

令和元年6月6日現在

機関番号：14603

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K12489

研究課題名(和文) 自然言語処理解析による認知症の低侵襲・早期発見に関する研究

研究課題名(英文) Using Natural Language Processing of Patient Short Narratives to Detect Cognitive Impairment

研究代表者

荒牧 英治 (ARAMAKI, Eiji)

奈良先端科学技術大学院大学・研究推進機構・特任准教授

研究者番号：70401073

交付決定額(研究期間全体)：(直接経費) 2,500,000円

研究成果の概要(和文)：研究では、言語関連疾患の患者の語り(10分程度の自由音声発話)を材料として、自然言語処理技術による患者の診断、将来的なリスクの予測を実現し、最終的にはリハビリ/治療という革新的な医療の可能性に挑む。言語関連疾患としては、語彙量との関連について、すでにエビデンスが存在するアルツハイマー型認知症と、当事者の自伝的語りから治療において重要視されている発達障害(自閉スペクトラム症；以降ASD)を主な2対象として扱い、加えてうつ病、統合失調症など、語りとの関連が深い疾患についても適応を広げていく。

研究成果の学術的意義や社会的意義

現時点で対象としているアルツハイマー型認知症のみならず、本研究により得られる知見や成果は、言語と関係した他の疾患についても有効である可能性が高い。本研究によって測定法が確立されることで、これら他の障害や疾患への応用も可能となり、大きな波及効果が期待される。

加えて、本研究の方法論である「患者の語りを引き出す」という行為そのものが、社会を変える可能性がある。本来口を閉ざしがちな認知症やASD者の語りを引き出し、その土台となる場を生み出すことによって、これらの疾患への社会的理解を促進し、結果的に疾患への社会的無理解からくる偏見や診断への抵抗感を大きく軽減できる可能性がある。

研究成果の概要(英文)：This study was conducted to ascertain whether a Natural Language Processing (NLP) based system can detect dementia or not. We obtained materials of two types from dementia patients: (1) speech samples, and (2) cognitive function level test scores from Hasegawa's Dementia Scale; Revised (HDS-R), a test used in Japan for dementia patients. Then, we investigated that the language based index is related to the level of dementia.

研究分野：自然言語処理

キーワード：自然言語処理 医療情報学 アルツハイマー病

1. 研究開始当初の背景

世界に類をみない超高齢化社会を迎える本邦にとって、高齢者への医療対策は重要な課題である。とりわけ、認知症は、その予備軍(MCI)も含めると4人に1人の割合となり、その医療費は10兆円規模と算定され、医療崩壊が危惧されている(厚生労働省)。認知症の対策には、治療法の確立もさることながら、一方で早急に症状を発見し、その進行を遅らせることで、健康な期間を延長し、介護が必要となる期間を短縮することも重要である。

早期に症状を発見する研究については、近年、血液検査や精密記憶検査など新たな方法が提案されつつあるが、多くは身体的侵襲、精神的負担、または、その両方を伴う。より簡便かつ負担のない検査として、欧米では記述した文章から認知症を予測する研究が注目を集めている。

これに対して、本邦においては、認知症と言語能力の関係を大規模に調べた研究は未だ存在しない。欧米の知見を導入しようにも、日本語は、言語体系が英語などの欧米の言語と大きく異なることから、日本語を母語とする話者の、言語能力と認知能力との関連を確認することが必要とされている。

以上のような研究開始当初の背景のもと、研究代表者は研究開始以前から、京都の特別養護老人ホームにて高齢者の語りを収集して解析を続けている。

この結果、14名の高齢者(認知症者予備軍8名を含む)を用いた予備実験結果では3~7分の発話において語彙量が認知症の検査値(HDS-R)と有意に相関する(相関係数0.47)との結果を得た。研究代表者は、これまでも、ブログ文章において、症状の進行とともに語彙能力が低下するとの結果を得ている(図1)。さらに、実用化の観点からも、語彙量は、音声認識ミスにより、単語を誤って別の単語と認識してしまった場合でも、同じ間違いが続く限りは正確な算出が可能であり、音声認識誤りに対して頑健であることを実証している。これら3つの結果を総合し、我々は、自動音声認識による認知症予備軍の早期検出が可能だと考えている。

2. 研究の目的

世界に類を見ない超高齢化社会を迎える日本にとって、高齢者への医療対策は重要な課題である。とりわけ、認知症は、その予備軍も含めると4人に1人の割合となり、その医療費は10兆円規模と算定され、医療崩壊が危惧されている(厚生労働省)。この危機をまっさきに迎えるのが、すでに高齢者率20%を超えようとする京都である。本研究は、京都のケア施設と京大病院をフィールドとして、自由音声発話を収集・コーパス化する。さらに、語彙量を測定するアルゴリズムを開発し、語彙量から認知症予備軍の早期発見手法を確立する。認知症患者の語彙や文法能力が低下することは、海外では、すでに明らかになっているが、日本語での測定方法は確立していない。本研究は、これを確立し、臨床的に効果を示すことまでを行う挑戦的な研究である。

3. 研究の方法

認知症患者の日本語の語りから 相関係数 0.50 以上で認知症の障害の度合いを推定する技術を確認する。このために必要となる認知症患者 30 名の自由発話音声データを医療施設病院にて収集する。また、データの書き起こしも行い、語彙量を始めた言語指標と症状レベルとの

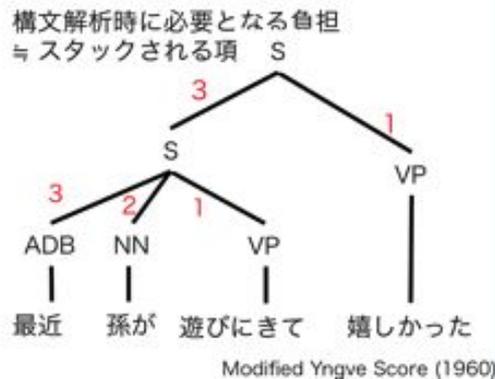


図1: 構文解析時に必要なスタック数( Modified Yngve Score ) は言語能力と相関する。このように様々な言語指標と、疾患との関連を調査する。

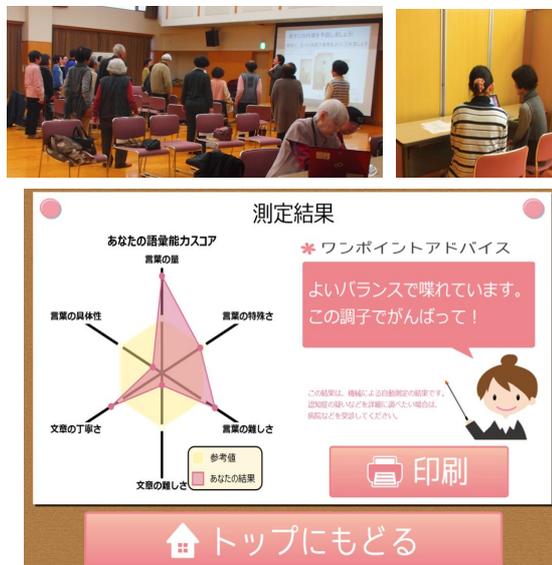


図2: 京都府の特別養護老人ホーム「修徳」において、言語による認知症測定を行っている様子(図上)。3~7つの質問が画面に提示され、答えた内容を音声認識し、リアルタイムに言語能力を提示する(図)。語られる内容に注目するのではなく、語彙数などの単語統計量を注目しており、現状の音声認識でも測定可能である。

関連を調査する。語彙量推定のアルゴリズムは、最も単純な Type / Token 比から、潜在語彙量推定まで、様々な指標がある。これらのうちもっとも症状レベルと相関するものを探索する。症状レベルは世界的に認知症の進行をスコア化した FAST を用いる。また、語彙量だけでなく、必要に応じて構文の複雑さ（図 1）や語の抽象性など様々な指標を探索する。

#### 4. 研究成果

本研究の結果、自由発話の音声から認知症の推定が一定の精度で実現できることが証明された。この成果は、ケアハウス（図 2）などの現場にて試験利用していただいている。しかし、臨床的な実用レベルには至らず、自由発話のみで、認知症を診断できるというレベルには至っていない。そこで、今後の課題としては、自由発話ではなく、ある程度固定した質問を用いる、音声特徴を利用する、経時的な測定などが挙げられる。これらについて継続して、研究していく予定である。

また、最終年度では、認知症だけでなく、がん患者における認知機能の低下など、さまざまな要因における認知機能の低下に本技術を適応する可能性も検討した。がん患者 100 名を超える実験（現在も進行中；雑誌論文 1）では、Type / Token 比によって、がん患者の認知機能の低下を捉えることができることを示した。

今後は、企業との共同研究を通じて、実証フェーズに移行し、研究を継続する予定である。さらに、本研究の成果物として、語の抽象性を計算するためのリソース（日本語抽象度辞書 AWD-J <http://sociocom.jp/data/2019-AWD-J/>）や、iPhone アプリ（語彙量チェッカー）を開発、公開してきた。これらは、研究代表者が参画している新学術研究などの他大学との共同研究にも利用している。本研究により言語処理技術が他分野に波及する可能性の一端を示せたと考えている。

最後に、現時点で対象としているアルツハイマー型認知症、ASD やうつ病のみならず、本研究により得られる知見や成果は、言語と関係した他の疾患についても有効である可能性が高い。例えば、学習障害や失字症などについても、自己モニタリングや早期発見が必要であるとされており、本研究で扱う障害や疾患と極めて似た状況にある。本研究によって測定法が確立されることで、これら他の障害や疾患への応用も可能となり、大きな波及効果が期待される。

加えて、本研究の方法論である「患者の語りを引き出す」という行為そのものが、社会を変える可能性がある。本来口を閉ざしがちである患者の語りを引き出し、その土台となる場を生み出すことによって、これらの疾患への社会的理解を促進し、結果的に疾患への社会的無理解からくる偏見や診断への抵抗感を大きく軽減できる可能性がある。

#### 5. 主な発表論文等

〔雑誌論文〕(計 3 件)

1. **Eiji Arameaki**, Mai Miyabe, Chihiro Honda, Seiko Isozaki, Shoko Wakamiya, Akira Sato, Isao Miyashiro: KOTOBAKARI Study: Using Natural Language Processing of Patient Short Narratives to Detect Cancer Related Cognitive Impairment, MEDINFO 2019: eHealth-enabled Health (採録済；DOI 未決定) (査読有り).
2. Daisaku Shibata, Kaoru Ito, Hiroyuki Nagai, Taro Okahisa, **Ayae Kinoshita**, **Eiji Aramaki**: Idea density in Japanese for the early detection of dementia based on narrative, PLoS One, Volume 13 (Number 12), pp. e0208418, 2018 (査読有り). (<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0208418>)
3. 柴田大作, 若宮翔子, **木下彩栄**, **荒牧英治**: 音声発話による認知症スクリーニング技術の開発, 医療情報学, Volume 37 (Number 6), pp. 303-312, 2018(査読有り) (<https://doi.org/10.14948/jami.37.303>).

〔学会発表〕(計 2 件)

1. **Eiji Aramaki**, Chihiro Honda, Shoko Wakamiya, Akira Sato, Isao Miyashiro: Quick Cognitive Impairment Test for Cancer Patients using Emotional Stroop Effect, MedInfo 2019 (Poster) (France, Lion), 2019 (2019/03/14).
2. Daisaku Shibata, Shoko Wakamiya, Kaoru Ito, Mai Miyabe, **Ayae Kinoshita**, **Eiji Aramaki**: VocabChecker: Measuring Language Abilities for Detecting Early Stage Dementia, In Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion (IUI Companion) (Tokyo, NII), 2018 (2018/03/10).

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

日本語抽象度辞書 AWD-J [http://sociocom.jp/ data/2019-AWD-J/](http://sociocom.jp/data/2019-AWD-J/)  
ソーシャル・コンピューティング研究室 <http://sociocom.jp/>  
医療言語処理グループ <http://mednlp.jp/>

## 6 . 研究組織

### (1)研究分担者

研究分担者氏名：木下彩栄

ローマ字氏名：KINOSHITA, Ayae

所属研究機関名：京都大学

部局名：医学研究科・医学部

職名：教授

研究者番号（8桁）：80321610

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。