃志(もやし:男)」「芽論(めろん:女)」「衣智瑚(いちご:女)」などは、正しくキラキラネームと判定可能になると考える。また、「我宝(わほう:男)」「我馬衣(がばい:男)」「鳳正(おうが:男)」「季夢汰(きむた:男)」などは、日本人の人名の音として聞き慣れないものであると考えられる。一般的な人名に多く存在する音素列であるかを評価する特徴量を用いることで、これらのキラキラネームは判定可能になると考える。外国人の名前として存在する読み方を評価することによって、「麻囚(ましゆ:男)」「落敏(ろびん:男)」「維吏弥(いりや:男)」「音呂(ねろ:男)」などを判定可能になるとも考えられる。しかしながら、グローバル化が進む現代においては外国人の人名の読み方の日本人名が定着していく可能性も考えられ、これらについては更なる議論が必要とされるであろう。この問題の解決法としては、文献[13]などを参照にして年代ごとの人名を学習し、年代ごとのキラキラネーム判定モデルを学習するアプローチやreCAPTCHA[16]のような、データを半自動的に収集可能なフレームワークを応用することでキラキラネームに対するダイナミカルな感性を捉える仕組みが考えられる。

表5 提案手法によって正しく判定できたキラキラネームと一般的な人名の例

括弧内には読みと性別を示す。

キラキラネーム		一般的な人名	
未来蘭	(あすらん:男),	快斗	(かいと:男),
織夢	(おりむ:男),	亮広	(あきひろ:男),
麗愛	(うるあ:男),	寛大	(かんた:男),
鳳凰翔	(おうが:男),	敦司	(あつし:男),
輪変統	(つづく:男),	歩	(あゆむ:男),
彩愛紗	(さあしや:女),	一花	(いちか:女),
業愛華	(なうしか:女),	久実	(くみ:女),
咲笑穂	(さくらんぼ:女),	里香	(りか:女),
子子	(にこ:女),	早梨	(さり:女),
七七七	(ななみ:女)	奈美	(なみ:女)

表4 提案手法の10分割交差検定の結果得られた精度、適合率、再現率

数値は%.

分割番号	精度	適合率	再現率
1	81.8	76.8	91.2
2	92.1	88.9	96.2
3	81.1	75.8	91.4
4	79.7	74.2	91.0
5	86.9	84.5	90.4
6	81.4	75.2	93.6
7	78.9	72.3	93.8
8	82.4	78.1	90.0
9	79.3	75.9	85.8
10	74.3	67.2	95.0
平均	81.8	76.9	91.8

表6 提案手法によって一般的な人名であると誤判定したキラキラネームの例

括弧内には読みと性別を示す。

無	(む:男),	莓歩	(いちほ:女),
我宝	(わほう:男),	笑亜	(えむあ:女),
麻囚	(ましゆ:男),	莓朗	(いちごろう:男),
燃志	(もやし:男),	祈履	(いのり:女),
芽論	(めろん:女),	鳳正	(おうが:男),
落敏	(ろびん:男),	音呂	(ねろ:男),
我馬衣	(がばい:男),	季夢汰	(きむた:男),
衣智瑚	(いちご:女),	亜美瑠	(あみる:女),
維吏弥	(いりや:男),	獲	(える:女),
衣舞姫	(いぶき:女),	亜里亜	(ありあ:女)

表7 提案手法によってキラキラネームであると誤判定した一般的な人名の例

括弧内には読みと性別を示す。			
愛結美	(あゆみ：女)，	咲緒梨	(さゆり：女)，
織登	(おりと：男)，	沙織子	(さおりこ：女)，
織英	(おりひで：男)，		

一般的な人名であるにも関わらず、キラキラネームであると誤判定された人名の例を表7に示す。これらの名前は、常識的に読むことが可能であるが、キラキラネームであると誤判定された。これらの名前は全て漢字の総画数が多い。提案手法では、名前に使用された漢字の総画数が多い場合にキラキラネームであると判定するモデルを構築したため、このように誤判定されたと考えられる。

6. お わ り に

本稿では、子どもにつけられた名前の中で、常識的には考えられないような名前をキラキラネームと定義し、キラキラネームの言語的特徴について予備的な調査を行った結果を報告した。そして、調査の結果明らかとなったキラキラネームの言語的特徴を用いたキラキラネーム判定手法を提案した。

キラキラネームの言語的特徴に関する予備的な調査では、キラキラネームの言語的特徴として8つの特徴が得られた。それぞれ、(1) 漢字の個数、(2) 読みの発音数、(3) 同一漢字の複数回使用、(4) 異体漢字の使用、(5) 漢字の音訓読みでない読み方の使用、(6) 漢字の総画数、(7) 漢字と性別の不一致、(8) 読み方がカタカナ単語として存在、であった。

調査の結果得られたキラキラネームの言語的特徴に基づいて、人名を特徴ベクトルとして表現し、SVMを用いて学習した。このキラキラネームの学習モデルに基づくキラキラネーム判定手法を提案した。10,000件の人名に対して評価実験を行った所、キラキラネーム判定の性能は、平均精度81.79%、平均適合率76.89%、平均再現率は91.84%であった。これらの性能評価実験の結果より、本稿の調査によって得られたキラキラネームの言語的特徴は、ヒトの人名に対する「一般的／奇妙」という直感的な感性をうまく捉えられていたことが確認された。提案手法を用いることで、客観的な視点で人名の「一般性／奇妙さ」を捉えることが可能となり、子どもの名付けにおける参照情報となることが期待される。

今後は、5.2節にて考察した、キラキラネームであるにも関わらず一般的な人名と誤判定した人名の特徴を追加した人名の言語的特徴モデルを提案し、キラキラネーム抽出性能の向上を目指す。一方で、本モデルでは、名乗り読みを考慮していないため、一般的な人名であるにも関わらず、「名乗り読み」を用いた人名については、キラキラネームとして誤判定してしまう問題も存在する。例えば、本稿の第1著者および第2著者の名前は、それぞれ「典」を「すけ」、「平」を「ぺい」と読む。

このような人名は、本稿のモデルではキラキラネームとして誤判定してしまう。名乗り読みのデータを収集し、適用することでこの問題の解決を図る。キラキラネームには、姓との組み合わせによってキラキラネームとなる一般的な人名も存在する。例えば、「金田一」(きんだいち はじめ：男)や「糸色望」(いとしき のぞむ：男)などがこれにあたる。この問題の解決には、名前のみならず、姓名を一貫して評価する言語的特徴のモデル化が必要となると考える。また、Web上で提案手法を実装したキラキラネーム判定システムを公開し、広く多くのユーザに使用してもらい、フィードバックを得ることで、システムの有用性の検証およびフィードバックに基づく手法の改善を目指していく。

謝 辞

本研究は一部、JSPS科研費26540047の助成のもと行われた。記して謝意を示す。

参 考 文 献

- [1] 法務省：戸籍—子の名に使える漢字。
<http://www.moj.go.jp/MINJI/minji86.html>
- [2] 小林康正：名づけの世相史「個性的な名前」をフィールドワーク、京都文教大学 文化人類学ブックレットNo.4, 2009.
- [3] J-CASTニュース：安倍総裁「キラキラネーム、いじめられる」北海道新聞「いじめ、いさめるのが教育」。
<http://www.j-cast.com/2012/11/23155095.html?p=all>
(2015年4月15日確認)
- [4] J-CASTニュース：「キラキラネーム」で患者取り違えの恐れ？救命医のツイッター苦言に賛否両論。
<http://www.j-cast.com/2013/08/01180677.html?p=all>
(2015年4月15日確認)
- [5] 関根聡：Web検索における人名の曖昧性解消技術の動向：同姓同名のクラスタリング、情報処理, 49(5), pp.573-578, 2008.
- [6] 上田洋, 村上晴美, 辰巳昭治：Web上の同姓同名人物識別のための職業関連情報の抽出、システム制御情報学会論文誌, 22(6), pp.229-240, 2009.
- [7] 杉山一成, 奥村学：半教師有りクラスタリングを用いたWeb検索結果における人名の曖昧性解消、自然言語処理, 16(5), pp.23-49, 2009.
- [8] Wolfgang Köhler: Gestalt Psychology, 2nd Ed. New York, Liveright, 1929.
- [9] 三浦智, 村田真樹, 保田祥, 宮部真衣, 荒牧英治：音象徴の機械学習による再現：最強のポケモンの生成、言語処理学会第18回年次大会発表論文集, pp.65-68, 2012.
- [10] 半田淳子：子どもの名づけに見られる日本文化の特徴、第九回国際日本語教育・日本研究シンポジウム予稿集, 2012.

人名の言語的特徴の分析に基づくキラキラネーム判定

- [11] 川岸克己：人名における漢字使用の変化とその誘因，安田女子大学紀要，41，pp.1-14，2013.
- [12] DQN ネーム (子供の名前@あー勘違い・子供がカワイソ)，
view-source : <http://dqname.jp/> (2015年4月15日確認)
- [13] 明治安田生命：名前ランキング2014，
<http://www.meijiyasuda.co.jp/enjoy/ranking/index.html>
(2015年4月15日確認)
- [14] C. Cortes, V. Vapnik: Support-Vector Networks, Machine Learning, 20(3), pp.273-297, 1995.
- [15] T. Joachims: SVM-Light, <http://svmlight.joachims.org/>
(2015年4月15日確認)
- [16] L.v. Ahn, B. Maurer, C. McMillen, D. Abraham, M. Blum: reCAPTCHA: Human-Based Character Recognition via Web Security Measures, Science, 321(5895), pp.1465-1468, 2008.

山西 良典 (正会員)

2007年 名古屋工業大学工学部知能情報システム学科卒業。2009年 同大学院工学研究科情報工学専攻博士前期課程修了。2012年 同研究科同専攻博士後期課程修了。同年、立命館大学情報理工学部助手。2013年同特任助教。2014年同助教、現在に至る。博士(工学)。感性情報処理、音楽情報処理、ヒューマンインタラクションなどに興味を持つ。情報処理学会、人工知能学会、知能情報ファジイ学会各会員。

大泉 順平 (非会員)

2015年 立命館大学情報理工学部卒業。学士(工学)。

西原 陽子 (非会員)

2003年 大阪大学基礎工学部卒業。2005年 同大学院基礎工学研究科博士前期課程修了。2007年 同研究科博士後期課程修了。博士(工学)。日本学術振興会特別研究員を経て、2008年 東京大学大学院工学系研究科助教、2009年 同講師、2012年 立命館大学情報理工学部准教授、現在に至る。コミュニケーション支援、対話デザイン研究に興味を持つ。

福本 淳一 (非会員)

1984年 広島大学工学部第2類卒業。1986年 同大学院工学研究科システム工学専攻博士前期課程修了。同年、沖電気工業(株)入社。1992-1994年 英国マンチェスター科学技術大学 Ph.D. コース在学。2000年 立命館大学理工学部情報学科助教授。2004年 同大学情報理工学部メディア情報学科助教授。2006年 同教授、現在に至る。Ph.D. 談話構造解析、質問応答技術、情報抽出、比喩表現理解などに興味を持つ。情報処理学会、電子情報通信学会、人工知能学会、言語処理学会、観光情報学会各会員。