

## 3分ほどの音声発話で認知症者は見つかるか？ 自動音声データの言語処理による認知症スクリーニングの試み

荒牧英治<sup>1,2</sup>, 四方朱子<sup>1</sup>, 宮部真衣<sup>3</sup>, 野田泰葉<sup>4</sup>, 木下彩栄<sup>4</sup>

【背景】認知症の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、その多くは、身体的侵襲、精神的な負担、または、その両方を伴う。そこで、認知症の症状である記憶能力や言語能力の低下を調査する研究が注目されている。【目的】本研究は、自然言語処理技術を用いて、認知症患者の発話を定量化し、認知症との関係を明らかにする。【材料と手法】HDS-RとMe-CDTテストにより認知症疑いがある高齢者(4名)と正常な高齢者(10名)の語りに対して、語彙の統計量を調査した。【結果と考察】語彙の難解さ(AUC 0.712)と特殊性(AUC 0.711)が認知症との関連を示し、音声発話から、認知症をスクリーニングできる可能性が示された。今後、より大規模な症例にて検討が望まれる。

### All It Takes Is 3 Minutes of Talking for Finding Dementia: A Challenge for a Simple Dementia Screening of Natural Language

Eiji ARAMAKI<sup>1,2</sup> Shuko SHIKATA<sup>1</sup> Mai MIYABE<sup>3</sup>  
Yasuha Noda<sup>4</sup> Ayae KINOSHITA<sup>4</sup>

[Background] Several methods have been established for the early detection of dementia, such as blood testing. However, such methods are often quite invasive both physically and mentally, and less invasive methods that may detect the deterioration in cognitive functions, including language abilities, of the dementia patients are awaited. [Objective] This report challenges to quantify the speech of dementia patients and to examine the relationship between dementia and language ability by using natural language processing technique. [Materials & Methods] We have examined the language abilities of the speeches of healthy elders and possible dementia individuals (derived by both HDS-R scores and Me-CDT scores). [Results & Discussion] The language difficulty (AUC 0.712) and uncommon words usage (AUC 0.711) showed the relationship with dementia. It may indicate the possibility of screening dementia by analyzing the speech of elders.

#### 1. はじめに

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人の割合と言われており、その医療費は10兆円規模と算定されている(平成22年厚生労働省)。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。

また、近年、さまざまな介入による予防に対する関心も高まりをみせ、1,000超規模の臨床試験も世界各地で開始されつつ

ある: FINGER<sup>a</sup>, HATICE<sup>b</sup>, DO-HEALTH<sup>c</sup>。近い将来、これらのいずれかが、認知症を予防、もしくは進行抑制できる可能性を示せば、当然のこととして、次に、いかに認知症を早期発見するかが次の課題となるであろう。

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、その多くは、身体的侵襲、精神的な負担、または、その両方を伴う。そこで、認知症の症状である記憶能力や言語能力の低下を調査する研究が注目されている。記憶において、すでに多くの研究がある [1, 2]。言語においても、コミュニケーション能力 [3]、文理解 [4]、失語 [5, 6]、話し言葉の変化 [7]、カテゴリ化 [8, 9]、単語想起 [10]、文の追唱 [11]など多くの研究がある。近年は、診断の補助的手段としての可能性も検討されはじめ [12]、さら

†1 京都大学学際融合教育研究推進センター

〒600-8815 京都府京都市下京区中堂寺栗田町 93 KRP6 号館 202 号室

†2 JST さきがけ「自然言語処理による診断支援技術の開発」

〒600-8815 京都府京都市下京区中堂寺栗田町 93 KRP6 号館 202 号室

†3 和歌山大学システム工学部

〒640-8510 和歌山県和歌山市栄谷 930

†4 京都大学医学系研究科人間健康科学系専攻

〒606-8507 京都府京都市左京区聖護院川原町 53

a

<http://ki-su-arc.se/longitudinal-studies/finnish-geriatric-intervention-study-to-prevent-cognitive-impairment-and-disability-finger/>

b <http://www.hatice.eu/>

c <http://do-health.eu/wordpress/>

には、診断のみならず、予防やリハビリへの応用可能性も調査されている [13]. このようにして、言語能力との関連が徐々に明らかになりつつある [14].

しかし、言語能力の多様な側面のうち、先行研究で着目されている言語能力には偏りがある。最もよく研究されているものは、語彙に関する能力であり、例えば、Letter Fluency (ある文字から始まる単語を羅列する課題) や Category Fluency (<動物>や<野菜>などあるカテゴリに属するものを可能な限り羅列する課題)、Boston Naming Test [15] や Graded Naming Test [16] といった課題によって調査される [14]. より高度な課題としては、word-picture matching [17] や、指示に従った操作をする token test [18]、語の定義の述べる課題 [19] 等がある。これらは、いずれも認知症との関連が認められるが、実験室環境と、一定の検査時間が必要となる。本来、言葉は日常的に使うものであることを考慮すれば、より簡便に検査できる可能性があると考えられる。さらに、簡便な検査であればあるほど、将来的に、スクリーニングなどの臨床応用を考えれば、患者負担が少なく望ましい。

このような背景の中、他の研究と比較し劇的に簡便な検査の可能性を提示したのが Nun Study [20] である。この研究では、若年時に記述した日記にみられる文の複雑さや語彙の特徴から将来的な認知症の傾向を予測できるとされ [21]、現在にいたるまで、Kemper を中心とした研究グループによる測定法などの緻密化に関する研究が継続している [22-24].

これに対して、本邦においては、認知症と日本語能力の関係を大規模に機械的に調べた研究は存在しない。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、日本語を測定する仕組みが必要となる。

そこで、本研究では、日本語における言語と認知症との関連を調べる。さらに、Nun Study より、さらに検査を簡便にするため、話し言葉を自動音声認識し、そこから認知症を診断することを試みる。本研究は、認知症における日本語能力を機械的に調べた初めての研究であるとともに、音声認識で認知症患者の発話を解析する点でも世界初の試みである。

## 2. 関連研究

本研究は、言語処理が、臨床場面への応用を目指す研究である。そこで、言語処理の医療応用例から (2.1 章)、自然言語処理を利用した言語能力測定 (2.2 章) について以下に関連研究をまとめる。

### 2.1 医療分野における言語処理の動向

医療における言語処理への関心は高く、世界中でワークショップが開催されている。例えば、アメリカにおける i2b2 NLP

workshops [25]、EU における CLEF e-health workshops [26]、日本における NTCIR MedNLP [27] 等、この 10 年ですべて立ち上がったものである。これらのワークショップでは、主にカルテ文章が対象となっているが、患者が記述した文章も扱われはじめている。患者文章の扱い方には以下の 2通りのアプローチがある [28].

1つは、患者データのある種のビッグデータとみなして、そこから情報抽出を行う集合知アプローチである。インフルエンザ流行 [29, 30]、West Nile ウィルス検出 [31]、薬剤副作用検出 [32] といった疾患特有のものから、疾患を限定しない汎用システム (BioCaster [33]、EpiSPIDER [34]、HealthMap [35, 36]) など多くの研究がある。

もう1つのアプローチは、各個人に焦点を当て、その個人のみ の身体状態や精神状態を推定するアプローチである。前者に比べて、後者のアプローチは数こそ少ないが、自殺者の日記から著者の感情を推定するもの [37] や、認知症予備軍 (mild cognitive impairment) を推定するもの [38]、失語症の診断 [39]、前頭側頭葉変性症の解析 [40, 41]、自閉スペクトラム症の解析 [42] など臨床応用に直結する研究が多い。これらの研究では、患者の言語能力が様々な観点から測定されている [43].

### 2.2 言語能力の測定

#### 構文能力

言語処理を用いて、構文の複雑さを測ることができる。英語の構文能力についてはすでに確立した尺度がある。D-Level [44] は、文の複雑さを 8 つのレベルに分割した尺度である。D-level は、すでに認知症の測定に用いられており [45]、同じく認知症を対象とする本研究に適する可能性のある尺度のうちの 1 つであるが、現時点では英語のみに対応しており、日本語版は存在しない。他にも Frazier score [46]、依存構造の係り受け距離 [47]、Yngve score [48] など多くの構文測定尺度があるが、そもそも、日本語に対応していないか、あるいは、句構造解析器、または、依存構造解析器を必要とし、話し言葉への導入が困難である。

#### 語彙能力

語彙能力に関しても、いくつかの尺度が提案されている。これらには、語彙の質に関するものと、量に関するものの 2通りがある。

- **語彙の質**: 語彙の質は、用いられる語彙の難易度と考えることができる。語彙の難易度は、日本での学校教育や外国語教育に用いられる語彙レベル辞書などを用いて測ることができる。

- **語彙の量**: 語彙の量は、理解できる語彙の量や使用する語彙の量を計測する。しかし、これを正確に測定することは難しい [49-51]。最も素朴な方法は、任意の単語を知っているかどうかを大量に問うことである。この方式は大量の時間がかかる[52]。そこで、一定語数に存在する語彙量を調べるタイプトークン比が代替として用いられることが多い [53] [54]。

### 2.3 先行研究における認知症における言語能力調査との差異

本研究と先行研究との差異は語の特殊性など新しい指標を使っていること、及び、先行研究が主に書き言葉を扱っていたのに対し、本研究は音声認識結果の解析を行う点が新しい。作文を毎日行う高齢者は稀であるが、音声による発話を行う高齢者は圧倒的に多く、収集が容易であることを考えれば、音声による予測の利用可能性は高い。

## 3. 材料

### 3.1 フィールド

本研究のフィールド特別養護老人ホーム「修徳」である。「修徳」とは、2001年に下京区元修徳小学校跡地に開設された複合施設であり、特別養護老人ホーム、デイサービスセンター、在宅介護支援センター、児童館から構成され、京都府指定事業者としてサービス提供を行っている。規模は、介護老人福祉施設（特別養護老人ホーム）定員 80 名、短期入所生活介護・介護予防短期入所生活介護定員 20 名、通所介護・介護予防通所介護定員 30 名である。この施設では、関連施設及び近隣の市民への啓蒙活動のために、「認知症あんしん相談会」というイベントを開催している。実験場所は、この「認知症あんしん相談会 第三回」にて行い、参加者のリクルートは、配布物とポスターにて行った。

このイベントの参加者は、会場に入って、認知症相談コーナー、2つの簡易検査コーナー、言語使用調査を受ける。すべてを受ける必要はなく、好きなコーナーを見て回ることができる。簡易検査では、最も広く使用されているファルメディオ社の iPad 対応の長谷川式簡易知能評価スケール（以降、HDS-R）による検査および、ヤンセンファーマ社の Windows 対応ソフト Me-CDT[55]による検査を実施した。今回は簡易検査コーナーおよび言語使用調査の双方のデータを提供された計 14 名（男性 4 名、女性 10 名、平均年齢 80.3 歳）について分析する。詳細は表 1 参照のこと。

### 3.2 長谷川式簡易知能評価スケール (HDS-R)

HDS-R は計算能力や、近時記憶を問う 9 問からなる試験で、およそ 10 分で認知症を判定する。30 点満点であり、20 点以下だと認知症疑いとしている。今回は最も広く使われているファルメディオ社の iPad 対応の無料アプリを用いて測定した [56]。

### 3.3 MeCDT

Me-CDT は、パソコン・ソフトによる音声ガイド付き検査で、被験者が、紙とペンで、パソコンから再生される指示にもとづき、時

表 1: データ提供者背景.

性別	男性 4 名, 女性 10 名
年齢	平均 80.3 歳 (64-92; SD 7.67)

表 2: 長谷川式簡易知能評価スケールの質問.

お歳はいくつですか? (2年までの誤差は正解)
今日は何年何月何日ですか? 何曜日ですか?
私たちがいまいるところはどこですか?
これから言う3つの言葉を言ってみてください。あとでまた聞きますのでよく覚えておいてください。
100から7を順番に引いてください。
私がこれから言う数字を逆から言ってください
先ほど覚えてもらった言葉をもう一度言ってみてください。
これから5つの品物を見せます。それを隠しますのでなにがあったか言ってください。
知っている野菜の名前をできるだけ多く言ってください。

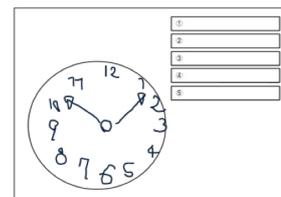


図 1: Me-CDT の検査用紙.

以下の6つの質問からなる。①「氏名」②「日付」③「今何階にいるか」④「最近気になったニュースはなにか」⑤「冒頭に提示した時間は何時か」⑥「時計の文字盤と⑤の時刻(10時10分)の針描写」



図 2: 収録光景

計描画テスト、近時記憶など 6 問からなる質問を 3 分程度で答え、それを人手で評価する指標である(図 1)。満点は 10 点であり、8.5 点以下の認知症疑いとする。

### 3.4 手続き

実験は、以下の手続きにそった。まず、実験協力者は、席に着席し、タッチパネル方式の入力で、年齢、性別を入力する。操作が分からない場合は、案内者が教示を行う。

次に、画面に質問が表示され、はじめてくださいとの指示のもと、1分前後を目安に発話を行う。ただし、今回は被験者へのプレッシャーや焦りを軽減するため、時間による強制的な発話量や発話時間の設定は下限、上限共に無く、発話の量や長さは任意である。ただし、一定の目安として、時間経過は画面に表示される。この質問-発話のプロセスは3回または7回行われる。質問は固定されており、以下の順で質問される。(てばやく測定モードは以下の1-3の質問、じっくり測定モードは以下の1-7すべての質問が順に表示される。)

1. 「今日はどうやってここまで来ましたか？できるだけ詳しく教えてください。」
2. 「最近、嬉しかったことは何ですか？」
3. 「子供の頃、どんな遊びをしましたか？その遊びの名前や遊び方などを教えてください」
4. 「あなたの趣味についてできるだけ詳しく教えてください」
5. 「今までで一番嬉しかったことは何ですか？」
6. 「小さいころ、何になりたかったですか？それはなぜですか？」
7. 「今一番したいこと、してもらいたいことは何ですか？」

音声認識、言語解析が行われる。音声認識には、Juliusを利用した。言語解析が終わると、結果がプリントアウトされ、実験は終了となる。実際の様子を図2に示す。

## 4. 手法

言語解析には、以下の尺度を用いた。

- 発話語数 (TOKEN)
- 語彙教育難易度 (Education Level; EL)
- 発話語数・発話種類数比 (Type Token Ratio; TTR)
- 固有名詞割合 (Named Entity Ratio; NER)
- 頻度・使用者数比 (Frequency per User Popularity; FPU)

### 4.1 TOKEN

発話した語数を集計した。単語の区切りには、日本語形態素解析器[57]を用いた。

### 4.2 Education Level (EL)

語彙の難易度を示す。難易度のスコアは、日本語学習辞書を用いた。日本語学習辞書には、主な日本語 17,928 語について、初級、中級、上級との分類がなされている[58]。この内の名詞を対象に、中級以上の割合を算出し、スコア化した：

$$EL = \frac{\text{中級以上の名詞数}}{\text{名詞数}}$$

### 4.3 Type Token Ratio

Type(異なり語数)とToken(延べ語数)の比率(Type/Token)であり、この値が大きいほど語彙量が多いと考えられる。

表 3: 言語指標と HDS-R, Me-CDT との相関係数。

	TOKEN*	EL	TTR	NER	FPU	CDT*
HDS-R	0.650	<u>0.712</u>	0.650	<u>0.775</u>	<u>0.711</u>	0.950
Me-CDT	0.510	<u>0.744</u>	0.555	0.640	<u>0.711</u>	-

\*は低値の場合 MCI と判定する指標を示す(これら以外は高値を MCI と判定する)。

$$TTR = \frac{\text{types 数}}{\text{Token 数}}$$

### 4.4 Named Entity Ratio

固有名詞の割合を示す。

$$NER = \frac{\text{固有名形態素数}}{\text{名詞形態素数}}$$

この値が大きければ、文章の内容がより具体的であることを示す。固有名詞の判定は、形態素解析器 JUMAN[57]の解析結果による。地名、数詞、固有名詞の割合を文ごとに算出し、平均した。

### 4.5 Frequency per User Popularity

語の出現頻度/語のユーザ数の文の平均値。

$$FPU = \sum_{w \in W} \frac{\text{語}w\text{の出現頻度数}}{\text{語}w\text{のユーザ数}}$$

この値が大きいほど、ユーザ数が出現頻度と比較し、少ない語であり特殊な語を多く用いていることを示す。例えば、スラングや専門用語は高い値を持つ。逆に、値が大きいと、一般的な語であることを示す。詳細は文献[59]を参照のこと。

## 5. 実験

### 5.1 デザイン

HDS-R と Me-CDT によって、認知症疑いとなった群を推定した。ただし、前述したように HDS-R は 30 点満点中 20 点以下だと認知症の可能性があるが、本サンプルで 20 点以下である群は 2 名しか存在しない。そこで、MCI(軽度認知障害)が疑われる 26 点以下の対象(4名)の識別を行った。また、Me-CDT 課題についても、カットオフ値 8.5 点以下の 5 名の識別を行った。評価は、ROC-AUC(以降 AUC)による。なお、一般に、AUC は 0.7 を上回ると妥当な検査とされる。

### 5.2 結果

各言語指標と HDS-R と Me-CDT の相関係数を表 3 に示す。ROC 曲線を図 5(HDS-R)と図 6(Me-CDT)に示す。TOKEN と MCI はほぼ無関係である。EL は高値になると、MCI を予測する。TTR 高値、NER 高値もわずかに MCI を予測する。FPU 高値も MCI を予測する。また、参考までに、CDT で HDS-R による MCI を識別すると AUC0.950 と高い性能が得られた。

## 6. 考察

### 6.1 TOKEN

実験結果より、TOKEN は MCI を予測しない可能性があることから、発話の量と MCI は無関係である可能性がある。実際の発話を観察すると、MCI および認知症疑いと簡易診断された者は、健常者と変わらないか、それよりも多い量の発話を行っているように見える。むしろ、発話の終了の判断がつきにくい傾向にある現象が観察された。

### 6.2 EL と FPU

EL が高いことが MCI である可能性を示しているが、これは MCI 者がより難易度の高い語を使用していることを示しており、直観と相反する。この点に関しては、音声認識誤りの影響の可能性が考えられる。不明瞭な発話が、音声認識の過程において難解な語、すなわち難易度の高い語へと誤変換されている可能性があり、この点についての調査が待たれる。

同様に、FPU も MCI との関連が示唆されており、MCI 者が特殊な語 (FPU 高値) を使っていることを示している。MCI および認知症疑い者の話の内容が、昔の話題中心となることで、使用語彙が、近年の平均的語彙と異なる事が考えられる。一方で、この現象についても、音声認識誤りによるバイアスの可能性もあり、今後の検討が望まれる。

### 6.3 TTR と MCI

TTR は MCI とわずかに関連している。TTR は多くの研究で認知症との関連が示唆されてきたが、本研究では、大きな傾向がみられなかった。本研究で得た発話は、ほとんどが 100 語程度の短いものであり、これが TTR の値を不安定にさせている可能性がある。また、従来の研究における TTR は書き言葉からの測定であったため、話し言葉による結果とは違った結果となっている可能性も考えられる。

### 6.4 CDT との比較

HDS-R と Me-CDT の AUC は 0.950 であるように、これらはよく関連する。これは、どちらも認知症に用いられる試験であり、かつ、同じ質問が含まれているため、高い関連が得られたと考えられる。Me-CDT は HDS-R よりも更に短い時間で簡単な評価を得るため開発された試験であるが、AUC0.950 は、十分代用たりうる精度と考えられる。

対して、EL (AUC=0.712) や FPU (ACU=0.711) は、妥当な検査となる値を越えているとはいえ、Me-CDT と比較すると低い値である。なお、3つの質問に必要な時間は3分前後であり、これは Me-CDT と同程度であって、効率の観点からは、Me-CDT を用いる事が最適である。しかし、一方で、今回用いたような言語使用調査の自動測定器による採取は簡便性があり、もっとも簡易な試験 (試験とさえ気づかれない) として、利用されうる可能性がある。

### 6.5 サンプル数の少なさからくる問題

本研究のサンプルは少なく、この実験だけから得られる強い結

論はなかった。したがって、3分ほどの音声発話で認知症者は見つかるか、については、その可能性がある、という結論に留まる。今後、症例を増加させることで、語彙の難解さや特殊性が、いかに認知症と関係するかの可能性を探ることが必要である。

### 6.6 今回の言語使用調査で調査できなかった問題

認知症患者の発話内容が、質問肢を何度も確認しているにもかかわらず、内容がずれてゆくなどの現象が見られた。今後、トピック抽出など研究で、このような現象を測定する指標を測定できるかもしれず、試みる価値がある。

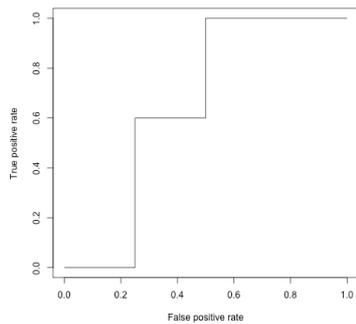
また、現時点での音声認識の精度の限界により、誤認識が多く認められた。話し言葉の自動認識の更なる精度改善が待たれる。

### 6.7 他の疾患への応用可能性

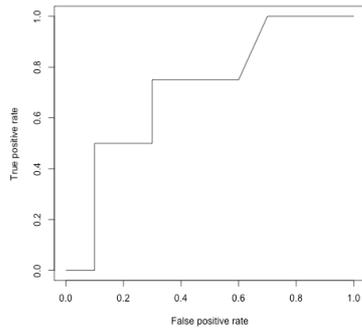
認知症のみならず、言語と関係する他の疾患についても、応用できる可能性がある。例えば、うつ病、発達障害 (自閉スペクトラム症や学習障害)、失語症などについても、自己モニタリングや早期発見が望まれている。本研究をすすめることによって、日本語測定法が確立されれば、これらへの応用も可能であり、大きな波及効果が期待される。

## 7. おわりに

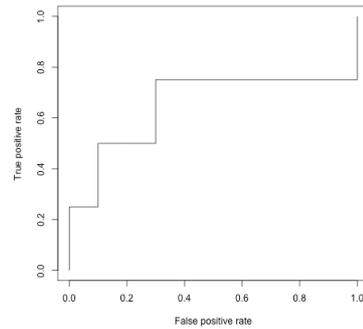
本研究は、自然言語処理技術を用いて、患者の語りを定量化し、認知症との関係を調査した。HDS-R と Me-CDT テストにより認知症疑いがある高齢者と正常な高齢者の語りに対して、音声認識を用いた後、言語解析し、語彙の統計量を調査した。語彙の難解さ (AUC 0.712) と特殊性 (AUC 0.711) が認知症との関連を示し、音声発話から、認知症を一定の精度でスクリーニングできる可能性が示された。今後の大規模な調査でこれを確認する必要がある。



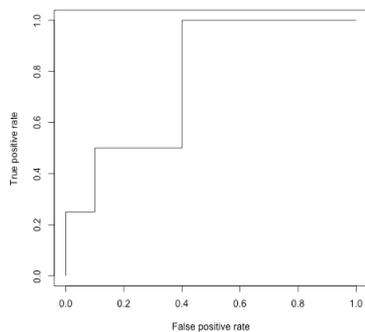
(a) TOKEN



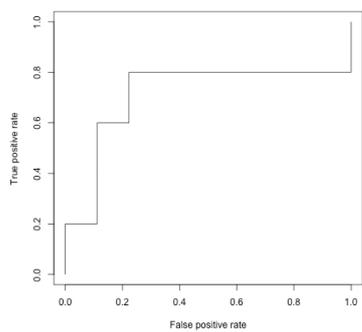
(b) EL



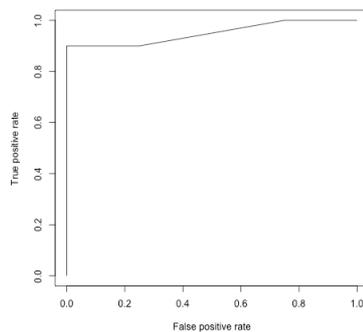
(c) TTR



(d) NER

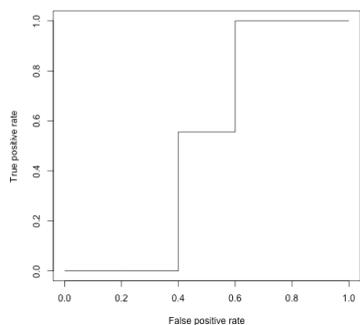


(e) FPU

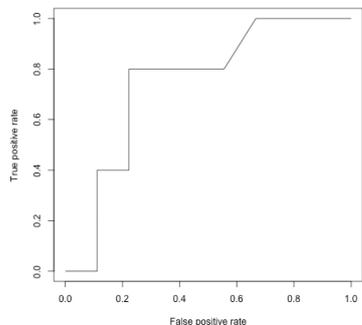


(f) CDT

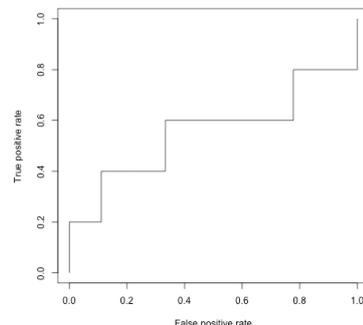
図 3: HDS-R と各言語指標の ROC.



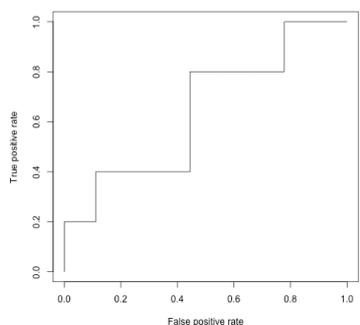
(a) TOKEN



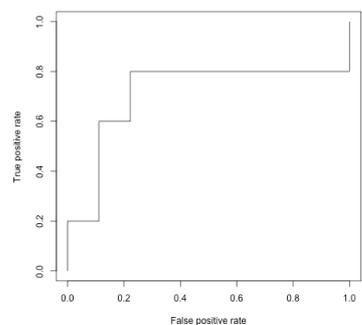
(b) EL



(c) TTR



(d) NER



(e) FPU

図 4: Me-CDT と各言語指標の ROC.

1. Machado, S., et al., *Alzheimer's disease and implicit memory*. Arq Neuropsiquiatr, 2009. **67**(2A): p. 334-42.
2. Randolph, C., M.C. Tierney, and T.N. Chase, *Implicit memory in Alzheimer's disease*. J Clin Exp Neuropsychol, 1995. **17**(3): p. 343-51.
3. Bayles, K.A., C.K. Tomoeda, and M.W. Trosset, *Relation of linguistic communication abilities of Alzheimer's patients to stage of disease*. Brain Lang, 1992. **42**(4): p. 454-72.
4. Kempler, D., et al., *Sentence comprehension deficits in Alzheimer's disease: a comparison of off-line vs. on-line sentence processing*. Brain and Language, 1998. **64**(3): p. 297-316.
5. Ecklin, R. and P.M. Schonenberger, *[Verbal ability of Alzheimer's disease patients and very elderly control patients using the Aachen aphasia test]*. Schweiz Arch Neurol Psychiatr, 1992. **143**(4): p. 371-9.
6. Nicholas, M., et al., *Empty speech in Alzheimer's disease and fluent aphasia*. J Speech Hear Res, 1985. **28**(3): p. 405-10.
7. Benke, T., et al., *[Speech changes in dementia]*. Fortschr Neurol Psychiatr, 1990. **58**(6): p. 215-23.
8. Nebes, R.D. and E.M. Halligan, *Instantiation of semantic categories in sentence comprehension by Alzheimer patients*. J Int Neuropsychol Soc, 1999. **5**(7): p. 685-91.
9. Price, C.C. and M. Grossman, *Verb agreements during on-line sentence processing in Alzheimer's disease and frontotemporal dementia*. Brain Lang, 2005. **94**(2): p. 217-32.
10. Salmon, D.P., W.C. Heindel, and K.L. Lange, *Differential decline in word generation from phonemic and semantic categories during the course of Alzheimer's disease: implications for the integrity of semantic memory*. J Int Neuropsychol Soc, 1999. **5**(7): p. 692-703.
11. Small, J.A., S. Kemper, and K. Lyons, *Sentence comprehension in Alzheimer's disease: effects of grammatical complexity, speech rate, and repetition*. Psychol Aging, 1997. **12**(1): p. 3-11.
12. Tsantali, E., D. Economidis, and M. Tsolaki, *Could language deficits really differentiate Mild Cognitive Impairment (MCI) from mild Alzheimer's disease?* Arch Gerontol Geriatr, 2013. **57**(3): p. 263-70.
13. Ramstrom, I., *Linguistic development in Alzheimer's disease: 12 months language training including use of a personal computer system: a pilot study*. Dev Neurorehabil, 2011. **14**(3): p. 156-63.
14. Taler, V. and N.A. Phillips, *Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review*. J Clin Exp Neuropsychol, 2008. **30**(5): p. 501-56.
15. Kaplan, E.F., H. Goodglass, and S. Weintraub, *Boston Naming Test*. 1983: Philadelphia.
16. McKenna, P. and E. Warrington, *The Graded Naming Test*. 1983: Windsor.
17. Adlam, A.L., et al., *Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease*. Cortex, 2006. **42**(5): p. 675-84.
18. De Renzi, E. and L.A. Vignolo, *The token test: A sensitive test to detect receptive disturbances in aphasics*. Brain, 1962. **85**: p. 665-78.
19. Hodges, J.R., D.P. Salmon, and N. Butters, *Semantic memory impairment in Alzheimer's disease: failure of access or degraded knowledge?* Neuropsychologia, 1992. **30**(4): p. 301-14.
20. Snowden, D.A., et al., *Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study*. JAMA, 1996. **275**(7): p. 528-32.
21. Kemper, S., et al., *On the preservation of syntax in Alzheimer's disease. Evidence from written sentences*. Arch Neurol, 1993. **50**(1): p. 81-6.
22. Mitzner, T.L. and S. Kemper, *Oral and written language in late adulthood: findings from the Nun Study*. Exp Aging Res, 2003. **29**(4): p. 457-74.
23. Small, J.A., S. Kemper, and K. Lyons, *Sentence repetition and processing resources in Alzheimer's disease*. Brain Lang, 2000. **75**(2): p. 232-58.
24. Kemper, S., *Metalinguistic judgments in normal aging and Alzheimer's disease*. J Gerontol B Psychol Sci Soc Sci, 1997. **52**(3): p. 147-55.
25. Uzuner, O., *Second i2b2 workshop on natural language processing challenges for clinical records*. AMIA Annu Symp Proc, 2008: p. 1252-3.
26. Kelly, L., et al., *Overview of the ShARe/CLEF eHealth Evaluation Lab 2014*, in *Information*

- Access Evaluation. Multilinguality, Multimodality, and Interaction.* 2014. p. 172-191.
27. Aramaki, E., et al., *Overview of the NTCIR-11 MedNLP-2 Task*, in *NTCIR-11*. 2014.
28. Prieto, V.M., et al., *Twitter: a good place to detect health conditions*. PLoS One, 2014. **9**(1): p. e86191.
29. Aramaki, E., S. Maskawa, and M. Morita, *Twitter catches the flu: detecting influenza epidemics using Twitter*, in *EMNLP*. 2011. p. 1568-1576.
30. Lamb, A., M. Paul, and M. Dredze, *Separating fact from fear: Tracking flu infections on Twitter*, in *Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 2013.
31. Sugumaran, R. and J. Voss, *Real-time spatio-temporal analysis of West Nile virus using Twitter data*, in *Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications*. 2012. p. 1-2.
32. Bian, J., U. Topaloglu, and F. Yu, *Towards large-scale twitter mining for drug-related adverse events*, in *Proceedings of the 2012 international workshop on Smart health and wellbeing*. 2012. p. 25-32.
33. Collier, N., et al., *BioCaster: detecting public health rumors with a Web-based text mining system*. Bioinformatics, 2008. **24**(24): p. 2940-1.
34. Lyon, A., et al., *Comparison of web-based biosecurity intelligence systems: BioCaster, EpiSPIDER and HealthMap*. Transbound Emerg Dis, 2012. **59**(3): p. 223-32.
35. Freifeld, C.C., et al., *HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports*. J Am Med Inform Assoc, 2008. **15**(2): p. 150-7.
36. Brownstein, J.S. and C.C. Freifeld, *HealthMap: the development of automated real-time internet surveillance for epidemic intelligence*. Euro Surveill, 2007. **12**(11): p. E071129 5.
37. Wang, W., et al., *Discovering Fine-grained Sentiment in Suicide Notes*. Biomed Inform Insights, 2012. **5**(Suppl. 1): p. 137-45.
38. Roark, B., et al., *Automatically derived spoken language markers for detecting mild cognitive impairment*, in *the 2nd International Conference on Technology and Aging (ICTA)*. 2007.
39. Fraser, K.C., F. Rudzicz, and E. Rochon. *Using text and acoustic features to diagnose progressive aphasia and its subtypes*. in *Interspeech 2013*. 2013.
40. Pakhomov, S.V., et al., *A computerized technique to assess language use patterns in patients with frontotemporal dementia*. J Neurolinguistics, 2010. **23**(2): p. 127-144.
41. Pakhomov, S.V., et al., *Computerized analysis of speech and language to identify psycholinguistic correlates of frontotemporal lobar degeneration*. Cogn Behav Neurol, 2010. **23**(3): p. 165-77.
42. Rouhizadeh, M., et al., *Detecting linguistic idiosyncratic interests in autism using distributional semantic models*, in *The Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 2014. p. 46-50.
43. Francis, W.N., H. Kučera, and A.W. Mackie, *Frequency analysis of English usage: lexicon and grammar*. 1982: Houghton Mifflin.
44. Covington, M.A., et al., *How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level scale*, in *CASPR Research Report*. 2006.
45. Cheunga, H. and S. Kemper, *Competing complexity metrics and adults' production of complex sentences*. Applied Psycholinguistics, 1992. **13**(1): p. 53-76.
46. Frazier, L., *Syntactic complexity*, in *Natural Language Parsing*, L.K. Dowty and A.M. Zwicky, Editors. 1985, Cambridge University Press.
47. Lin, D., *On structural complexity*, in *COLING96*. 1996.
48. Yngve, V.H., *A model and an hypothesis for language structure*, in *he American Philosophical Society*. 1960. p. 444-466.
49. D'Anna, C.A., E.B. Zechmeister, and J.W. Hall, *Toward a meaningful definition of vocabulary size*. Journal of Reading Behavior, 1991. **23**: p. 109-122.
50. Goulden, R., P. Nation, and J. Read, *How Large Can a Receptive Vocabulary Be?* Applied

- Linguistics, 1990. **11**(4).
51. Lorge, I. and J. Chall, *Estimating the size of vocabularies of children and adults: an analysis of methodological issues*. Journal of Experimental Education, 1963. **32**(2): p. 147-157.
  52. Nation, P., *Using dictionaries to estimate vocabulary size: essential, but rarely followed, procedures*. Language Testing, 1993. **10**(1).
  53. Maas, H., *Zusammenhang zwischen Wortschatzumfang und Länge Eines Textes*. Zeitschrift für Literaturwissenschaft und Linguistik, 1972. **8**: p. 73-79.
  54. Richards, B., *Type/Token Ratios: what do they really tell us?* J Child Lang, 1987. **14**(2): p. 201-9.
  55. Kihara, T., et al., *Development of Memory-entailed Clock Drawing Test (Me-CDT) a brief screening tool for the detection of Alzheimer's disease (in Japanese)*. progress in medicine, 2013. **33**(2): p. 193-199.
  56. 加藤伸司, et al., *改訂長谷川式簡易知能評価スケール (HDS-R) の作成*. 老年精神医学雑誌, 1991. **11**: p. 1339-1347.
  57. Kurohashi, S., et al. *Improvements of Japanese Morphological Analyzer JUMAN*. in *The International Workshop on Sharable Natural Language Resources*. 1994.
  58. Sunakawa, Y., *学習辞書編集支援データベース作成について -『学習辞書科研』プロジェクトの紹介-*. 日本語教育連絡会議論文集, 2012. **24**.
  59. Aramaki, E., et al. *A Word in a Dictionary is used by Numerous Users*. in *International Joint Conference on Natural Language Processing (IJCNLP2013)*. 2013. Japan.