

令和 2 年 5 月 20 日現在

機関番号：10101

研究種目：基盤研究(C)（一般）

研究期間：2017～2019

課題番号：17K00294

研究課題名（和文）ユーモアの面白さの評価手法及び標準データセットの開発

研究課題名（英文）Development of Evaluation Method and Standard Data Set for Fun of Humor

研究代表者

荒木 健治（ARAKI, KENJI）

北海道大学・情報科学研究院・教授

研究者番号：50202742

交付決定額（研究期間全体）：（直接経費） 2,600,000円

研究成果の概要（和文）：新型コロナによる外出自粛により精神面での健康の維持が社会的な問題となっている。本研究課題では、ユーモアの面白さのメカニズムを科学的に解明し、工学的に実現することにより人間と話すのと同等あるいはそれ以上に楽しいユーモアを認識・生成するための基盤となるユーモアデータベースを構築し、駄洒落の分析・認識・自動生成、対話システムによるユーモアによる対話破綻の回避、オノマトペの分析、認知科学的面白さのメカニズムの分析、皮肉の検出など種々のユーモアに関する研究を行った。また、研究成果は67,000件を面白さの度合いなどの情報を持つ駄洒落データベースとしてHP上で一般に公開し、広く利用していただいている。

研究成果の学術的意義や社会的意義

本研究の学術的意義としては、これまでユーモアの研究については、大規模なデータベースが存在せず、各研究者が独自に収集した小規模なデータを用いて評価実験を行ってきた。このような現状に対して、本研究課題では面白さのスコア、駄洒落の分類、構造などの種々の情報を付与した67,000件の駄洒落データベースを構築し、公開することによりユーモア研究の研究基盤を確立したことにある。また、社会的意義としては、新型コロナによる外出自粛などにより最近顕著な問題となっている精神的な健康を維持するために、人間を超える面白いユーモアを生成する技術の開発を促進することに貢献することである。

研究成果の概要（英文）：Maintaining mental health has become a social issue due to the refraining from going out with the new Corona. In this research project, a humor database, which is the basis for recognizing and generating humor that is as fun as or more than talking to humans by scientifically elucidating the mechanism of the fun of humor and realizing it through engineering, is provided. We constructed and analyzed various humors such as pun analysis, recognition and automatic generation, avoiding dialogue breakdown due to humor by dialogue system, onomatopoeia analysis, analysis of mechanism of cognitive science fun, and detection of irony. In addition, we have been released to the public on HP as the pun database (67,000 puns) with information such as the degree of interest, and are widely used.

研究分野：自然言語処理

キーワード：ユーモア 駄洒落 面白さ オノマトペ 皮肉

1. 研究開始当初の背景

日本社会は、少子高齢化と独身若年層の急増により今後これまで人類が経験したことのないような高齢者単身世帯が急増する事態に直面する。高齢者単身世帯で重要となるのが、高齢者の孤独感を緩和することである。すなわち、心のケアが最重要課題となる。心のケアのためには高齢者単身世帯を訪問する人を多く雇用する必要がある。しかし、このためには膨大な経費が必要となり、昨今の医療、福祉予算の急増と政府の緊縮財政を考えると現実的ではない。

そこで、我々はこの問題を解決するためにこれまで非タスク指向型の対話システム(雑談システム)の研究を行ってきた。しかし、応答精度の低さや会話が盛り上がらないという点等依然として多くの問題が残されていた。

これらの問題を解決するために、対話システムのさらなる精度の向上がもちろんであるが、より人間らしい振る舞いのできる対話システムを開発することが必須である。人間の対話と対話システムの対話との違いは種々あるが、ユーモアを認識したり、生成したりすることが人間並にできないことが大きな要因の一つである。また、ユーモアにより発する笑いは医療的な効果が科学的に実証されている。しかし、ユーモアを自動的に認識・生成するという工学的な面での研究は少なく、学問として確立されているとは言い難かった。特に、その原因の大きな要因は、ユーモアの面白さを評価するための標準的なデータセットも手法も確立されていないということであった。

2. 研究の目的

本研究の目的は、ユーモア研究における標準的なデータベース及び評価手法を確立することにより、ユーモア研究の基盤を整備することである。このことにより、ユーモア研究を行う際に問題となる、ユーモアの面白さに対する評価手法がないということや、評価を行うための標準的なデータセットがないという問題の解決を目指す。これまでは、ユーモア研究における実験は、その都度複数の被験者が主観的な評価を行っているが、評価者は少数であり、他研究との比較も難しい。そこで、本研究課題では、ユーモア全般の面白さに関する評価手法を確立し、実験の際に用いる標準的なデータセットを定め、ユーモア工学確立のための基盤を整備することを目的としている。

3. 研究の方法

研究の方法としては、ユーモアの面白さの度合いを決定するためのデータベースを開発することである。67,000件という大規模な駄洒落データベースを開発し、種表現、変形日表現、それらの対応関係、駄洒落の分類、面白さのスコアを人手により付与することにより、開発を行った。

この駄洒落データベースは HP 上で広く一般に公開され、ユーモア研究に利用されているが、我々も駄洒落の分析・認識・自動生成、駄洒落中のオノマトペの分析、認知科学的見地からの面白さの分析、駄洒落の面白さの予測など種々の研究を行っている。また、駄洒落以外のユーモアである皮肉についても研究を進め、皮肉の検出などの研究を行った。

4. 研究成果

(1) 2017年度は、ユーモアを収集し、ユーモアの面白さの評価手法を検討し、その開発を行う計画であった。そこでその第一段階としてインターネット上の駄洒落を収集した9つのサイトから駄洒落を大量に自動的に収集し、駄洒落データベースの構築を行った。その成果を2017年12月に開催された第56回ことば工学研究会において発表を行った。駄洒落データベースは51,000件を収録し、ツールを用いて形態素解析を行ったものに人手で種表現、変形表現、駄洒落の種類などのタグ付けを行ったものである。

さらに面白さの評価手法の確立としては、駄洒落データベースを用いて認知科学的見地から駄洒落の面白さの要因の分析を行った。この結果について前述の研究会において発表を行った。また、駄洒落データベースに含まれるオノマトペの分析を行い、その成果を2018年3月に開催された言語処理学会年次大会で研究発表を行った。さらに、駄洒落データベースなどを用いて駄洒落生成システムの開発を行い、その性能評価結果を2018年2月に開催された第57回ことば工学研究会において発表を行った。

このように2017年度は、代表的なユーモアである駄洒落について、大規模な駄洒落データベースの構築を行い、開発された駄洒落データベースを用いて、面白さの要因の分析、駄洒落に含まれるオノマトペの分析、駄洒落生成システムの開発とその性能評価を行った。また、駄洒落以外のユーモアとしてストーリージョーク、皮肉などを収集し、面白さの要因分析を行った。

当初の計画であったユーモアの収集によるデータベースの構築については、当初の予定をほぼ遂行でき、面白さの評価手法の検討については、開発された駄洒落データベースを用いて、認知科学的な観点からの面白さの要因分析、駄洒落に含まれるオノマトペの分析、駄洒落生成システムの開発を行ったので、これについても当初の予定をほぼ遂行できたと考えられる。

(2) 2018年度は、2017年度に開発した駄洒落データベースをさらに拡張し、分析を行い、その拡張された駄洒落データベースを用いて面白さの要因の分析をさらに進める計画であった。そこで2017年度にインターネット上の駄洒落を収集した9つのサイトから、その後追加された駄洒落を大量に自動的に収集し、駄洒落データベースの拡張・分析を行った。その成果を2018年9月に開催された第58回ことば工学研究会において発表を行った。

さらにこの拡張された駄洒落データベースを用いて認知科学的見地から駄洒落の面白さの要因の分析を行った。この結果について前述の研究会において発表を行った。また、駄洒落データベースに含まれるオノマトペの分析をさらに進め、その成果を知能と情報(日本知能情報ファジィ学会誌)において発表した。また、さらに駄洒落を収集するためにニコニコ動画のコメントに含まれる駄洒落を自動的に収集するシステムを開発し、その検出精度を考察した。この成果は、情報処理学会第81回全国大会において発表を行った。また、皮肉検出における感情生起要因の有効性の検討および駄洒落の自動生成を雑談対話システムにおける対話破綻回避に使用しその有効性を確認した。これらの成果は、人工知能学会第9回対話システムシンポジウム研究会および第17回情報技術フォーラム(FIT2018)において発表を行った。

当初の計画であったさらなるユーモアの収集によるデータベースの拡張については、当初の予定をほぼ遂行でき、面白さの評価手法の検討については、拡張された駄洒落データベースを用いて、認知科学的な観点から面白さの要因分析を行ったので、当初の予定をほぼ遂行でき、さらに、駄洒落に含まれるオノマトペの分析、動画コメントからの駄洒落の自動抽出、皮肉の自動検出、対話システムにおける駄洒落を用いた対話破綻回避を行ったので、これらについても当初の予定をほぼ遂行できたと考えられる。

(3) 2019年度は、本研究課題の最終年度であることから2017,2018年度に開発した駄洒落データベースに面白さのスコアを付与し、面白さを付与された駄洒落データベースを用いて、面白さの評価の傾向及び面白さの要因の分析をさらに進める計画であった。そこで2017,2018年度に開発した67,000件を有する駄洒落データベースにクラウドソーシングで雇用した3名の評価者により5段階の面白さのスコア付けを行い、面白さのスコアの平均値の推移、評価者ごとのスコアの分布、評価者間の一致度、平均スコアの上位のものなどの分析を行った。この結果、駄洒落の面白さの感じ方には個人差が大きいことや対話形式で表現されている場合に面白さを感じるなどことが明らかとなった。これらのことより文脈情報、背景知識が駄洒落の面白さの感じ方の大きな要因となっていることが確認された。また、この駄洒落データベースは、すでに<http://arakilab.media.eng.hokudai.ac.jp/~araki/dajare.htm>において公開し、広く一般に利用されている。また、この駄洒落データベースに駄洒落検出用のフレームワークを開発しGitLabリポジトリにおいて公開を行った。また、ニコニコ動画のコメントデータから駄洒落を自動的に抽出し、抽出された駄洒落の分析を行った。その結果、種表現を明示的に持たない重畳型駄洒落が4割を占めることが明らかになった。また、ユーモアの種類としての皮肉についても人手による皮肉の核のタグ付けを行い、その特徴を分析した。また、その結果を用いて、皮肉の核の自動抽出手法を構築した。これらの成果は、第63回ことば工学研究会において発表を行った。

さらにこの面白さを付与された駄洒落データベースを用いて深層学習を使ってAIに駄洒落の面白さを区別させようとするとき、テキストに加え、どのような教師データがあれば、AIに区別かが可能になるかを考察した。この成果は、第64回ことば工学研究会で発表を行った。

また、面白さを付与した駄洒落データベースを用いてオノマトペに着目した面白さの分析を行った。具体的には、駄洒落に出現するオノマトペの出現傾向を定量的に示し、オノマトペの出現位置や音の変化について考察を行った。その結果、独特な音韻的特徴や音の変化への柔軟性、新語の作りやすさなど、オノマトペの特徴が駄洒落の成立に貢献していることが明らかになった。その成果を知能と情報(日本知能情報ファジィ学会誌) Vol.32, No.1及び第35回ファジィシステムシンポジウム講演論文集(FSS2019)において発表を行った。また、面白さを付与された駄洒落データベースを学習データとしてSVMを用いて駄洒落の面白さの予測を行った。この成果は、令和元年度電気・情報関係学会北海道支部連合大会において発表され、2019 IEEE Sapporo Section Student Paper Contest Encouraging Prizeを受賞した。

本研究課題の目的である駄洒落データベースに面白さを付与し、その駄洒落データベースを用いて、面白さの要因を分析し、面白さのメカニズムを解明するということは、ほぼ遂行できたと考えられる。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 内田ゆず	4. 巻 30,4
2. 論文標題 計算機で言葉の意味を扱う難しさーオノマトペのあいまいさを例にー	5. 発行年 2018年
3. 雑誌名 知能と情報（日本知能情報ファジィ学会誌）	6. 最初と最後の頁 28-30
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.3156/jssoft.32.1_507	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 内田ゆず, 荒木健治	4. 巻 32,1
2. 論文標題 駄洒落に使用されるオノマトペの特徴分析	5. 発行年 2020年
3. 雑誌名 知能と情報（日本知能情報ファジィ学会誌）	6. 最初と最後の頁 507-511
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.3156/jssoft.32.1_507	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計16件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 谷津元樹, 荒木健治
2. 発表標題 動画コメントデータ及びブログ記事における駄洒落の抽出
3. 学会等名 情報処理学会第 81 回全国大会講演論文集, 23-24
4. 発表年 2019年

1. 発表者名 魚住ゆい, 内田ゆず, 荒木健治
2. 発表標題 皮肉検出における感情生起要因の有効性
3. 学会等名 第 17 回情報技術フォーラム(FIT2018)講演論文集, 163-164
4. 発表年 2018年

1. 発表者名 徐雲帆, 荒木健治
2. 発表標題 駄洒落を用いた雑談対話システムにおける対話破綻回避の有効性
3. 学会等名 第17回情報技術フォーラム(FIT2018)講演論文集, 161-162
4. 発表年 2018年

1. 発表者名 徐云帆, 荒木健治
2. 発表標題 雑談対話システムにおけるLSTMを用いた駄洒落による対話破綻回避の有効性
3. 学会等名 第9回対話システムシンポジウム研究会資料, 143-148
4. 発表年 2018年

1. 発表者名 佐山公一, 荒木健治
2. 発表標題 駄洒落理解過程のモデル化 -駄洒落の事例を手続きで分類する-
3. 学会等名 第58回ことば工学研究会, 17-27
4. 発表年 2018年

1. 発表者名 荒木健治, 佐山公一, 内田ゆず, 谷津元樹
2. 発表標題 駄洒落データベースの拡張及び分析
3. 学会等名 第58回ことば工学研究会, 1-15
4. 発表年 2018年

1. 発表者名 荒木健治, 内田ゆず, 佐山公一, 谷津元樹
2. 発表標題 駄洒落データベースの構築及び分析
3. 学会等名 第56回ことば工学研究会資料, 13-24
4. 発表年 2017年

1. 発表者名 佐山公一, 荒木健治
2. 発表標題 コンピュータが駄洒落で笑わせる? 駄洒落の面白さを認識する理解の手続きー
3. 学会等名 第56回ことば工学研究会資料, 25-32
4. 発表年 2017年

1. 発表者名 荒木健治
2. 発表標題 駄洒落データベースを用いた駄洒落生成システムの性能評価
3. 学会等名 第57回ことば工学研究会資料, 39-48
4. 発表年 2018年

1. 発表者名 内田ゆず, 荒木健治
2. 発表標題 駄洒落に含まれるオノマトペの特徴分析
3. 学会等名 言語処理学会第23回年次大会発表論文集, 741-744
4. 発表年 2018年

1. 発表者名 魚住惟, 内田ゆず, 荒木健治
2. 発表標題 皮肉の核の分析および自動抽出手法
3. 学会等名 第63回ことば工学研究会, 17-23
4. 発表年 2019年

1. 発表者名 荒木健治, 内田ゆず, 佐山公一, 谷津元樹
2. 発表標題 駄洒落データベースを用いた面白さの評価及び分析
3. 学会等名 第63回ことば工学研究会, 63-76
4. 発表年 2019年

1. 発表者名 谷津元樹, 荒木健治
2. 発表標題 駄洒落検出用フレームワークの公開と動画コメントデータへの適用
3. 学会等名 第63回ことば工学研究会, 77-78
4. 発表年 2019年

1. 発表者名 佐山公一, 荒木健治, 片岡駿
2. 発表標題 AIが駄洒落の面白さを識別する 面白さの識別に必要な心理的要因
3. 学会等名 第64回ことば工学研究会, 65-70
4. 発表年 2020年

1. 発表者名 内田 ゆず, 荒木 健治
2. 発表標題 オノマトペに着目した駄洒落の面白さの分析 駄洒落の自動生成に向けて
3. 学会等名 第35回ファジシステムシンポジウム講演論文集(FSS2019), 332-336
4. 発表年 2019年

1. 発表者名 王格格, 荒木 健治
2. 発表標題 SVM を用いた駄洒落の面白さの予測
3. 学会等名 令和元年度電気・情報関係学会北海道支部連合大会, 143-144, 【2019 IEEE Sapporo Section Student Paper Contest Encouraging Prize】
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>(1) 駄洒落データベース http://arakilab.media.eng.hokudai.ac.jp/~araki/dajare.htm 67,000件を有する駄洒落データベースにクラウドソーシングで雇用した3名の評価者により5段階の面白さのスコア付けを行ったものを一般公開したものである。</p> <p>(2) djr_wpsm https://gitlab.com/m-yatsu/djr_wpsm (1)の駄洒落データベースを用いて駄洒落検出用のフレームワークを開発しGitLabリポジトリにおいて公開を行ったものである。</p>
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	内田 ゆず (UCHIDA Yuzu) (80583575)	北海学園大学・工学部・准教授 (30107)	

6. 研究組織（つづき）

	氏名 (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	佐山 公一 (SAYAMA Kohichi) (90271733)	小樽商科大学・商学部・教授 (10104)	
研究分担者	谷津 元樹 (MOTOKI Yatsu) (30805015)	青山学院大学・理工学部・助教 (32601)	