Vol.2018-EC-47 No.5
2018/3/16

単語間の意味的類似度を用いた数字語呂自動生成システムの実装と評価

袴田 はるか[1,a)]　磯山 直也[1,b)]　寺田 努[1,2,c)]　塚本 昌彦[1,d)]

概要：
覚えておかなければならない数字列を覚える際に，数字語呂合わせと呼ばれる手法がよく用いられる．しかし，既存のツールでは自分と関連のある事柄や，意味を他と関連づけた事柄はより記憶に残りやすいのに対し，そのような事柄と関連した数字語呂合わせは生成できない．そこで本稿では，語呂合わせに使われている単語間の類似度を計算することで，関連付けたい事柄との関連度に着目した数字語呂合わせの自動生成システムを提案する．実装したシステムを用いて生成された数字語呂合わせの有用性について，評価実験を行った．実験の結果，4 字程度の短い数字列に対しては，提案システムにより生成された語呂合わせが高順位となることが分かった．

Hakamata Haruka[1,a)]　Isoyama Naoya[1,b)]　Terada Tsutomu[1,2,c)]　Tsukamoto Masahiko[1,d)]

1. はじめに

パスワードや年号，友人の誕生日など，覚えておかなければならない数字列は数多く存在する．一般的に，数字列を覚える際には，数字語呂合わせと呼ばれる方法が用いられることが多い．数字語呂合わせとは，数字に読み仮名を与えて意味のある単語や文に置き換えることで，それ自体には意味をもたない数字列を，記憶に残りやすい形へ変換する手法である．

数字語呂合わせを自動的に生成するツールは既に複数開発されているが [1]，既存のツールで生成された数字語呂合わせは，使用頻度の低い単語や脈絡がない単語によって構成され，記憶に残りにくい場合がある．青木らは，数字語呂合わせの生成に日本語形態素解析を取り入れ，語呂合わせの文としての妥当性を計る手法を提案している [2]．しかし，この手法では，数字列と生成した語呂合わせ結果に関連をもたせることに目を向けられてはいない．例えば，服屋の電話番号「1129」には「いい服」を当てはめる，食べる事が好きな人にとっては「いい肉」の方が覚えやすい，というように，覚える意味を対象と関連づけた事柄や，自分と関連のある事柄は，そのまま覚えるよりも記憶に残りやすいとされているが [3], [4]，そのような関連度を重視した語呂合わせを生成する手法は，筆者の知る限り存在しない．

そこで本研究では，入力されたキーワードと語呂合わせに使われている単語間の類似度を計算することで，関連付けたい事柄と意味が類似した単語を用いた数字語呂合わせを生成するシステムを提案する．提案システムでは，語呂合わせの生成時に，ユーザが数字列だけでなく任意のキーワードも入力することで，そのキーワードに意味が類似した単語を用いた数字語呂合わせを自動生成する．この時，単語間の意味的類似度の計算には，Word2Vec [5] で取得した単語ベクトルを使用している．ユーザは，このシステムを用いることで，覚えたい事柄に関連する語呂合わせや，ユーザ自身が興味のある分野に関連する，記憶に残りやすい語呂合わせを生成できる．評価実験では，状況設定のある 9 つの数字列に対して，提案システムで生成した語呂合わせ 2 つと，既存手法で生成した語呂合わせ 3 つの合計 5 つの語呂合わせを被験者に順位付けさせ，システムの有用性の評価を行う．

以降では，2 章で関連研究を紹介し，3 章では提案システムについて述べ，4 章で実装について説明する．5 章では評価実験を行い，6 章で本論文をまとめる．

[1] 神戸大学工学研究科
Graduate School of Engineering, Kobe University
[2] 科学技術振興機構さきがけ
JST PRESTO
[a)] haruka-hakamata@stu.kobe-u.ac.jp
[b)] isoyama@eedept.kobe-u.ac.jp
[c)] tsutomu@eedept.kobe-u.ac.jp
[d)] tuka@kobe-u.ac.jp

2. 関連研究

2.1 暗記支援に関する研究

ユーザの暗記を支援するシステムは，数多く存在している．鈴木らは，軽運動の導入による暗記型学習支援シリアスゲーム Alpha-Motion を提案している [6]．Alpha-Motion では，マイクロソフト社の kinect を用いて，学習者の動きを計測し，英単語のスペルを大きくなぞるような運動を繰り返し行うことで，英単語の暗記型学習を支援している．この研究では，軽運動を実施中の語彙学習は，運動をしていないときの語彙学習よりも効果があることに着目している．數原らは，アニメーションを利用した英単語暗記支援システムを開発している [7]．この研究では，語源暗記法と呼ばれる，2 種類の英単語を組み合わせて複合英単語を覚える暗記方法に，アニメーションを導入することで，子どもの英単語の暗記を支援している．伊藤らは，暗記学習のための替え歌を自動で生成するシステムを提案している [8]．このシステムでは，学習者が暗記したい単語のリストを入力することで，慣れ親しんだ楽曲中の歌詞の区切りごとに譜割りや韻を考慮して単語を割り当てた，暗記に適した替え歌を生成できる．どちらの研究も，ある単語を記憶する際に，その形や発音だけについて考えるよりも，単語の意味について考えたり，既知の事柄と結びつけて考えたりするなどのより深い処理をする事で，再認や再生の記憶成績が向上する処理水準効果を利用している [4]．いずれの研究も，学習者の暗記を手助けするために，様々な手法が取り入れられているが，学習に時間がかかったり，道具や場所を要したりするなどの課題がある．

2.2 語呂合わせに関する研究

本研究で扱う数字列や専門用語などの簡便かつ効果的な暗記法として，一般的に語呂合わせと呼ばれる手法が知られている．語呂合わせを自動で生成するツールは数多く存在している．例えば語呂合わせジェネレータ [1] では，最大 11 桁までの数字列に対して語呂合わせを生成し，複数の候補を表示する．しかし，これらのツールによって生成された語呂合わせは，使用頻度の低い単語や日本語として意味をなさない単語の羅列になる場合があり，暗記に効果的な語呂合わせとはいえない．

青木らの研究では，形態素解析に用いられるコスト最小法のアルゴリズムを利用して，数字列に対する語呂合わせの自動生成を行っている [2]．このシステムは，数字 1 つ 1 つに読み方を割り当て，入力された数字列に対応する単語を辞書から探し，その組み合わせのパスコストを計算することで，語呂合わせの文としての妥当性を評価している．岡安らは，暗記したい単語列から頭文字を抜き出し，それらの文字を全て含む単語を辞書に登録している単語から探し出す日本語語呂合わせの自動生成手法を提案している [9]．入力単語数が長い場合は，頭文字を数個ずつに分け，それぞれで単語を決定したのち，文章らしくなるよう助詞などを付与することで，最終的な結果を出力している．また，語呂合わせに似た手法として，覚えておくことが難しいランダムな文字列で構成されたパスワードを，パスワード内の文字を用いたニーモニックと呼ばれる短文に変換して覚える暗記法がある．例えば，“jpwjaop”という文字列は，“Jill’s pet wolf just ate our pizzas. (ジルのペットの狼は，ちょうど私たちのピザを食べました．)”というニーモニックで覚えることができる．Jeyaraman らは，テキストコーパスに基づいた，パスワードに対するニーモニックの自動生成システムを提案している [10]．この研究では，新聞記事の見出し文をコーパスとして用い，意味辞書を用いた類義語の置き換えを行うことで，よりユーザの記憶に残りやすいニーモニックを生成している．Juang らは，自動生成したニーモニックをユーザに提示する際に，ニーモニックに関連する画像の検索結果をブラウザで表示したり，ログイン時に表示するヒントの画像をユーザ自身に描画させたりすることで，さらにユーザの記憶の定着を支援している [11]．いずれの研究も，数字列や文字列を意味の通るフレーズに変換することで，より暗記に効果的な語呂合わせやニーモニックを生成している．しかし，事柄と生成された語呂合わせの間の関連度に着目した語呂合わせの自動生成手法は，筆者の知る限り存在しない．

3. 提案システム

3.1 システム要件

覚えやすい語呂合わせの条件の 1 つとして，任意の事柄と関連しているかどうかがあげられる．例えば，服屋の電話番号「1129」の語呂合わせとしては,「いい肉」よりも「いい服」が適切である．これは，意味の関連について考えたり，既知の概念と組み合わせて覚えたりする深い処理を行うことで，2 章で述べた処理水準効果が現れると考えられるためである．また，食べる事が好きな人にとっては，「1129」というランダムパスワードに対して,「いい服」よりも「いい肉」というパーソナライズされた語呂合わせの方が印象に残ると思われる．その理由の 1 つとして，自己関連付け効果があげられる．自己関連付け効果とは，暗記時に自己に関連させた処理を行うと，意味的な処理や他者に関連した処理を行った場合と比較して，記憶保持が優れる現象である [3]．この効果により，システムによってパーソナライズされた語呂合わせの中に，ユーザが自身との関連を見出すことで，その記憶はより強固なものになると考える．

このように，事柄と語呂合わせの間の関連度を考慮することで，覚えたい事柄に関連する語呂合わせや，ユーザ自身が興味のある分野に関連する語呂合わせを自動生成できれば，よりユーザにとって暗記しやすい語呂合わせを生成

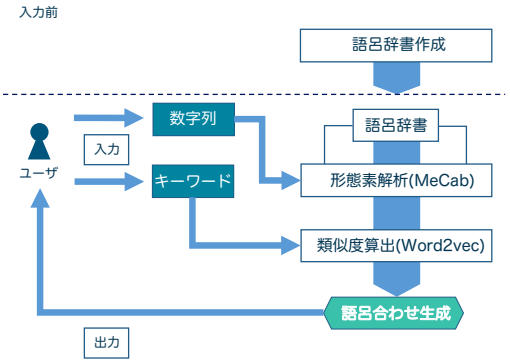


図 1　システム構成

できると考えられる．Jeyaraman らは，研究の中でニーモニックのパーソナライズに関して，新聞記事のコーパスには内容をカテゴリ分けするタグが付けられている事に着目し，サッカーファンのユーザに対して，株式取引よりもサッカー関連の見出し文から作られたニーモニックを優先的に提示できると述べている [10]．実際にサッカーに興味のある人にサッカー関連のニーモニックが有用かどうかかは研究内で検証されていないが、このような記憶と事柄の関連付けは、暗記のしやすさにつながると想定できる．

これらの関連研究をふまえて，ユーザにとって覚えやすい語呂合わせとは以下の条件を満たす単語列であると考える．

- 日本語として意味の通る単語列である．
- 任意の事柄と関連がある単語列である．

本稿では，それぞれの条件に対して，日本語形態素解析に用いられるコスト最小法と，Word2Vec を用いて取得した単語の概念ベクトルを利用する手法を取り入れた語呂合わせの生成システムを提案する．

3.2　システム構成

提案システムでは，ユーザが入力した，語呂合わせを生成したい数字列と関連させたい事柄を示すキーワードの 2 つから，キーワードに関連した数字語呂合わせを生成してユーザに提示する．システムの流れを**図 1** に示す．提案システムは，日本語形態素解析による語呂合わせ候補の絞り込み，Word2Vec によるキーワードと単語間の類似度算出の 2 つのフェーズで構成されている．語呂合わせの候補には，事前に語呂を登録した辞書（語呂辞書）を用いる．

まず，システム設計者は日本語辞書と各数字にふられた読み仮名のリストから，1–4 桁の数字列に対する語呂となる単語を辞書に登録しておく．その後，システムは次にこの語呂が登録された辞書を用いて，ユーザが入力した数字列に日本語形態素解析器 Mecab[12] のコスト最小法を利用して，コスト合計の小さい語呂の組み合わせパターンを語呂合わせの候補として選出する．さらに，その中から入力されたキーワードとの意味的類似度が高い単語を用いた数字語呂合わせを最終的にユーザに提示する．

表 1　数字の読み仮名

	0	1	2	3	4	5	6	7	8	9	10
音読み	レイ	イチ イ	ニ	サン ザン	シ	ゴ コ	ロク ロ	シチ	ハチ ハ バ パ	キュウ ク グ	ジュウ ジッ
訓読み	ゼロ	ヒト ヒ ビ ピ	フタ フ ブ プ	ミ	ヨン ヨ ョ	イツ	ム ムイ	ナナ ナノ ナ	ヤ ヨウ ャ	ココノ ココ コ	トオ ドオ
英語		ワン	ツー ズ		フォー						テン
その他	オ マ ワ エン		ジ								

以下に，語呂辞書の作成・形態素解析による語呂合わせの絞り込み・単語間の類似度を用いた語呂の評価，それぞれについて詳細を示す．

3.2.1　語呂辞書の作成

日本語形態素解析に用いられる IPA 辞書に載せられている単語の数は，表記揺れを除いても 28 万を超える膨大な数になる．数字列から語呂合わせを生成する際に，この IPA 辞書の中から読みに合致する単語をその都度検索する方式を取ると，桁数が長くなるにつれ生成に多くの時間がかかる．そのため，数字列の入力の前に，数字列の読みに当てはまる単語のみを語呂として登録した独自の語呂辞書を作成しておくことで，システムの高速化を計る．この時，語呂辞書に登録する語呂については，0–9 の 10 個の数字を並べた 1–4 桁の数字列に対してそれぞれ読み仮名を与え，IPA 辞書の発音に一致した単語を語呂と定義した．読み仮名を与える数字については，1 桁の数字「0123456789」の 9 つと，2 桁の数字「10」の 1 つとした．各数字に与えた読み仮名の一部を**表 1** に示す．

本研究では，一般的に数字語呂合わせで用いられている読みを参考に，以下の基準で数字の読み仮名を設定した．

- 読み仮名の基本
  - 各漢数字の音読み (1=イチ，2=ニ，3=サン)
  - 各漢数字の訓読み (1=ヒト，2=フタ，3=ミ)
  - 各数字の英語のうち音節数が 1 つのもの (1=ワン，2=ツー，4=フォー)
- 読み仮名の最初の一文字 (1=イ，ヒ)
- 濁点，半濁点の付け外しを行った読み仮名 (1=ビ，ピ)
- 促音，長音，撥音を付け外しした読み仮名 (1=ヒッ，ヒー，ヒン)
- 拗音に変換した読み仮名 (4=ョ)
- 意味や形から連想した読み仮名 (2=次=ジ，0=輪=ワ)

以上から設定された読み仮名は，計 180 個となる．この読み仮名を使って変換した読み仮名の組み合わせが，IPA 辞書内の発音に一致した単語の中から，人名，地域名，数詞，記号を除いた単語を，語呂として辞書に登録した．この時，後述する日本語形態素解析のフェーズでの使用を想定し，語呂辞書の表層形を語呂に対応する数字列に変換し

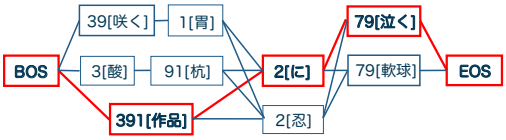


図 2　数字列のラティス構造と最適解

ておくことで，Mecab がそのまま参照できる形にしておく．語呂辞書に登録された語呂の総数は，33487 個である．

3.2.2　形態素解析による語呂合わせの絞り込み

入力された数字列に対して，適用できる語呂合わせのパターンは無数に存在する．しかし，ある程度日本語として成立している語呂合わせでないと，ユーザにとって覚えやすく，思い出しやすい語呂合わせであるとはいえない．そのため，無数の語呂の組み合わせパターンの中から，文として妥当といえる語呂合わせを絞り込む必要がある．提案システムでは，日本語形態素解析器 Mecab に実装されているコスト最小法を用いて，文としての妥当性を評価し，語呂合わせの候補を選出する．

日本語形態素解析とは，文を日本語の最小単位である形態素へ分割する技術である．Mecab は，文章として最も確からしい単語の並びを推定するために，コスト最小法を採用している．コスト最小法とは，ラティスと呼ばれる文に対しての全ての単語の候補を列挙した構造の中から，最もコストの合計が小さくなる単語の組み合わせパターンを最適解とする手法である．このコスト最小法を，日本語の文章ではなく，数字列に対して利用することで，数字語呂合わせの文としての妥当性を評価できると考えられる．

このフェーズでは，ユーザが入力した数字列に対して，先頭の数字から一桁ずつ語呂辞書内を検索して数字列に当てはまる語呂を取り出し，日本語形態素解析と同様にラティス構造を構築する．数字列のラティス構造のイメージを図 2 に示す．そして，その中の全ての語呂合わせパターンについて，日本語として意味の通る単語列であるかどうか，コスト最小法を用いて文としての整合性を評価し，順位付ける．この時，各単語に割り当てられる単語の出現頻度を示す生起コストと，品詞による前後の単語間のつながりやすさを示す連接コストは，Mecab の IPA 辞書データベースのコスト値を利用した．また，ラティス内のコスト合計を求めるアルゴリズムについても，Mecab 内のアルゴリズムをそのまま用いることで，桁数によらない高速な計算を可能にしている．本研究では，コストの小さい上位 500 個の語呂の組み合わせを，日本語として自然な語呂合わせとして選出する．

3.2.3　単語間の類似度を用いた語呂の評価

提案システムでは，ユーザに関連させたい事柄としてキーワードを入力させる．語呂合わせを文と考えた時，文とキーワード間の関連度は，文中で使われている単語の意味がキーワードにどれだけ似ているかを算出することで判別できると考える．本研究では，Word2Vec という自然言語処理の手法を用いて，語呂合わせに使われている単語と，入力されたキーワードとの類似度を計算し，類似度の高い単語を用いた語呂合わせを関連している語呂合わせとしてユーザに提供する．

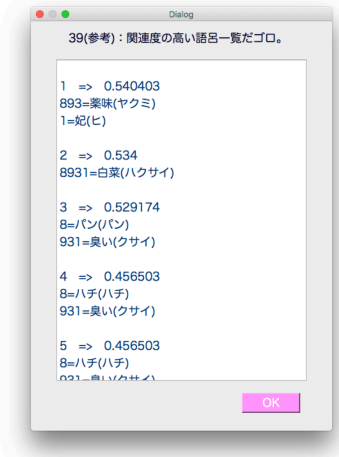
Word2Vec とは，単語に分けられた文章を学習させることで，ある単語の文章中での使われ方を 200 次元程度のベクトルの集合に変換し，その単語の特徴量とする手法である．似た意味を持つ単語は，文章内でも同じような文脈で使われることが多いため，Word2Vec で表した際の単語の特徴ベクトルも似た値をとる．そのため，Word2Vec で表した単語の特徴ベクトル間のコサイン類似度を計算することで，単語間の意味の類似度を計算することができる．今回単語間の類似度計算を行う Word2Vec の学習モデルは，鈴木らが配布している日本語 Wikipedia エンティティベクトル [13] を使用した．

このフェーズでは，日本語形態素解析にかけて選出した 500 個の語呂合わせに対して，Word2Vec の学習モデルから算出した，入力されたキーワードの特徴ベクトルと，語呂合わせ内の単語の特徴ベクトルとのコサイン類似度を求める．類似度を求める単語については，文の意味を正確に表すことを考え，語呂合わせに使われている単語の中でも名詞のみを対象とした．また，類似度算出の対象になっていても，特徴ベクトルがモデル内で定義されていない単語については，キーワードとの類似度を 0 として計算している．語呂合わせ内の単語についてコサイン類似度を求めた後，それぞれの語呂合わせ内の単語の類似度の平均を算出する．この時，数字語呂合わせについては，より桁数の多い数字列を変換する語呂のほうがユーザに与える影響が大きいと考えられるため，各単語が変換する数字列の桁数を重みとして，類似度の加重平均を求める．語呂合わせ内の類似度を求める対象の単語の数を n，対象となった単語のキーワードとのコサイン類似度を s_i，類似度を求めた単語の変換する数字列の桁数を l_i とすると，語呂合わせ全体のキーワードとの平均コサイン類似度 S は以下のように表せる．

$$S = \frac{\sum_{k=1}^{n} s_i * l_i}{\sum_{k=1}^{n} l_i} \tag{1}$$

例えば，8 桁の数字列に対して生成した語呂合わせ内で，類似度を求める対象となる単語が 4 桁に対する語呂と 2 桁に対する語呂の 2 つが存在した場合，4 桁の語呂のコサイン類似度が 0.4，2 桁の語呂のコサイン類似度が 0.6 の時は，平均類似度 $S = 0.4 * 4 + 0.6 * 2/4 + 2 = 0.4667$，4 桁の語呂のコサイン類似度が 0.6，2 桁の語呂のコサイン類似度が 0.4 の時は，平均類似度 $S = 0.6 * 4 + 0.4 * 2/4 + 2 = 0.5333$ となる．全ての語呂合わせについて平均類似度を算出した後，平均類似度の高い順に語呂合わせのソートを行い，平均類似度が最も高かった語呂合わせを最適な数字語呂合わ

(a) 出力画面



(b) 語呂一覧画面

図 3 プロトタイプシステム画面

せとしてユーザに提示する．

4. 実装

3 章で述べた提案手法をもとにして，数字語呂合わせの自動生成システムのプロトタイプを作成した．

図 3 にスクリーンショットを示す．図 3(a) 中の「数列」の入力欄に語呂合わせを生成したい数字列，「言葉」の入力欄に関連させたい事柄を表すキーワードを入力して「Create」ボタンを押すと，最も平均類似度が高かった語呂合わせが出力される．さらに，画面右部の「GORO」ボタンを押すことで，平均類似度の高かった上位 10 つの語呂合わせを確認できる．また，「Reset」ボタンを押すと入力欄と出力欄がクリアされ，「Exit」ボタンでシステムを終了する．

表 2 に，プロトタイプシステムによって生成された語呂合わせの上位 10 つの生成例を示す．表のように，同じ数字列に対しても，入力したキーワードによって，生成される数字語呂合わせは異なる．さらに例えば，「食べ物」というキーワードに対しては「薬味」「白菜」「パン」「胃」，「化学」というキーワードに対しては「薬剤」「錆」「剤」など，

表 2 システムによる生成例

数字列		キーワード	
8931		食べ物	化学
語呂合わせ (類似度)	1	薬味 妃 (0.54)	薬剤 (0.46)
	2	白菜 (0.53)	約 錆 (0.33)
	3	パン 臭い (0.53)	パクン 未 ヒト (0.32)
	4	ハチ 臭い (0.46)	パクン 未 胃 (0.29)
	5	パクン 未 胃 (0.39)	野球 剤 (0.23)
	6	パン粉 サイン (0.37)	野球 錆 (0.21)
	7	野球 味 ピッ (0.37)	約 雑品 (0.19)
	8	野球 味 ピン (0.37)	約 産院 (0.18)
	9	ハクサイ (0.36)	パクン 産院 (0.18)
	10	野球 味 ヒト (0.35)	ハクサイ (0.18)

数字列		キーワード	
1024		音楽	スポーツ
語呂合わせ (類似度)	1	一円 舞踊 (0.27)	イレブン よ (0.36)
	2	一連 風刺 (0.23)	イレブン よう (0.27)
	3	陶磁 よ (0.22)	テンプ よ (0.26)
	4	一連 諷刺 (0.20)	イレブン 余 (0.25)
	5	陶磁 よう (0.22)	トップ 誌 (0.24)
	6	一円 風刺 (0.19)	道場 (0.22)
	7	純情 (0.18)	トップ よ (0.22)
	8	同情 (0.18)	トップ ヨ (0.22)
	9	一円 に 詩 (0.16)	トップ ヨー (0.22)
	10	イン は ズシン (0.16)	トップ よー (0.22)

キーワードに意味が似ていたり，同じカテゴリに属したりしている単語が語呂合わせに用いられているのが分かる．なお，ソフトウェア開発には PC は Apple 社の MacBook Pro (CPU: Core i5，メモリ: 8GB) を使用して，Python 統合開発環境である PyCharm を用い，システム GUI の作成にはアプリケーションフレームワークの Python 用バインディングライブラリ PyQt5 を用いた．

5. 評価実験

本稿では，事柄と語呂合わせに用いられる単語との類似度を評価基準に取り入れる手法を提案した．提案システムで生成した語呂合わせと，既存手法で生成した語呂合わせに対して被験者に順位付けを行わせることで比較を行い，その結果を考察することで提案システムの評価を行う．

5.1 実験内容

実験は，「好きなアイドルの誕生日」などの状況を設定した 9 つの数字列 ((a)–(i)) に対して，提案システムで生成した語呂合わせ 2 つと既存手法で生成した語呂合わせ 3 つを合わせた 5 つの語呂合わせを被験者に提示し，良い語呂合わせだと思う順に，1 位から 5 位まで順位付けさせる．被験者には，数字列・語呂合わせと共に，その数字列を使う状況を想定した状況説明を提示する．さらに，各数字列に

表 3　数字列と語呂合わせの一覧

数字列	(a)0820	(b)3979	(c)1320
状況設定	好きなアイドルの誕生日	好きなアイドルの誕生日	大学工学部棟パスワード
手法 1	オヤジ は	サンキュー なく	ピッ 雑音
手法 2	音盤 連れ	ミック ナック	一座 プレー
手法 3	マッハ 連れ	未 コーナー 郡	移民 連れ
手法 4	お初穂	山内 新築	悲惨 プレイ
手法 5	おやつ は	巫女 泣く	秘密 は

数字列	(d)6619	(e)9497	(f)2654268
状況設定	大学工学部棟パスワード	大学工学部棟パスワード	ペットショップの電話番号
手法 1	労々い 工	工 良くな	フッ 無ゴシップ ロバ
手法 2	無論 ビーコン	講師 コーナー	フッ 向こう 要心炉 ハア
手法 3	ローム 一個	グッ 至高な	フッ 無 紅葉 ぶろ ハア
手法 4	無力 ピンク	屈伸 苦難	ブーム 腰 プロパン
手法 5	ROM 行く	来よ 来な	ふむ いつ 世 に 産む や

数字列	(g)5753854	(h)7276136103	(i)2284071200
状況設定	大学工学部棟パスワード	ケーキ屋の電話番号	大学工学部棟パスワード
手法 1	こんな ゴミ や 雇用	ナズナ 炉 ビザ 炉 ジューサー	フッ 突っ走れ ない 不 練磨
手法 2	こんな 講座 番号 よ	ナズナ 浪費 寒い おっさん	フッ 突っ走れ ない 不 応援
手法 3	こんな ゴミ箱 よ	ナズナ 炉 一座 炉 登山	ブンブン 走れ ない ジレンマ
手法 4	困難 ゴミ 箱師	夏 な ムービー 魅力 倒産	辻番 よ 学び ジレンマ
手法 5	粉 ゴミ箱 よ	何 南無 悲惨 武藤 さん	ジジ 橋 罠 いつ 折れ

おける語呂合わせの順位付けについて，判断基準や選定理由などを自由に記述できるコメント欄を用意した．被験者は 20 代の男女 20 名で，筆者と同じ研究室に所属する学生である．

提示した数字列と語呂合わせの一覧を**表 3** に示す．用意した数字列は，0–9 の数字をランダムに組み合わせており，4 桁の数字列を 5 つ，7 桁の数字列を 2 つ，10 桁の数字列を 2 つの計 9 つである．この時，誕生日や電話番号など，与えた状況説明に常識上そぐわない数字列も対象に含まれている．与えた状況説明のうち，「大学工学部棟パスワード」については，研究室に通う学生が頻繁に使うと思われるため，実験に使用した．5 つの語呂合わせの生成手法については，以下である．

手法 1　提案システム使用-設定とダイレクトに関連するキーワード

手法 2　提案システム使用-設定と広義的に関連するキーワード

手法 3　日本語形態素解析器 Mecab のみ使用

手法 4　文献 [1] の語呂合わせジェネレータを使用

手法 5　読み仮名の一覧のみを見ながら筆者が作成

手法 1，2 については，各数字列に与えた状況設定に結びつけやすいと考えられる単語を 2 つ事前に設定し，入力するキーワードとした．提案システムに入力したキーワードの一覧を**表 4** に示す．なお，手法 1–4 に関しては，結果の候補に上がった上位 10 個の中から，最も優れていると思う語呂合わせを筆者が選んで提示する．特に，手法 1，2 では，優れていると感じる語呂合わせがどちらの候補にも含まれていた場合，類似度平均が高かった方の手法にて提示する．また，提示される語呂合わせは被験者によって異なる順序で提示されるようにしており，提示順による回答の偏りがなくなるようにする．

表 4　キーワードの一覧

状況設定	手法 1	手法 2
好きなアイドルの誕生日	アイドル	音楽
ペットショップの電話番号	ペット	生物
ケーキ屋の電話番号	ケーキ	食べ物
大学工学部棟パスワード	工学	大学

5.2　結果

実験結果については，被験者が答えた語呂合わせの順位に対し，1 位から 5 位まで，順に 5，4，3，2，1[pt] の得点をつけて評価する．それぞれの数字列に対して，被験者が語呂合わせにつけた得点をまとめた結果の箱ひげ図を**図 4** に示す．箱の中央にある太い線が得点の中央値，箱の上部下部が 25 パーセンタイル，75 パーセンタイル，ひげの上端・下端が最大値・最小値を表し，○は外れ値，▲は箱の高さの 3 倍を超える極端な外れ値を示す．

まず，(a)–(i) の数字列ごとに，提案システムである手法 1，2 と従来手法である手法 3，4，5 を比べた結果，数字列 (a) では提案システムである手法 1 の得点が，従来手法 3，4 よりも多かった．数字列 (b) では手法 1 の得点が手法 3，4，5 より高く，手法 2 の得点が手法 3，4 よりも高かった．これより，数字列 (a)，(b) においては，提案システムが優れていることがわかった．しかし，それ以外の数字列については，提案システムである手法 1，2 が，手法 3，4，5 よりも得点が高いという結果は得られなかった．そのため，提案システムと従来手法の間には，明確な差はなかったと考えられる．

5.3　考察

提案システムである手法 1，もしくは手法 1，2 の両方が高順位であった数字列に関して，数字列 (a) は「0820(好きなアイドルの誕生日)」で，手法 1 による語呂合わせが「オヤジ は」，数字列 (b) は「3979(好きなアイドルの誕生日)」で，手法 1 による語呂合わせが「サンキュー なく」，手法 2 による語呂合わせが「ミック ナック」である．まず数字列 (a) について，手法 1 の「オヤジ は」に高順位をつけた被験者のコメントから，オヤジという単語とアイドルのイメージを組み合わせて考え，設定された状況との関連度を重視した被験者が複数いることが分かった．一方，数字列 (b) では，手法 1 の「サンキュー なく」については「本来の数字の読み方に近いため覚えやすい」，手法 2 の「ミック ナック」については「意味は分からないけれど語感が良い」というコメントが複数見られ，事柄との関連度が評価に結びついてはいなかった．しかし，数字列 (a)，そして同じ 4 桁の数字列 (c)「1320(大学工学部棟パスワード)」，(d)「6619(大学工学部棟パスワード)」において，他の手法より順位が高かった従来手法である手法 5 によって作られ

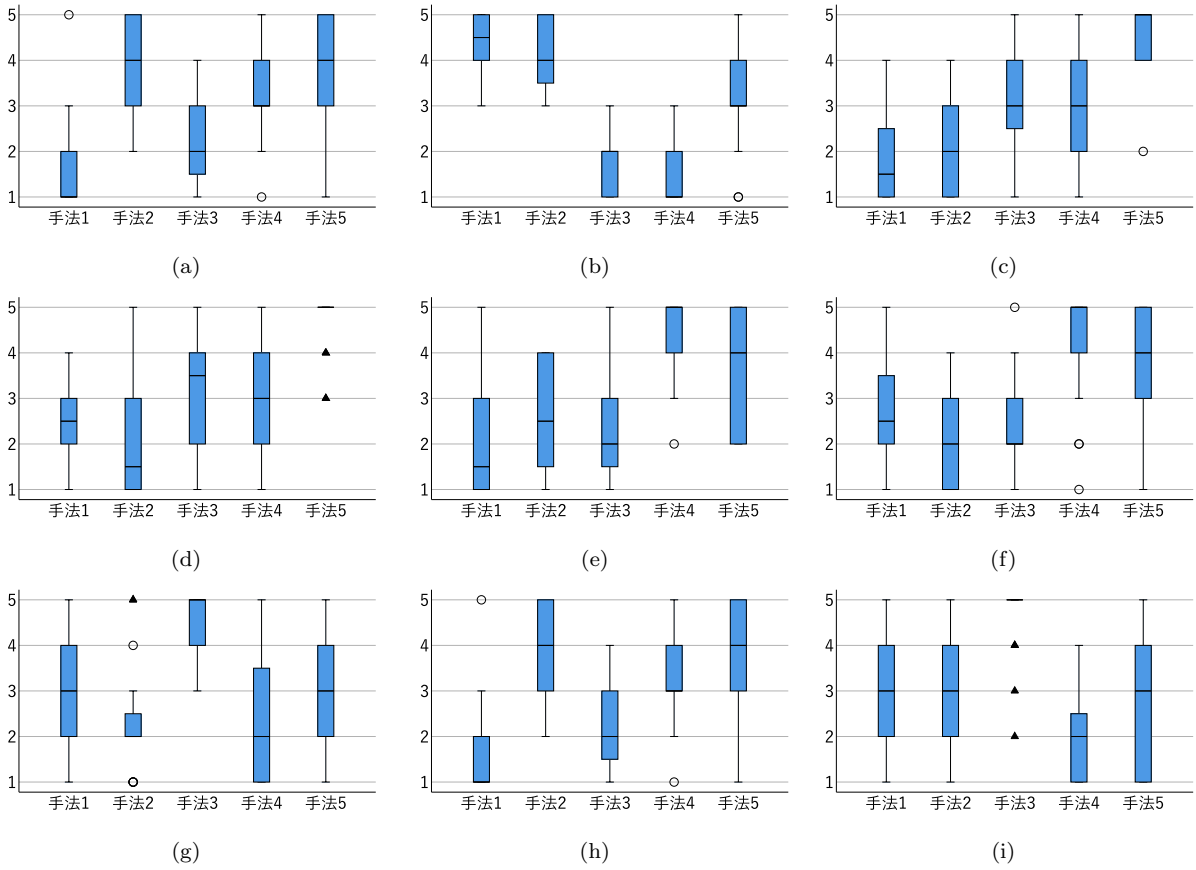


図 4 順位付けによる数字列別得点

た語呂合わせ「おやつ は」,「秘密 は」,「ROM 行く」について，それぞれ「おやつとアイドルのスイーツ感を結びつけることができた」,「パスワードと秘密という単語は関連していて覚えやすい」,「ROM という単語に工学部との関連を感じた」というような，事柄との関連度に注目した被験者のコメントが複数あった．これらより，このような比較的短い数字列に対しては，ユーザは使われている単語の意味と覚えたい事柄との関連を考え，関連がある語呂合わせを覚えやすいと感じる傾向があると推測できる．

また，7 桁の数字列 (g)「5753854(大学工学部棟パスワード)」，10 桁の数字列 (i)「2284071200(大学工学部棟パスワード)」においては，日本語形態素解析，つまり文の日本語としての整合性のみを評価した従来手法である手法 3 の語呂合わせ「こんな ゴミ箱 よ」,「ブンブン 走れない ジレンマ」が，他の手法よりも順位が高かったことが確認できた．この結果の理由として，被験者のコメントから，ユーザが特に長い数字列を覚える際に，事柄との関連度よりも，文として意味が通っているかを重視して語呂合わせを評価したためであると推測できる．そのため今後は，入力された数字列の桁数が長くなるのに応じて，日本語形態素解析で得た順位の重みを大きくしていくなどの方向にシステムを改良していく．

本実験では，各語呂合わせの評価基準について，事柄との関連度や文としての妥当性の他にも，読み仮名が数字の発音に近いかどうか，使われているのが馴染みのある単語かどうか，口に出して読みやすいかどうか，前後の単語と合わせて意味が通るかどうかなどの様々な意見が見られた．今後は，事柄との関連度だけでなく，このような評価項目について定量的に評価するために，例えば数字の本来の発音により近い読み仮名を用いている単語はスコアを高くする，コスト最小法に用いられる単語の出現頻度を示す生起コストの値の重みを大きくする，入力されたキーワードだけでなく語呂合わせ内の単語同士でも特徴ベクトルの類似度を計算するなど，どのような手法をシステムに取り入れて対応していくかを考えていく必要がある．

実験全体を通して，自分が普段数字語呂合わせに使用している読み仮名から外れた読み方をする語呂合わせに対して，順位を低くつける被験者が多く見られた．ユーザによって数字に与える読み仮名は個人差が大きいため，各読み仮名を各々が自由に選択した上で，個人にカスタマイズされたユーザ語呂辞書を用いることによって，より良い語呂合わせをユーザに提供できると考えられる．

6. まとめ

本研究では，単語の特徴ベクトルのコサイン類似度を利用して，任意の事柄と関連を持った数字語呂合わせの生成

システムを実装した．ユーザはシステムを使って，覚えたい事柄や自身が興味のある分野に関連する事柄をキーワードとして入力することで，キーワードと意味の似た単語を用いた記憶に残りやすい語呂合わせを生成できる．提案システムでは，日本語形態素解析に用いられるコスト最小法と，Word2Vec を用いて取得した特徴ベクトルを利用する手法を取り入れた．

評価実験では，9 つの状況設定のある数字列に対して，提案システムで生成した語呂合わせ 2 つと，既存手法で生成した語呂合わせ 3 つの比較を順位付けによって評価した．実験の結果，提案手法と既存手法との間には，差は見られなかった．しかし，短い数字列に対しては，ユーザは語呂合わせに使われている単語の意味を考え，覚えたい事柄と関連した語呂合わせを高順位にしていたことが分かった．また，長い数字列に対しては，語呂合わせの評価基準として，文として意味が通っているかを重視する傾向があることが分かった．

今後の課題としては，評価実験で得た知見から，さらに覚えやすい語呂合わせを生成できるよう，システムの評価基準を再考，追加することがあげられる．例えば，覚える対象の数字列の桁数が大きくなるにつれ，日本語形態素解析の順位の重みを大きくすることや，キーワードとの類似度だけでなく，前後の単語についても類似度の計算を行って語呂合わせを評価することなどが考えられる．また，ユーザによって数字に与える読み仮名の個人差が大きいために，各読み仮名を各々が自由に選択した上で，カスタマイズされたユーザ語呂辞書を使用することもあげられる．システムによって生成された語呂合わせを使用して，ユーザが実際に数字列を長期間覚えていられるかどうかを調べることも必要である．

参考文献

[1] 語呂合わせジェネレータ: http://seoi.net/goro/.
[2] 青木賢太郎: コスト最小法を用いた言葉遊び–数字語呂合わせの自動生成システム, ことば工学研究会 (第 10 回), pp. 31–35 (2002) .
[3] T. B. Rogers, N. A. Kuiper, and W. S. Kirkier: Self-reference and Encoding of Personal Information, *Journal of Personality and Social Psychology, 35*, pp. 677–688 (1977).
[4] F. I. M. Craik and E. Tulving: Depth of Processing and Retention Og Words in Episodic Memory, *Journal of Experimental Phychology: General*, 104, pp. 268–294 (1975).
[5] T. Mikolov, K. Chen, G. Corrado, and J. Dean: Distributed Representations of Words and Phrases and Their Compositionality, *In Advances in Neural Information Processing Systems*, pp. 3111–3119 (2013).
[6] 鈴木雄次郎 , 江袋天亮 , 小林篤史 , 粟飯原萌 , 古市昌一: 軽運動の導入による暗記型学習支援シリアスゲームの試作と初期評価, 情報処理学会 第 78 回全国大会講演論文集, pp. 739–740 (2016).
[7] 數原綾華, 角 薫: アニメーションを利用した英単語暗記支援システム, インタラクション 2013 論文集, pp. 727–729 (2013).
[8] 伊藤悠真, 寺田 努, 塚本昌彦: Mnemonic DJ: 暗記学習のための替え歌自動生成システム, 情報処理学会論文誌, Vol. 56, No. 11, pp. 2165–2176 (2015).
[9] 岡安優弥, 高田雅倫, 渡辺邦浩, 濱川 礼: 品詞による文評価を用いた日本語語呂自動生成手法, 情報処理学会創立 50 周年記念 (第 72 回) 全国大会, pp. 511–512 (2010).
[10] S. Jeyaraman and U. Topkara: Have the Cake and Eat It Too- Infusing Usability into Text-password Based Authentication Systems, *Proceedings of the 21st Annual Computer Security Applications Conference (ACSAC'05)*, pp. 473–482 (2005).
[11] K. A. Juang, S. Ranganayakulu, and J. S. Greenstein: Using Systemgenerated Mnemonics to Improve the Usability and Security of Password Authentication, *Proceedings of the 56th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 506–510 (2012).
[12] MeCab, "MeCab:Yet Another Part-of-Speech and Morphological Analyzer" : http://taku910.github.io/mecab/.
[13] 日本語 Wikipedia エンティティベクトル: http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/.