

## 対照群付き高齢者コーパスの構築とそれを用いた認知症予備軍スクリーニング技術の開発

柴田, 大作  
奈良先端科学技術大学院大学

伊藤, 薫  
Kyushu University

若宮, 翔子  
奈良先端科学技術大学院大学

荒牧, 英治  
奈良先端科学技術大学院大学

<https://hdl.handle.net/2324/7183423>

---

出版情報 : Transactions of the Japanese Society for Artificial Intelligence. 34 (4), pp.B-J11\_1-9, 2019-07-01. 人工知能学会

バージョン :

権利関係 : Notice for the use of this material. The copyright of this material is retained by the Japanese Society for Artificial Intelligence (JSAI). This material is published on this web site with the agreement of the author(s) and the JSAI. Please be complied with Copyright Law of Japan if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) The





# 対照群付き高齢者コーパスの構築と それを用いた認知症予備軍スクリーニング技術の開発

## Detecting Early Stage Dementia based on Natural Language Processing

柴田 大作  
Daisaku Shibata  
奈良先端科学技術大学院大学  
Nara Institute of Science and Technology  
shibata.daisaku.rr8@is.naist.jp, <https://sites.google.com/view/daisaku-s/home>

伊藤 薫  
Kaoru Ito  
九州大学  
Kyushu University  
ito@flc.kyushu-u.ac.jp, <http://flc.kyushu-u.ac.jp/~ito/>

若宮 翔子  
Shoko Wakamiya  
奈良先端科学技術大学院大学  
Nara Institute of Science and Technology  
wakamiya@is.naist.jp, [http://sociocom.jp/~wakamiya/index\\_Japanese.html](http://sociocom.jp/~wakamiya/index_Japanese.html)

荒牧 英治  
Eiji Aramaki  
(同上)  
aramaki@is.naist.jp, <http://luululu.com/aramaki/>

**Keywords:** natural language processing, dementia, corpus, Alzheimer's disease, screening

### Summary

We construct an elderly corpus with a control group (EC) comprising narratives of elderly people with Mild Cognitive Impairment (MCI), healthy elderly people, and younger people in order to develop a method to classify the elderly into healthy and MCI by analyzing the corpus. To do so, we carry out three tasks (picture description task: PDT, episode picture description task: EDT, and animation description task: ADT) to participants ( $n=80$ ) and their voices to the tasks are recorded and manually transcribed. 60 out of the participants are the elderly and classified into MCI and healthy control based on Mini Mental State Examination (MMSE). Then, language features such as Type Token Ratio and Idea Density are extracted by analyzing the elderly people's data and machine learning models are built with the extracted features. In the experiments, our classification model using combined language features obtained from all tasks' data achieved the highest performance (AUC = 0.85). The results indicate that it would be important to carry out multiple tasks to detect the elderly with MCI.

## 1. はじめに

我が国における急激な高齢化に伴い認知症高齢者の数が増加している。特に、治療が困難なアルツハイマー型認知症が大部分をしめるため、早期発見によりその進行を遅らせることが重要である。医療機関における認知症の検査方法として、Positron Emission Tomography (PET) 検査やMagnetic Resonance Imaging (MRI) 検査がある。また、これらの前に用いる迅速かつ簡易な検査法(以降、スクリーニングと呼ぶ)として、Mini Mental State Examination (MMSE) による認知機能検査や長谷川式認知症スケール (HDS-R) がある。しかし、これらの検査にはいくつかの問題が存在する。PET 検査は体

内に放射性物質を注入する必要がある、身体的侵襲を伴う。MRI は長時間の身体拘束が必要であり、MMSE やHDS-R は拘束時間は短い、精神的な負担を強いるものとなっている。そのため、これらの検査を日常的に受診することは難しい。

そこで近年、言語能力に基づく認知症スクリーニングが注目されている。一般的に認知症患者は健常者と比較して、言語能力が低下することが知られている。具体的に、認知症者は発話において使用する単語数が減少すること[荒牧 16b]や代名詞の使用が増加すること[Shibata 16]が知られており、自然言語処理技術によりこれらの特徴を測定することで、認知症のスクリーニングが可能となる。言語能力による認知症のスクリーニングは、医療

機関で行う診断に比べて、精度は低下するが、身体的侵襲や長時間の拘束を必要としないことが考えられる。また、日常的に調査できるため、わずかな変化を察知し、早期発見に繋がる可能性がある。

今日に至るまで言語能力に基づく認知症のスクリーニングに関する研究は数多くなされてきた。言語に注目した研究は1996年のNun Study [Snowdon 96]に始まり、機械学習により認知症者と健常者の識別に関する研究 [Jarrold 10, Orimaye 14]が英語圏では盛んに報告されている。英語圏ではDementia Bank<sup>1</sup>から健常者と認知症者の画像説明課題における発話データを取得することができる。この発話データは実験参加者に cookie theft picture と呼ばれる画像について説明をさせるタスクを実施したものであり、多くの先行研究で使用されている [Fraser 16, Karlekar 18, Orimaye 14]。このようなパブリックデータセットが存在することが英語圏において自然言語処理を用いた認知症のスクリーニングに関する研究が盛んである理由だと考えられる。一方で、日本語においてこのようなデータは利用することができず、このことが日本語で自然言語処理を用いた認知症のスクリーニングに関する研究の障壁となっている。

そこで我々は、日本語における認知症者と健常者の発話データセットの構築とそれを用いた軽度認知症者と健常高齢者の自動識別を行う。なお、認知症者の発話を取得することは困難であるため軽度認知症を対象とする。以後、対照群付き高齢者コーパスと呼ぶ。対照群付き高齢者コーパスは80人（若年者:  $n = 20$ , 高齢者:  $n = 60$ ）に対して3種類の刺激課題（先行研究で広く使用されている画像説明課題、これに加えてエピソード説明課題とアニメーション説明課題）を実施し、取得した発話データと高齢者のMMSEの結果から構成される。本研究ではさらに基礎的な調査として、コーパスから高齢者のデータを抽出し、MMSEに基づき健常高齢群と認知機能障害群に分類し、両者の自動識別に関する実験と評価を行う。

## 2. 認知症について

認知症とは「生後いったん正常に発達した種々の精神機能が慢性的に減退・消失することで、日常生活・社会生活を営めない状態」のことであり、後天的原因により生じる知能の障害である。また、軽度認知障害はアルツハイマー病など認知症の前駆状態を意味する状態のことを指す<sup>2</sup>。

認知症の診断は米国精神医学会による診断マニュアルであるDiagnostic and Statistical Manual of Disorders-5 (DSM-5)、国際疾病分類第10版 (ICD-10)、またはNational Institute on Aging-Alzheimer's Association

(NIA-AA)の診断基準に基づいて行われる。認知機能障害だけでなく社会生活の障害を確認することが大切である。生活機能の低下があれば認知症を疑い、概ね自立している場合は、軽度認知障害 (MCI: Mild Cognitive Impairment) を疑う。

認知症の診断方法の1つにMini-Mental State Examination (MMSE) がある。MMSEは認知機能検査の1つであり、MMSEは時間の見当識、場所の見当識、3単語の即時再生と遅延再生、計算、物品呼称、文章復唱、3段階の口頭命令、書字命令、文章書字、図形模写の計11項目から構成される30点満点の認知機能検査である。MMSEは23点以下が認知症疑いであり（感度81%、特異度89%）、27点以下は軽度認知障害 (MCI) が疑われる（感度45-60%、特異度65-90%）<sup>3</sup>。

本研究では上記に基づき、MMSEが28点以上の高齢者を健常高齢者、23点以上かつ27点以下の高齢者を軽度認知障害とみなす。

## 3. 関連研究

自然言語処理を用いた認知症のスクリーニングで対象とされている疾患は大きく分けて3つあり、認知症 (Dementia)、アルツハイマー型認知症 (Alzheimer's Disease) と軽度認知障害 (Mild Cognitive Impairment: MCI) である。関連研究を以下に示す。

Snowdonらは93人の修道女が20歳代のときに書いた自叙伝から、意味密度（文章中の命題数を単語数で除した値）や文法の複雑さを算出した [Snowdon 96]。そして、若年期の意味密度と文法の複雑さの値が低いと、晩年期の認知機能が低いことを示し、これらの値と認知機能の間には一定の関係があることを報告した。また、Kemperらは、文法の複雑さは認知症の有無にかかわらず年齢と共に低下するが、意味密度は認知症群のみで低下することを報告した [Kemper 01]。Jarroldらは認知症者 ( $n = 22$ )、健常者 ( $n = 23$ ) の発話データを用いて、発話中の品詞数、意味密度や語彙カテゴリ（語彙をカテゴリ化するツールであるLinguistic Inquiry and Word Count (LIWC) を使用）などを抽出し、機械学習の特徴量として用いた結果、認知症者と健常者を最大73%の精度で分類できたことを報告した [Jarrold 10]。OrimayeらはDementia Bankで提供されている認知症者 ( $n = 314$ ) と健常者 ( $n = 242$ ) から構成されるコーパスからいくつかのSyntactic featuresとLexical featuresを抽出し、機械学習による分類を行い、0.74のF値を達成した [Orimaye 14]。田中らはコンピュータアバターとの対話により、Lexical featuresとSpeech featuresに加えて、顔の表情の特徴量であるImage featuresを抽出するシステムを用いて、認知症者 ( $n = 14$ ) と健常者 ( $n =$

<sup>1</sup> <https://dementia.talkbank.org/access/>

<sup>2</sup> [https://www.mhlw.go.jp/kokoro/speciality/detail\\_recog.html](https://www.mhlw.go.jp/kokoro/speciality/detail_recog.html)

<sup>3</sup> [https://www.jpn-geriat-soc.or.jp/tool/pdf/tool\\_02.pdf](https://www.jpn-geriat-soc.or.jp/tool/pdf/tool_02.pdf)

15) を機械学習により分類した。その結果、最大 0.93 の AUC 値が得られ、高い精度で分類できることを確認した [Tanaka 17]。Fraser らは認知症患者 (n=240) と健常者 (n=233) から構成されるコーパス (Dementia Bank) から、Language features と Acoustic features を抽出し機械学習による分類を行ったところ、81%を超える精度での分類を実現した [Fraser 16]。近年では深層学習を用いた手法が提案されており、Karlekar らは Dementia Bank を用いた認知症患者と健常者の分類を 91%の精度で達成したことを報告した [Karlekar 18]。また、コーパスに画像説明課題を用いて実験を行っている論文の内容について調査を行ったレビュー論文も報告されている [Mueller 18]。

健常者と軽度認知症者の識別を行う研究についても報告されている。Roark らは軽度認知障害 (n=37) と健常者 (n=37) から構成されるデータセットの書き起こしデータと音声データを用いて、Language features と Acoustic features を抽出し機械学習による分類を行ったところ、最大で 0.86 の Area under the curve (AUC) を達成した [Roark 11]。Fraser らは軽度認知障害 (n=37) と健常者 (n=37) の実験参加者に対して、自由発話でなく、画像についての説明課題 (Picture Description Task) を行った。得られた発話データから Language features を抽出し、識別を行った結果、最大で 0.87 の AUC 値を達成 [Fraser 18]した。Asgari らは軽度認知障害 (n=14) と健常者 (n=21) の実験参加者から取得した発話データ (daily chat) を LIWC を用いることで、HC と MCI を最大で 84%の精度で識別できることを報告した [Asgari 17]。

先行研究の多くは英語を対象とした研究であり、日本語での研究は少ないのが現状である。これは日本語で利用可能なデータセットが存在しないことが一因であると考えられる。そこで本研究では、日本語における認知症患者と健常者の発話データセットの構築とそれを用いた軽度認知症患者と健常高齢者の識別実験を行う。

## 4. 対照群付き高齢者コーパスの構築

対照群付き高齢者コーパス (以下、高齢者コーパス) は高齢者と若年者の実験参加者に対して節 4.3 から 4.5 節で述べる 3 つの課題を行い、その発話内容を書き起こしたコーパスと心理検査の結果から構成される。心理検査は 65 歳以上の実験参加者に対して、MMSE の測定を行った。

### 4.1 倫理的配慮

本研究の実施にあたっては「人を対象とする医学系研究に関する倫理指針」を順守し、任意参加であること、参加・不参加による不利益がないこと、いつでも参加の

取り消しができることなどを対象者および代諾者にわかりやすい言葉を用いて文書で説明した。

取得したデータは、第三者がアクセスできないように、パスワード設定、施錠等で適切に管理し、個人が特定されることがないように加工したデータ、または実験参加者から公開の許諾を得たデータを公表した。なお、本研究は奈良先端科学技術大学院大学の「研究科倫理審査委員会」の倫理審査の承認<sup>4</sup>を経て実施された。

### 4.2 実験参加者

実験参加者の基本的なデータを表 1 と表 2 に示す。

実験参加者のリクルート、実験内容の説明やデータの収集は株式会社アイアールアルトによって行われた。リクルートの際には男女比は 1:1 とし、可能な限り各年代で男女比が均等になるように調整した。発話データの収録は防音室にて行い、MMSE は臨床心理士などの訓練を受けた人間によって実施された。

本研究では 65 歳以上で MMSE が 28 点以上の実験参加者 (45 人) を健常高齢者、65 歳以上で MMSE が 23 点以上かつ 27 点以下の参加者 (15 人) を軽度認知症とし、それ以外の実験参加者を健常者とする。

表 1 実験参加者の詳細。括弧は標準偏差を示す。

|            | 全体          | 健常者         | 健常高齢者      | 軽度認知症      |
|------------|-------------|-------------|------------|------------|
| 参加者数       | 80          | 20          | 45         | 15         |
| 年齢         | 65.2 (16.3) | 39.6 (11.4) | 73.8 (4.4) | 73.5 (3.5) |
| 性別 (男性/女性) | 40/40       | 10/10       | 23/22      | 8/7        |
| 教育年数       | 14.2 (2.1)  | 15.1 (2.2)  | 14.0 (4.4) | 14.0 (1.7) |
| MMSE       | 21.3 (12.5) | -           | 29.3 (0.7) | 25.9 (1.0) |

表 2 実験参加者の年齢区分の詳細

| 年齢区分  | 参加者数 (男性/女性) | 平均年齢 (標準偏差) |
|-------|--------------|-------------|
| 20-29 | 5 (3/2)      | 24.6 (3.2)  |
| 30-39 | 5 (2/3)      | 36.0 (2.8)  |
| 40-49 | 5 (3/2)      | 44.0 (2.8)  |
| 50-59 | 5 (2/3)      | 54.0 (2.9)  |
| 60-69 | 10 (5/5)     | 67.1 (1.9)  |
| 70-79 | 50 (25/25)   | 75.1 (3.1)  |

### 4.3 画像説明課題 (PDT)

Picture Description Task (以下、PDT) は、画像に描かれている内容について説明を行うタスク [Goodglass 83]である。PDT の有名なデータセットとして Dementia Bank コーパスが存在する。これは TalkBank プロジェクト [MacWhinney 11]の一部であり、健常者と認知症患者 (MCI を含む) に対して PDT (画像は図 1 に示す Cookie theft picture) などを行ったもので、その発話データと書き起こしデータから構成されており、英語圏の先行研究では広く使用されている。しかし日本では普及

<sup>4</sup> 承認番号: 2017-I-20 (承認日: 平成 29 年 12 月 7 日)

していない。この理由としては Dementia Bank が英語のデータであり、日本語で広く利用可能なデータが存在しないことが一因であると考えられる。

本研究では、実施する PDT の画像には先行研究で使用されている Cookie theft picture を使用した。実験参加者に対して画像 (図 1) に描かれている内容について 1 分以上 2 分以内で具体的に話すように説明 (話す内容がない場合には 1 分以下でも可とした) し、説明を終える際には「以上です」と言うように指示した。実際の発話例を表 3 の PDT に示す。

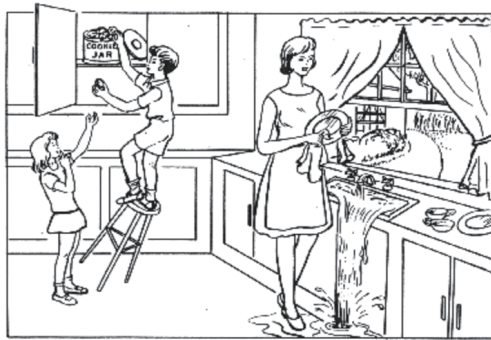


図 1 Cookie theft picture [Goodglass 83]

#### 4.4 エピソード説明課題 (EDT)

エピソード説明課題 (以下, EDT) は、以下に示す 8 つのテーマについて、参加者自身のエピソード記憶について 1 分以上 2 分以内で説明を行う課題である (話す内容がない場合には 1 分以下でも可とした)。本研究では「最近あった楽しい出来事」のみを用いた分析を行う。

1. 最近あった悲しい出来事
2. 最近不安を覚えた出来事
3. 最近怒りを覚えた出来事
4. 最近嫌悪感を覚えた出来事
5. 最近驚いた出来事
6. 最近あった楽しい出来事
7. あなたの尊敬する人
8. あなたが最近夢中になっていること

実験参加者にはテーマに沿ったエピソードについて具体的に話すように説明し、説明を終える際には「以上です」と言うように指示した。実際の発話例を表 3 の EDT に示す。

#### 4.5 アニメーション説明課題 (ADT)

前述の PDT における画像のかわりに、映像を用いる場合も存在する。映像を刺激とした場合のタスクは PDT との対比のため、便宜的に本研究では Animation Description Task (ADT) と呼ぶ。ADT は会話分析において用いられることがあり、the Pear Film [Chafe 80] やルーニー・テューンズなど著作権が消滅した映像作品の一部抜粋が頻用されている。

本研究では、再生時間がより短く (54 秒)、著作権が放棄された CG 作品である NAIST dog story を用いる。

ADT は、NAIST dog story を視聴し、起こった内容について 1 分以上 2 分以内で説明を行う課題である (話す内容がない場合には 1 分以下でも可とした)。NAIST dog story の時系列ごとのスクリーンショットを図 2 に示し、以下に各シーンの内容について述べる。

- ① 青い犬と赤い犬が散歩している
- ② 大きい未確認生物が小さい未確認生物を虐めている現場に犬が出くわす
- ③ 青い犬が大きい未確認生物を驚かせる
- ④ 助けられた小さい未確認生物は犬達に感謝し、青い犬の背中に乗って去っていく

実験参加者には動画の内容について具体的に話すように説明し、説明を終える際には「以上です」と言うように指示した。実際の発話例を表 3 の ADT に示す。

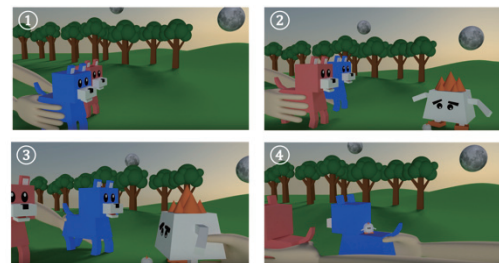


図 2 NAIST dog story

#### 4.6 データの利用

本研究により構築した対照群付き高齢者コーパスは、言語資源協会から音声データとして入手することが可能である<sup>5</sup>。

### 5. 機械学習による自動識別

機械学習モデル、評価方法、使用する特徴量と実験結果を以下に示す。

#### 5.1 モデルと評価方法

高齢者コーパスから健常高齢者 (Healthy Elderly: HE;  $n = 45$ ) と軽度認知症 (MCI;  $n = 15$ ) のデータを抽出し、それを用いた両者を自動識別するための機械学習モデルの構築と評価を行う。モデルには先行研究で多く使用されているサポートベクターマシン (SVM; 線形カーネル) とロジスティック回帰 (LR) を使用する。モデルの評価は 5-fold 交差検証を用いて、ROC (Receiver Operator Characteristic) 曲線を描画し、AUC (Area Under the Curve) の平均値を算出することで行う。本実験におけるモデルの実装、学習と評価は Python の機械学習ライブラリである Scikit-learn [Pedregosa 11] を用いて行った。また、本実験における SVM と LR のパラメーター値は Scikit-learn で設定されているデフォルト

<sup>5</sup> <https://www.gsk.or.jp/catalog/gsk2018-a>

表 3 各課題の発話例

|     |                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PDT | 台所の様子なんですけれども、お母さんと思われる方が洗い物をしているんですが、シンクから水があふれ出てしまっています。ただ、それにその女性は気づいていないというような状況に見えます。それから、子どもたちが男の子が踏み台に乗って棚の上にあるお菓子か何かを取ろうと。それで下にいる女の子に渡そうとしてるんですけれども、その踏み台がもう倒れそうになって、恐らくこのあと、この男の子は台から転げ落ちる。で、右手を入れ物にかけてるので、それも多分、落っことして下にぶちまけるようなことになるのではないかと思います。そのぶん、窓の外は何となく静かに流れていて平穏な感じですけど、家の中は大騒ぎになっているというような状況に見えます。以上です。                                                                      |
|     | お母さんがお皿を洗ってるんだけれども、水を流しながら、何を考えてるのか流しっ放しなので、水がシンクからあふれて床にこぼれております。だけど、視線はその方向に向かっていません。その一方で子どもは、戸棚にあるお菓子か何かを取り出そうと、長椅子の上に乗って倒れそうになってますが、妹はそれを助けようとしてないし、お母さんも振り向こうとしていません。外は何かよくわかりませんが、道路が見えて、農村のような気がします。大体そういうところでしょうか。以上です。                                                                                                                                                                |
| EDT | 身近なことばかりで申し訳なく思いますけれども、私の友人は、ニューヨークに住んでいまして、ごく最近の電話では 3 月から 4 月にかけて、こちらに日本のほうに戻ってくるという電話がありました。もうやはりなかなか会えないので、本当にそれは楽しみで、今からどうしようかいろいろと考えています。ニューヨーク東京の間は、飛行機に乗ればノンストップで来られるし、そんな遠いわけではないですけども、電話ではよく話しますがなかなか会うことができないので、これはもう今からとっても楽しいことで、待っているところです。以上です。                                                                                                                                  |
|     | 私は着物のほうのことをやっております、和の物が好きなんです。で、とある骨董市に行きまして、いろいろな物をいつも見るんですけども、その中でこういう柄行きの物が欲しいなというのを、ずっと探していたんですけども何軒目かのところで、それに近いようなものが出会えたのでそれがとても嬉しかったです。以上です。                                                                                                                                                                                                                                            |
| ADT | 青い、え、なんか、ブロックでできた、青い犬と茶色い犬が、犬を、あの、後ろから人が操っていて、一緒に、二、三歩進めていくと、え画面右の方から、あ、あ、あ、ば、場面が変わって、あなんか新しいキャラクターの、ちょ、ちょっと訳のわかんないのが出て来て、えっともう一人の人が、それを、えーっとなんか、ちい小さな…なんか白いもじゃもじゃの様なのをたたいて、えーっとい板の薄い板が曲がったようなやつで、たたいて・・・、んでまたその、その所に青い犬と茶色い犬が、二つ出て来て、えっと青い犬・・・と、その、へ、も一つの新しい変なキャラクターが、挨拶をして、それで、えっと、また、青い犬と茶色い犬・・・だけになって、その、白いもじゃもじゃのちっちゃなキャラクターと、挨拶をして、えー、それで、えっと白いもじゃもじゃのは消えて、青い犬と茶色い犬が二人で帰って行く、という感じだと思います。 |
|     | もういいですか？…ええっと、広い、草原が、がありまして、そこに、大きな木が、何本か、生えてます…その外には、その 1 群の林以外には、木は特になくて、え、そ空には、え晴れか曇りか分からないような、あー色ですが…その中に、月なのか、えーなんか丸いものがう、うー浮かんでいます…左の隅には、青い犬と、赤い犬が、居ますが、その犬がひょこひょここと出て来て…えーいました…ところが、そその、中に、えー途中で…将棋みたいなの、あー、出て来て、それが、生物なのかどうかのかはよく分かりません…それで、じゃれるかと思うと、別に遊んでる風でもなかったと思います、えーお菓子を、えもらったような感じがしましたけれども、それもよく、理解はできません、えー…、え、また、戻って行きました…まあそういうところ、終わります、15 秒や…お終わります…以上です                  |

ト値を使用した。

ROC 曲線はモデルの診断性能の特性を表現する方法で縦軸に偽陽性、横軸に真陽性の割合をプロットすることで描画される。モデルの検出がランダムである場合には 45 度の直線になり、診断性能が高いほど弓なりの曲線となる。この曲線下の面積の割合が AUC であり、診断性能が高いほど 1 に近い値となる [Lang 11]。

## 5.2 特徴量

特徴量の抽出においては、日本語形態素解析器 MeCab [Kudo 2004] を用いてコーパスの形態素解析を行った。抽出する形態素は品詞が名詞、動詞、形容詞と副詞であるものに限定した。

## 1 基礎特徴

基礎特徴として次の 6 つの指標を用いた。

- **Token**  
コーパス中の単語数を示す。
- **Type**  
コーパス中の異なり語数を示す。
- **Type Token Ratio (TTR)**  
Type と Token の比率 (Type/Token) で算出され、一般的にこの値が大きいほど語彙量大きい。
- **語彙難易度 (Lexical Educational Level: LEL)**  
語彙の難易度を示す指標である。難易度は日本語学習辞書<sup>6</sup>に記載されている語彙レベルを用いた。語

<sup>6</sup> <http://jhlee.sakura.ne.jp/JEV.html>

彙レベルは、1 (初級前半)、2 (初級後半)、3 (中級前半)、4 (中級後半)、5 (上級前半)、6 (上級後半) に分けられる。語ごとに算出し、全単語の平均を JEL スコアとする。

- 潜在語彙量 (Potential Vocabulary Size: PVS)  
潜在語彙量は、先行研究 [Aramaki 16a] によって示されるように、ある人物が無制限時間発話した際に使用する可能性のある語彙量の予測であり、対象者の用いた延べ単語数を計測し、それにジップ則を用いることで推測される。
- 品詞割合  
コーパス中の名詞、動詞、形容詞と副詞の使用割合を示す。コーパス中の名詞、動詞、形容詞、副詞の使用回数を算出し、Token 数で除すことで算出される。

## 2 意味密度

意味密度は、先行研究で認知機能との関連が報告されている指標であり、話者が発話した文がどの程度豊かな意味内容を伴っているかを測定するための指標である。意味密度の計算方法は、通常 1 つの文章に含まれる命題数を文章の異なり単語数で除した値とされる。意味密度の算出には専門家によるアノテーションが必要でありコストが大きい。そのため英語圏では意味密度の自動計算を研究がいくつか報告 [Brown 08, Sirts 17] されているが、日本語における意味密度の算出方法は長らく確立されていなかった。しかし、日本語コーパスを機械翻訳を用いて英語に翻訳し、英語における意味密度の算出方法である Dependency-based propositional idea density (DEPID) [Sirts 17] を適用することで、近似的な値を算出できることが報告されている [Shibata 18]。そのため本研究においても上記の手法を用いることで意味密度の算出を行う。機械翻訳には先行研究と同様に Google 翻訳<sup>7</sup>を使用する。また、先行研究で報告されている 3 つの指標 (D-ID, DR-ID, DRA-ID) の他に、意味密度の増減割合を示す指標である Density-Reduction を提案し、追加した。

- D-ID  
D-ID は DEPID (英語における意味密度の自動計算手法の一つ) で算出される意味密度で、コーパス中の命題数を単語数で除すことで算出される。

$$DID = \frac{\text{Number of Ideas}}{\text{Number of Tokens}}$$

- DR-ID  
認知症者の発話の特徴として、同じ内容を繰り返すことがある。同じ内容を繰り返すと命題数が増えてしまい、意味密度が高くなる。そこで、全ての命題数ではなく、延べ命題数を用いて意味密度の算出を

行うものが DEPID-R である。

DEPID-R により算出した命題数を DR-Ideas、算出される意味密度を DR-ID と表記し、以下に意味密度の算出方法を示す。

$$DRID = \frac{\text{Number of DR Ideas}}{\text{Number of Tokens}}$$

- DRA-ID  
DEPID-R に曖昧な文章に含まれる命題をカウントしない機能を追加したものが DEPID-R-ADD である。認知症者は、具体的な命題を含まない曖昧な文章を多く発話する特徴がある。そのため Speciteller により文章の具体性を算出し、具体性が低い曖昧な文章を除外する。先行研究と同様に Speciteller's specificity rating が 0.01 以下の文章に含まれる命題はカウントしない。

DEPID-R-ADD により算出した命題数を DRA-Idea と、算出される意味密度を DRA-ID と表記し、以下に意味密度の算出方法を示す。

$$DRAID = \frac{\text{Number of DRA Ideas}}{\text{Number of Tokens}}$$

- Density-Reduction1 (DR1)  
DR1 は DR-ID を D-ID で除した値であり、D-ID から DR-ID への変換時に減少した意味密度の値を示す指標である。
- Density-Reduction2 (DR2)  
DR2 は DRA-ID を DR-ID で除した値であり、DR-ID から DRA-ID への変換時に減少した意味密度の値を示す指標である。

基礎特徴と意味密度に加えて、単語ベクトル特徴である Bag of Words と分散表現を用いる。

- Bag of Words (BoW)  
BoW はコーパスの全体集合から、単語から成り立つ語彙を作成し、各コーパスでの各単語の出現頻度を含んだ特徴ベクトルを構築するモデルである。
- 分散表現  
分散表現はベクトルを密な固定長のベクトルに変換する手法であり、本研究では Mikolov らによって考案された Word2Vec [Mikolov 13] を使用する。また本研究では、鈴木らによって開発された学習済みの分散表現ベクトルである日本語 Wikipedia エンティティベクトル<sup>8</sup>を使用 [鈴木 16] し、コーパスから品詞が名詞である単語を全て抽出し、抽出した単語の分散表現ベクトルの平均値をそのコーパスの単語ベクトルとした。

<sup>7</sup> <https://translate.google.co.jp/?hl=ja>

<sup>8</sup> [http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki\\_vector/](http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/)

### 5.3 単一課題による識別

各課題から算出した特徴量を用いた健常者高齢者と軽度認知症者の自動識別実験を行った。その結果を表 4 に示す。PDT においては Word2Vec を特徴量とした SVM が最も精度が高いことが確認された。また、EDT においては基礎特徴と意味密度を合わせた特徴量である Language Ability Scores (LAS) を特徴量とした SVM が最も精度が高く、ADT では LAS と Word2Vec において LR での精度が最も高いことが確認された。

表 4 単一課題の実験結果

| Task | Feature  | SVM         | LR          |
|------|----------|-------------|-------------|
| PDT  | BoW      | 0.57        | 0.66        |
|      | Word2Vec | <b>0.74</b> | 0.69        |
|      | LAS      | 0.31        | 0.60        |
| EDT  | BoW      | 0.49        | 0.42        |
|      | Word2Vec | 0.40        | 0.44        |
|      | LAS      | <b>0.72</b> | 0.72        |
| ADT  | BoW      | 0.40        | 0.59        |
|      | Word2Vec | 0.59        | <b>0.72</b> |
|      | LAS      | 0.61        | <b>0.72</b> |

### 5.4 複数課題による識別

先行研究では 1 つのタスクでの識別精度を調査した研究が多い。しかし 1 つのタスクだけではそのときの体調や

気分などの影響を受ける可能性がある。例えば、緊張などでうまく話すことができず、通常能力を発揮することができないことが考えられる。そこで本研究では、タスクを組み合わせた場合の識別精度についても調査を行う。

PDT, EDT と ADT をそれぞれ組み合わせた場合の識別実験を行った。実験結果を表 5 に示す。Concat はそれぞれの課題の特徴量を結合したものを特徴量とした場合の実験結果を、Average はそれぞれの課題の特徴量の平均値を特徴量とした場合の実験結果を示す。また、PEA は全課題の組み合わせ、PE は PDA+EDA、PA は PDT+ADT の組み合わせ、AE は ADT+EDT の組み合わせである。全課題の LAS を結合した値を特徴量とした SVM の AUC が 0.85 と最も大きいことが確認された。

表 5 複数課題の実験結果。PEA は全タスクの組み合わせ、PE は PDA+EDA、PA は PDT+ADT の組み合わせ、AE は ADT+EDT の組み合わせを示す。また W2V は Word2Vec を示す。BoW (Average) は課題ごとに使用される単語が異なり、比較対象として適切ではないため、実験していない。

|               | PEA         |      | PE          |      | PA   |             | AE   |             |
|---------------|-------------|------|-------------|------|------|-------------|------|-------------|
|               | SVM         | LR   | SVM         | LR   | SVM  | LR          | SVM  | LR          |
| BoW (Concat)  | 0.44        | 0.59 | 0.59        | 0.62 | 0.59 | 0.76        | 0.45 | 0.46        |
| W2V (Concat)  | 0.70        | 0.76 | 0.63        | 0.64 | 0.74 | 0.72        | 0.69 | 0.65        |
| W2V (Average) | 0.50        | 0.69 | 0.51        | 0.73 | 0.75 | <b>0.81</b> | 0.41 | 0.55        |
| LAS (Concat)  | <b>0.85</b> | 0.71 | 0.76        | 0.78 | 0.64 | 0.66        | 0.71 | 0.80        |
| LAS (Average) | 0.73        | 0.83 | <b>0.83</b> | 0.74 | 0.55 | 0.71        | 0.72 | <b>0.84</b> |

## 6. 考察

実験結果の考察を以下に示す。

### 6.1 単一課題

PDT は特定の画像について説明を行うタスクであることから、それぞれの実験参加者が話す内容はある程度類似することが考えられる。そのため PDT において重要な特徴は使用している単語の種類ではないかと思われる。例えば、クッキーという単語はビスケット、お菓子など様々な単語で表現されるが、お菓子はビスケットに比べてより抽象的であり、単語の種類が一つの重要な特徴なのではないかと考えられる。そのためこのような表現の違いを考慮することができる Word2Vec において最も高い識別性能が得られたと考えられる。

EDT は個人のエピソード記憶について話すタスクであり、それぞれの実験参加者が話す内容は異なっていた。そのため単語ベクトルではうまく特徴を捉えることができない。そこで発話を統計的に扱える特徴である言語特徴 (TTR や意味密度) の識別性能が最も高いと考えられる。

ADT は特定の動画の内容について説明を行うタスクである。説明課題であると同時に、動画に登場するキャラクターが何であるかの判断は参加者の想像に依存する。例えば、ある参加者が「兄弟の犬」と表現したキャラクターを、別の参加者は「夫婦の犬」と表現した。そのため参加者ごとに使用する単語がある程度異なることが考えられ、ADT は PDT と EDT の 2 つの課題をマージした課題であると言える。そのため、Word2Vec と LAS で同程度の識別性能が得られたと考えられる。

### 6.2 複数課題

全課題の LAS を結合した値を特徴量としたときの識別性能が、PDT, EDT と ADT 各々の特徴量を入力した場合より最大で 0.11 上昇していることが確認された。

多くの先行研究では PDT のみ (単一課題) で実験を行

っていたが、ここに他の課題 (EDT や ADT) を加えて特徴量を作成すると精度が上がる可能性が示唆された。

1 つのタスクだけではそのときの体調や気分などの影響を受ける可能性がある。本研究において複数タスクの結果の平均値もしくは結合した値を特徴量とした場合の識別性能は単一タスクのそれより高いことが確認された。そのため識別性能の向上という面においては、複数のタスクを実施することが重要である可能性がある。

## 7. まとめ

本研究では、対照群付き高齢者コーパスの構築とそれを用いた認知症予備軍の自動識別に関する実験を行った。

これまで日本語において、認知症のスクリーニング研究に用いることができるコーパスが存在せず、研究開発の障壁となっていた。そこで本研究では、軽度認知症を含む高齢者と若年者の発話コーパスから構成される対照群付き高齢者コーパスの構築を行った。

そして得られたデータを用いて軽度認知症者と健常高齢者の自動識別について検討した結果、全ての課題で得られた特徴量の平均値を特徴量としたときの識別性能が最も高く、複数課題による識別性能向上の可能性を示すことができた。

しかし、本研究のサンプル数は 60 人と先行研究と比較した場合に少なく、信頼性に問題がある。そのため継続したデータ収集を行い、サンプル数を増やし、結果の信頼性をあげていくことが重要である。今後は得られた結果を他の精神疾患に応用していく予定である。

最後に、本コーパスにより日本語における認知機能と発話の関連を調査する研究が加速することを祈念している。

## 謝辞

本研究は、JST CREST (JPMJCR16E2), JSPS 新学術 (計画班 JP16H06395, D01 班 JP16H06399), JSPS 萌芽 JP16K12489 の支援を受けたものである。

## ◇ 参考文献 ◇

- [Aramaki 16a] Aramaki, E., Shikata, S., Miyabe, M., and Kinoshita, A.: Vocabulary size in speech may be an early indicator of cognitive impairment, *PLOS ONE*, Vol. 11, No. 5, pp. 1–13 (2016).
- [荒牧 16b] 荒牧英治, 若宮翔子, 四方朱子, 木下彩栄: 話の冗長性で アルツハイマー病をみつける, 人工知能学会全国大会論文集 2016 年度人工知能学会全国大会 (第 30 回) 論文集, pp. 4D11–4D11 (2016).
- [Asgari 17] Asgari, M., Kaye, J., and Dodge, H.: Predicting mild cognitive impairment from spontaneous spoken utterances, *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, Vol. 3, No. 2, pp. 219–228 (2017).
- [Becker 94] Becker, J. T., Boiler, F., Lopez, O. L., Saxton, J., and McGonigle, K. L.: The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis, *Archives of Neurology*, Vol. 51, No. 6, pp. 585–594 (1994).
- [Brown 08] Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., and Covington, M. A.: Automatic measurement of propositional idea density from part-of-speech tagging, *Behavior Research Methods*, Vol. 40, No. 2, pp. 540–545 (2008).
- [Chafe 80] Chafe, W. L.: *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, Ablex (1980).
- [Fraser 16] Fraser, K. C., Meltzer, J. A., and Rudzicz, F.: Linguistic features identify Alzheimer's disease in narrative speech, *Journal of Alzheimer's Disease*, Vol. 49, No. 2, pp. 407–422 (2016).
- [Fraser 18] Fraser, K. and Fors, K. L.: Improving the Sensitivity and Specificity of MCI Screening with Linguistic Information, in Kokkinakis, D. ed., *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France (2018), European Language Resources Association (ELRA).
- [Goodglass 83] Goodglass, H. and Kaplan, E.: *The Assessment of Aphasia and Related Disorders*, Lea & Febiger (1983).
- [Jarrold 10] Jarrold, W. L., Peintner, B., Yeh, E., Krasnow, R., Javitz, H. S., and Swan, G. E.: Language analytics for assessing brain health: Cognitive impairment, depression and pre-symptomatic Alzheimer's disease, in *International Conference on Brain Informatics*, pp. 299–307, Springer (2010).
- [Jarrold14] Jarrold, W., Peintner, B., Wilkins, D., Vergry, D., Richey, C., Gorno-Tempini, M. L., and Ogar, J.: Aided diagnosis of dementia type through computer-based analysis of spontaneous speech, in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 27–37 (2014).
- [Karlekar 18] Karlekar, S., Niu, T., and Bansal, M.: Detecting linguistic characteristics of Alzheimer's dementia by interpreting neural models, 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. 701–707, Association for Computational Linguistics (2018).
- [Kemper 01] Kemper, S., Thompson, M., and Marquis, J.: Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content, *Psychology and Aging*, Vol. 16, No. 4, p. 600 (2001).
- [Kudo 04] Kudo, T., Yamamoto, K., and Matsumoto, Y.: Applying conditional random fields to Japanese morphological analysis, in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (2004).
- [Mueller 18] Mueller, K.D., Hermann, B., Mecollari, J., Turkstra, L.S. (2018) Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks, *Journal of Clinical and Experimental Neuropsychology*, Vol. 40, No. 9, pp. 917–939, DOI: 10.1080/13803395.2018.1446513.
- [MacWhinney 11] MacWhinney, B., Fromm, D., Forbes, M., and Holland, A.: AphasiaBank: Methods for studying discourse, *Aphasiology*, Vol. 25, No. 11, pp. 1286–1307 (2011).
- [Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J.: Distributed representations of words and phrases and their compositionality, in *Advances in Neural Information Processing Systems*, pp. 3111–3119 (2013).

- [Orimaye 14] Orimaye, S. O., Wong, J. S.-M., and Golden, K. J.: Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances, in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 78–87 (2014).
- [Pedregosa11] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V., et al.: Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, Vol. 12, No. Oct, pp. 2825–2830 (2011).
- [Roark 11] Roark, B., Mitchell, M., Hosom, J.-P., Hollingshead, K., and Kaye, J.: Spoken language derived measures for detecting mild cognitive impairment, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 19, No. 7, pp. 2081–2090 (2011).
- [Shibata 16] Shibata, D., Wakamiya, S., Kinoshita, A., and Aramaki, E.: Detecting Japanese patients with Alzheimer's disease based on word category frequencies, in *Proceedings of the Clinical Natural Language Processing Workshop (ClinicalNLP)*, pp. 78–85 (2016).
- [Shibata 18] Shibata, D., Ito, K., Nagai, H., Okahisa, T., Kinoshita, A., and Aramaki, E.: Idea density in Japanese for the early detection of dementia based on narrative speech, *PLOS ONE*, Vol. 13, No. 12, pp. 1–12 (2018).
- [Sirts 17] Sirts, K., Piguet, O., and Johnson, M.: Idea density for pre- dicting Alzheimer's disease from transcribed speech, in *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pp. 322–332, Association for Computational Linguistics (2017).
- [Snowdon 96] Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., and Markesbery, W. R.: Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study, *Journal of American Medical Association*, Vol. 275, No. 7, pp. 528–532 (1996).
- [鈴木 16] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎: Wikipedia 記事に対する拡張固有表現ラベルの多重付与, 言語 処理学会第 22 回年次大会, pp. 797–800 (2016).
- [Tanaka 17] Tanaka, H., Adachi, H., Ukita, N., Ikeda, M., Kazui, H., Kudo, T., and Nakamura, S.: Detecting dementia through interactive computer avatars, *IEEE Journal of Translational Engineering in Health and Medicine*, Vol. 5, pp. 1–11 (2017).
- [Lang 11] Lang, T. A., Secic, M., 大橋晴男, 林健一訳, わかりやすい医療統計の報告-医学論文作成のためのガイドライン 第2版, 中山書店, 2011.

## 著 者 紹 介

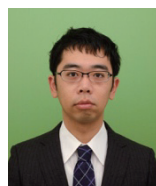
### 柴田 大作

2018 年奈良先端科学技術大学院大学情報学研究科博士前期課程修了。現在、同大学先端科学技術研究科博士後期課程に在籍。医療情報学会会員。



### 伊藤 薫

九州大学言語文化研究院助教。博士（人間・環境学）。2012 年京都大学大学院人間・環境学研究科修士課程修了。2015 年京都大学大学院人間・環境学研究科博士後期課程研究指導認定退学。2016 年奈良先端科学技術大学院大学博士研究員。2019 年より現職。言語学，自然言語処理に関する研究に従事。専門は認知言語学，談話・テキスト言語学。言語処理学会，日本認知言語学会，日本語用論学会各会員。



### 若宮 翔子

2013 年兵庫県立大学大学院環境人間学研究科博士後期課程修了。同年，日本学術振興会特別研究員 PD。2014 年京都産業大学コンピュータ理工学部研究員。2015 年奈良先端科学技術大学院大学博士研究員，2018 年より同大学特任助教，現在に至る。博士（環境人間学）。主にソーシャルメディア分析とその応用に関する研究に従事。情報処理学会，日本データベース学会各会員。



### 荒牧 英治（正会員）

奈良先端科学技術大学院大学特任准教授。博士(情報理工学)。2005 年東京大学大学院情報理工学系研究科修了。2005 年東京大学医学部附属病院特任助教，2008 年東京大学知の構造化センター特任講師，2011 年京都大学デザイン学ユニット特定准教授を経て 2015 年より現職。医療情報学，自然言語処理の研究に従事。情報処理学会，言語処理学会，日本認知科学会，医療情報学会等各会員。



〔担当委員：平 博順〕

2019 年 1 月 9 日 受理