

音声認識による認知症・発達障害スクリーニングは可能か？ - 言語能力測定システム“言秤”の提案 -

宮部 真衣^{1,a)} 四方 朱子¹ 久保 圭² 荒牧 英治¹

概要：近年，日本において認知症や発達障害は身近なものとなっている．そして，認知症・発達障害のいずれも，言語能力に何らかの特徴が表出する可能性があることはよく知られている．言語能力を測り，それらの兆候を捉えることができれば，早期発見や療養に役立つ可能性がある．これまでに，言語能力の測定に関する取り組みはあるものの，いずれも人手を介して測定を行うためコストが高く，気軽に測定することは難しい．そこで本研究では，認知症や発達障害などにおける初期判断や自己把握・状況改善での利用を想定し，手軽に言語能力を測定可能なシステム「言秤（コトバカリ）」を提案する．本提案システムでは，(1) 音声認識システムの組み込み，および(2) テキストデータから定量的に言語能力を測定する指標の採用を行うことで，従来人手で行っていたテキスト化および言語能力スコアの算出を自動化し，コストの軽減と手軽な測定を実現する．また，「被測定者自身による自己把握・状況改善（用途 1）」および「被測定者以外による初期判断（用途 2）」という観点から，言語能力スコア（Type・Token 比）算出における音声認識システムの利用可能性について検証を行った．検証実験の結果，閾値との比較のような，単純な言語能力スコアの対比による初期判断（用途 2）は難しいが，被測定者の言語能力スコアを継続的に測定し，その変化を観察することによる初期判断（用途 1）や言語能力の現状把握・維持・改善（用途 2）ができる可能性があることを示した．

1. はじめに

日本は世界に先駆けて超高齢社会に突入した．2013 年の高齢化率は 25.1%にのぼり^{*1}，世界でも例を見ないスピードで高齢化が進行している．高齢化の進行に伴い，認知症高齢者の増加も見込まれる．2012 年 8 月の厚生労働省発表によると，2010 年における日常生活自立度 II^{*2}以上の認知高齢者数^{*3}は 280 万人にのぼり，将来推計として 2025 年には 323 万人，65 歳以上の人口比率にして 9.3%にまで上昇するだろうと予測されている^{*4}．また，日本における人口 10 万人当たりの若年性認知症者数（18～64 歳）は，47.6

人にものぼるとされている^{*5}．今や，認知症は我々にとって非常に身近なものとなっている．

発達障害もまた，日本において身近になりつつある．発達障害は，「言葉や認知の面など，様々な領域において発達に遅れがみられる障害」であり [1]，文部科学省が発表した，小中学校の教職員に対する調査によると，通常学級に在籍する児童のうち，発達障害の可能性のある児童数の割合が 6.5%に達するという結果が報告されている^{*6}．

厚生労働省の調査によると^{*7}，最初に認知症に気づききっかけとなる症状の一つとして，言語障害がある．言語能力は長期におよぶ学習や経験によって発達するものであり，一定レベルまで発達した後は，加齢によっても衰えにくいとされる [2]．一方で，構文をあやつる能力は，70 代後半を境に低下しはじめるという報告もある [3]．また，前述したように，発達障害における症状の一つに言語発達の

¹ 京都大学学際融合教育研究推進センター
Center for the Promotion of Interdisciplinary Education and Research, Kyoto University

² 大阪大学日本語日本文化教育センター
Center for Japanese Language and Culture, Osaka University

^{a)} mai.miyabe@gmail.com

^{*1} 内閣府 平成 26 年 5 月「選択する未来」委員会，
[http://www5.cao.go.jp/keizai-](http://www5.cao.go.jp/keizai-shimon/kaigi/special/future/chuukanseiri/04.pdf)

^{*2} 「日常生活に支障を来すような症状・行動や意志疎通の困難さが多少みられても，誰かが注意していれば自立できる状態」を指す．

^{*3} 65 歳以上を指す

^{*4} 認知症高齢者数について，
<http://www.mhlw.go.jp/stf/houdou/2r985200002iaiu1.html>

^{*5} 若年性認知症の実態等に関する調査結果の概要および 厚生労働省の若年性認知症対策について，
<http://www.mhlw.go.jp/houdou/2009/03/h0319-2.html>

^{*6} 通常の学級において発達障害の可能性のある特別な教育的支援を必要とする児童生徒に関する調査結果，
http://www.mext.go.jp/a_menu/shotou/tokubetu/material/1328729.htm

^{*7} 厚生労働省の認知症施策等の 概要について，
<http://www.mhlw.go.jp/file/05-Shingikai-12301000-Roukenkyoku-Soumuka/0000031337.pdf>

遅れが挙げられ、認知症・発達障害のいずれも、言語能力に何らかの特徴が表出する可能性をもつものである。そこで、もし言語能力を測り、その兆候を捉えることができれば、早期発見や療養に役立つのではないかと考えた。

近年、大規模なコホート研究によって、数十年の言語能力の経過を観察する試みが行われている。その結果、老化や認知症などと加齢によるさまざまな能力との関係は徐々に明らかになりつつある [4], [5]。また、発達障害当事者の <語り> を分類し (コーディング)、定量的に扱うことで、病態の理解、または診断や治療に用いる動きもある [6]。しかし、これらの取り組みには、人手でのテキストデータの作成・収集、テキストの分類、各評価スコアの算出などが必要であり、時間・金銭的成本が高い。また、テキストの分類には訓練をつんだ専門家が必要であるなど、言語能力測定ハードルの高い。

本研究では、認知症・発達障害スクリーニングのための、言語能力測定システム「言秤 (コトバカリ)」を提案する。言語能力測定における大きな課題の一つとして、分析対象となるテキストデータの作成がある。従来は、音声データの書き起こしなど、人手でのテキスト化が必要であり、多大な時間を要していた。提案システムでは、音声認識システムにより、言語能力の被測定者の発話データを自動的にテキスト変換することで、コストの大幅な軽減を行う。ただし、音声認識システムの認識精度には限界があり、常に正しい認識結果が得られるとは限らない。

我々はこれまでに、テキストデータに基づいて定量的に言語能力を測定する指標 (以降、言語能力指標と表記する) を提案してきた [7]。提案指標を用いたテキスト分析の結果、提案指標によって認知症者の特徴的な傾向などを観察できる可能性を示した [7], [8]。この提案指標の一部には、語の内容ではなく、出現回数のみに基づいて算出されるものがある。この場合は、たとえ語の内容の認識結果が誤っていても、算出されるスコアに問題は生じにくいと考えた。よって、提案システムでは、この提案指標を採用し、音声認識による認識誤りの影響を受けにくく、さらに人手作業を排除した定量的なスコア算出を実現する。

本提案のポイントを以下に整理する。

分析テキスト作成コストの軽減 音声認識システムを組み込むことにより、音声を録音/入力するだけで、言語能力の測定を可能にする。

スコアリングコストの軽減 定量的に算出可能な言語能力指標を採用することにより、認識誤りを許容し、さらにスコア算出のための人間の介入を省略可能にする。

本稿では、採用する言語能力測定指標と提案システムについて概説した後、低コストな言語能力測定の要となる音声認識システムの利用可能性について、検証実験の結果から議論する。

2. 関連研究

本章では、言語能力と老化/疾患に関する先行研究および音声認識の活用事例について述べる。

2.1 言語能力と老化/疾患

一般に、老化と言語能力とは相関しないといわれている。例えば、一部の言語能力は、高齢者においても伸び続けるため、老人が理解している語彙数は、若者の約 1.3 倍だという報告もある [9]。その一方で、人間が構文をあやつる能力は、70 代後半を境に低下しはじめるという報告もある [3]。

失語症などの言語との関連が深い疾患を除けば、罹患中の言語能力の変化についての調査は少ない。しかし、85 歳以上の 40% が罹患している認知症 [10] については近年研究が進んでいる。Snowdon は、認知症者は認知症の症状が認められ始める 50 年前から語彙能力が低いと報告した [5]。また、Kemper によると、英文における認知症の進行度は語彙能力、構文能力ともに相関関係にあり、症状が進行するにつれ、構文能力の顕著な低下がみられるという [3]。

ただし、ある人間が認知症か否かを判定するための項目に、言語能力に関するものがあることを考慮すれば、認知症患者が言語能力の低下を示すことは、ある種のトートロジーであり、言語能力と認知症、さらに老化の関係は、複雑な様相を呈しているといえる。

文章から認知症などの疾患を推定することが可能となれば、測定の侵襲度が低く、かつ、超早期からの検出が可能となる。

2.2 音声認識の活用事例

近年、音声認識技術が進展し、様々な場面で音声認識が利用されている。例えば、スマートフォンの普及に伴い、ブラウザ上での音声検索なども利用できるようになった。また、旅行対話を対象とした音声翻訳 [11], [12], [13] や、コールセンターの自動案内のための電話音声認識 [14]、英語字幕の生成 [15]、議事録の作成 [16] など、音声認識を利用した多様なシステムがこれまでに提案されている。

旅行対話など、分野を限定した場合、音声認識によって高精度な認識を実現することは可能であるが、分野を制限せずに精度のよい認識結果を得ることは未だ容易ではない。従来の音声認識を利用したシステムは、限定的な分野を対象にするなどして精度改善の工夫を行い、正しい認識結果を取得し、その結果を利用することが前提となっている。

本研究では、言語能力の測定において、音声認識を利用する。精度の良い認識結果が得られれば、より正確な言語能力測定が可能であるが、認識誤りが生じていても、その影響を受けにくいと考えられる言語能力指標を採用することで、頑健な言語能力測定の実現を目指している。

表 1 本研究で用いる言語能力指標

指標	略記	概要	単位 / 方法
(1) Type・Token 割合	TTR	語彙の量を示す	文章指標 / 統計量
(2) 頻度・使用者数比	FPU	語彙の特殊性を示す	単語指標 / 統計量
(3) 日本語学習語彙レベル	JEL	語彙の難しさを示す	単語指標 / 経験則
(4) 機能表現難度	FNC	難易度を示す	単文指標 / 経験則
(5) ポライトネス	PLT	丁寧さを示す	単文指標 / 経験則
(6) 具体性	NER	固有名詞の割合を示す	単文指標 / 経験則

3. 言語能力指標

本章では、提案システムで採用する言語能力指標について概説する。

一般に、言語能力は「話す (speech)・聞く (listening)・読む (reading)・書く (writing)」の4つに大別される。本研究は、人間からの出力に関わる、<話す>能力および<書く>能力に注目する。<話す>能力および<書く>能力は、大別すると語彙に関する能力(以降、語彙能力とよぶ)と文法に関する能力(以降、文法能力とよぶ)の2つになる [17], [18]。本研究で採用する言語能力指標は、主に語彙能力に関わるものであり、以下の2つの観点で分類できる。

単位: 語単位で算出され、それを全文で平均するタイプのもの(単語指標)、文単位で算出され、それを全文で平均するタイプのもの(単文指標)、および使用語彙数など、全文から導出されるタイプのもの(文章指標)
方法: 経験則によりレベル付けされたもの(経験則)、および大規模コーパスから統計的に算出されるもの(統計量)

これらの観点から指標を分類した結果を表 1 に示す。以下、各指標の定義を述べる。

(1) Type・Token 割合 (Type Token Ratio; TTR):

Type(異なり語数)とToken(延べ語数)の比率(Type / Token)を示す。この値が大きいほど、語彙量が多いことを意味する。文章全体で集計した値を TTR スコアとする。

(2) 頻度・使用者数比 (Frequency per User Popularity; FPU):

語の特殊性を示す指標である。語の特殊性は、語の出現頻度を語のユーザ数で割った値と定義する。この値が低いほど、その語は一般的であり、高いほど、その語のユーザ数が少ない語(特殊な語)であることを示す。例えば、スラングや専門用語などは高い値を持つ。ソーシャルメディア上の10万人の発言を8ヶ月間調査して得たデータをもとに、語のユーザ数および出現頻度を算出し、各語の頻度・使用者数比コーパスを構築した [19]。このコーパスを参照して語ごとに頻度・使用者数比を算出し、全単語の値を平均した値を FPU スコアとする。

(3) 日本語学習語彙レベル (Japanese Educational Lexicon Level; JEL):

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^{*8}に記載されている語彙レベルを用いた。語彙レベルは、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 [20]。語ごとに算出し、全単語の平均を JEL スコアとする。

(4) 機能表現難度 (Difficulty of Functional Expression; FNC):

機能表現^{*9}の難易度を示す。この値が大きいほど、文章内で用いられている機能表現の難易度が高いことを意味する。難易度の定義は「日本語機能表現辞書つづじ」[21]で設定されている難易度による。難易度は A1, A2, B, C, F の5段階に分かれており、これを1(A1)から5(F)に変換した。文ごとに算出し、平均した値を FNC スコアとする。

(5) ポライトネス (Politeness of Functional Expression; PLT):

機能表現の丁寧さの度合いを示す。この値が大きいとき、機能表現が丁寧であることを表す。ポライトネスは「日本語機能表現辞書つづじ」[21], [22]の分類を採用した。分類では機能表現が常体(normal)、敬体(polite)、口語体(colloquial)、堅い文体(stiff)の4種類に分けられており、口語体(colloquial) = 1、常体(normal) = 3、敬体(polite) = 5、堅い文体(stiff) = 5に変換した。文ごとに算出し、平均した値を PLT スコアとする。

(6) 具体性 (Named Entity Ratio; NER):

固有名詞の割合を示す。具体性は、固有名形態素数を全名詞形態素数で割った値と定義する。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^{*10}[23]を用いて行う。地名、数詞、固有名詞の割合を文ごとに算出し、平均した値を NER スコアとする。

以降、これらの言語能力指標をもとに算出される値を

^{*8} <http://jishokaken.sakura.ne.jp/DB/>

^{*9} 日本語の文を構成する要素のうち、機能語(助詞や助動詞といった、主に文の構成に関わる要素)と複合辞(複数の語から構成され、かつ、全体として機能語のように働く表現)を総称したものの。

^{*10} <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

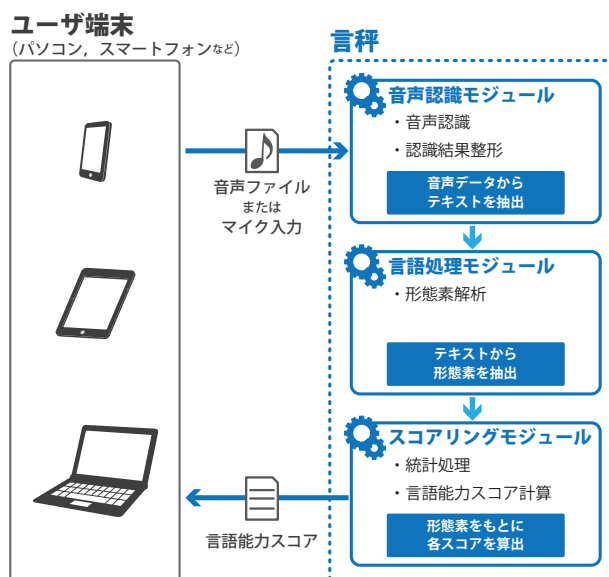


図 1 システム構成

「言語能力スコア」と総称する。なお、これらの指標は、人間の言語能力によって表出したテキストの特徴を表すものの一つであると見なし、本稿では「言語能力指標」と呼ぶこととしているが、これらの指標は人間の言語能力を直接的に表すものではない。

4. 提案システム：言秤（コトバカリ）

本章では、まず、想定するシステムの用途を整理し、提案システムの構成について述べる。

4.1 想定するシステムの用途

本研究における言語能力の測定対象として、認知症や発達障害を抱える人々を想定している。

認知症予防回復支援のアプローチの一つとして、認知的アプローチがある [24]。これは、知的活動と社会的ネットワークの構築により、認知症になると衰える認知機能を必要とする認知活動を行い、認知機能の低下を遅らせるものである。一方、発達障害の場合、言葉として症状が観察される障害であるため、当事者の「語り」が注目されることも多い。例えば、うつ患者が自伝的語りを行うことによって症状が回復するなどの報告事例もあり [6]、言語情報が病態の理解・診断・治療に用いられる動きがある。本人が認知症や発達障害といった自分の状態を認識している場合、その状態の維持や改善に取り組みたいと考えている場合、被測定者自身が前向きかつ積極的にシステムを利用する可能性がある。

一方で、認知症の疑いのある人の言語能力を測定する場合などは、上述したケースとは異なると考えられる。日本イーライリリー株式会社が行った調査^{*11}によると、認知症

を疑うきっかけとなる変化に気づいてから、最初に医療機関を受診するまでにかかった期間は平均 9.5 か月、変化に気づいてから確定診断までにかかった期間は平均 15.0 か月であるとの調査結果がある。認知症には様々な原因があり、早期診断で治療可能なものがあるが、上述の調査では、確定診断までに時間がかかったことによる患者や家族の負担に関して、「適切な治療がなされなかった」という回答が 36.7%を占めており、早期診断の重要性が見て取れる。しかし、それに関わらず、認知症の疑いのある人自身が受診を嫌がる^{*12}ことなども多い。このような場合、被測定者の周囲の人々（家族など）は測定に対して前向きであっても、言語能力の被測定者がシステムの利用に後ろ向きである可能性がある。

そこで本提案システムでは、以下の 2 種類の用途を想定し、システムの検討を行う。

用途 1： 被測定者 自身 による自己把握・状況改善

被測定者が自発的にシステムを利用し、その結果を自分自身で確認することにより、言語能力の現状を把握し、改善などに役立てることを想定している。

用途 2： 被測定者 以外 による初期判断

例えば、認知症の疑いのある人（被測定者）の家族などが、被測定者の発話を音声データとして保存しておき、そのデータから言語能力を測定することにより、被測定者の言語能力傾向を把握するために用いることを想定している。

4.2 システム構成

本研究では、3 章で述べた 6 指標に基づき、言語能力を測定するシステム「言秤」を提案する。

言秤のシステム構成を図 1 に示す。

4.1 節で述べたシステムの用途を考慮し、音声データの inputs は、入力方式 (1) ユーザ端末に保存された音声データ (wav ファイルなど)、または入力方式 (2) マイクによるリアルタイム入力を想定する。入力方式 (1) は、用途 2 のように、被測定者以外がシステムを利用する場合に必須となり、また用途 1 でも利用可能である。入力方式 (2) については、用途 1 のように自発的な測定を行う際に、音声データの作成の手間を減らし、その場で直接入力するために必要であると考えた。

音声データあるいはマイク入力に対し、以下の 3 つのモジュールでの処理を行い、言語能力スコアを出力する。

- (A) 音声認識モジュール 音声認識ソフトウェアにより、入力音声を変換する。
- (B) 言語処理モジュール 音声認識モジュールにより変換したテキストデータに対し、形態素解析を行う。

^{*11} https://www.lilly.co.jp/pressrelease/2014/news_2014_033.aspx

^{*12} http://www.fukushihoken.metro.tokyo.jp/kourei/ninchi/suishin_kaigi/haifushiryocarepass5.files/23carepass5_sankou3.pdf

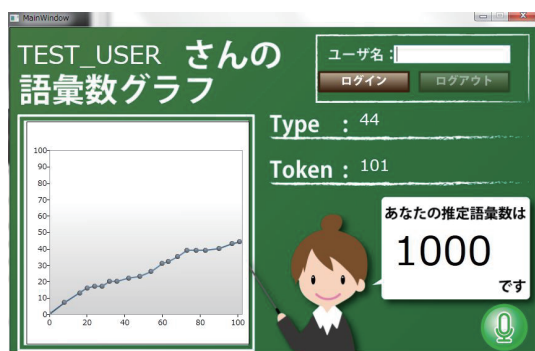


図 2 言程による測定結果フィードバック画面のイメージ

(C) スコアリングモジュール 言語処理モジュールによって得られた解析結果をもとに, 3章で述べた言語能力スコアを算出する.

(2) マイクによるリアルタイム入力の場合, 入力結果を随時可視化しながらフィードバックする(図2)こともできるようにする.

5. 検証実験

音声認識システムによる言語能力の測定可能性を検証するために, 音声データを用いた言語能力スコアの比較実験を行う. 本章では, 検証仮説と評価に用いるデータについて説明する.

5.1 検証仮説

音声認識システムの認識精度が高い場合, 認識結果が発話内容と完全一致することも考えられる. もし, 完全一致するならば, 算出される言語能力スコアには何の問題もない. しかし, 現在の音声認識システムでは, 常に高精度な認識ができるとは限らない. では, 認識結果と実際の発話内容にずれがある場合, 正しい言語能力スコアの算出はできないのだろうか.

これまでに行った分析の結果, 我々が提案・採用する6つの言語能力指標のうち, 認知症および発達障害の傾向把握に大きく寄与しうる指標は, TTR (Type・Token 割合) および JEL (日本語学習語彙レベル) であることがわかっている [7]. このうち, 特に TTR は, Type (異なり語数) と Token (延べ語数) の比率 (Type / Token) であり, TTR スコアの算出において, 発話されたのがどのような語であるかという点については関与しない.

我々はこの点において, 音声認識システムによる認識結果が不完全であっても, TTR スコアを算出できる可能性があると考えた. 具体的には, ある語の認識を誤った (例えば, 「機械」という発話を「機会」と変換したり, 全く異なる語として認識するなど) としても, 発話者によって語の発生の仕方が変わらないと仮定すれば, 音声認識システムは毎回同じ誤りとして出力し, 総じて異なり語数, 述べ語数が大きく変わらない可能性があると考えている. 異な

り語数, 述べ語数が大よそ合っていれば, 仮説上は TTR スコアも実際の発話内容から算出されるものと類似するはずである.

そこで, 本実験では, 「音声認識システムの認識結果が正しくなくとも, Type 数, Token 数は実際の発話データと相関する」という仮説を立て, 言語能力スコアの算出・比較を行う.

なお, 今回は, 音声認識することを想定していない録音音声データを評価用音声データとし, 検証を行う. マイクに向かい, 音声認識することを意図して入力するよりも認識精度が低くなると考えられ, より劣悪な環境下を想定した検証結果になると考えている.

5.2 評価用音声データ

評価用音声データとして, 模擬面接の設定で収録された音声データ [25] を用いた. 今回用いたデータは, 模擬面接の設定において, 5名の実験協力者 (男性2名, 女性3名であり, 年齢は20~40代である) が各10回収録したデータ (合計50回分) である.

就職活動を前提とした模擬面接の設定で, 実験協力者はあらかじめ考えてきた「学生生活で力を入れてきたこと」についての発話 (3分間程度) を行ってもらった. なお, 収録時, 偶数回のみ聴衆 (面接官役) を配置したが, 聴衆には聴いていることを表すためにうなづくことのみを許可し, 話者への質問や意見など, 発話は一切行わないようにしている.

発話内容は, ボイスレコーダーおよびビデオカメラにおいて収録した. 収録環境のイメージを図3に示す. ボイスレコーダーは机の上に置き, 被験者からは少し距離のある状態で録音している. 評価用音声データとしては, ボイスレコーダーで収録したものを利用した. なお, このデータの収録においては, 同じ内容を話してもらっているが, テキストを読み上げるのではなく, 面接の設定でその場で話してもらっているため, 実験協力者の10回の発話内容はそれぞれ異なっている^{*13}.

発話者には, 発話内容を録音していることを伝えているが, 音声認識によるテキスト化を前提として収録したものではない. したがって, このデータは, 録音を意識している可能性はあるが, 音声認識することを意識した発話データではないといえる.

5.3 評価用テキストデータの作成

音声認識による言語能力測定の可能性を検証するためには, 5.2節で述べた各評価用音声データに対して, テキストデータを作成する必要がある.

まず, 音声認識結果と比較するための正解データを作成

^{*13} 10回収録しているが, 同内容を繰り返すことや何回依頼するかは実験協力者に知らせていない.

表 2 各テキストデータの一部

書き起こし	Julius	AmiVoice
えー、ただ、その、卒業した後は、またそれまでであった分野では全くないんですけども、情報系のほうに就職したいなということで、えー、その分野について自分で勉強しました。で、これまでいろいろな分野にチャレンジしてきたんですけども、そのたびに、あの、学校では習ったことがないような新しい分野について、えー、たくさん、いや、いろいろ結構、自分なりに勉強してきました。	てたな。が出土したとは、多数の狙った分野では全くないんです、別の情報家の方に就職したいなということで、です。分野について、自分で勉強しました。てとメールの分野にチャレンジしたんですけども、そのために今の自覚じゃなかった事あなたが新しい分野について、でさ、対〇でも、自分なりに勉強してきました。	ただの卒業した後は集まれあった分野では全くないんですけども、情報系の方に就職したいなということでその分野について一面で勉強しましたってこれまでいろいろな分野にチャレンジしてきたんですけども、そのために近くでなったことがないあなたが心についてで沢山弥生ロケット自分なりに勉強してきました

Julius, AmiVoice のテキストについては、音声認識システムから出力された結果をそのまま掲載している。

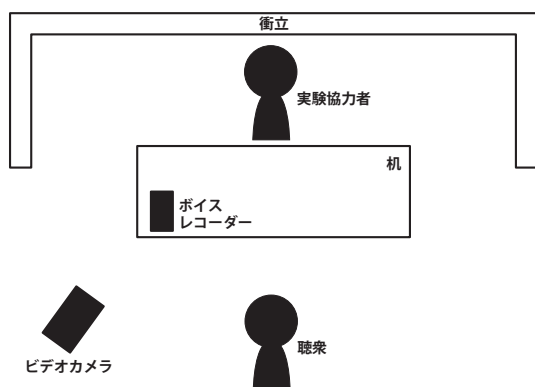


図 3 音声データの収録環境

した。正解データについては、音声を聞きながら、人手で書き起こし作業を行い作成した。なお、書き起こしの際は、「えー」や「あー」などの言いよどみ(フィラー)もテキストとして書き起こしている。

次に、音声認識システムによるテキストデータの生成を行った。現在、様々な音声認識システムが公開されているが、その精度はシステム毎に異なると考えられる。そこで、実験結果に対する音声認識システムの精度の影響を考慮し、異なる 2 種類の音声認識システムを用いて実験を行うこととした。本実験においては、以下の 2 種類の音声認識システムを用いて、音声認識結果を生成する。

- 大語彙連続音声認識エンジン Julius[26]
- アドバンスド・メディア社の AmiVoice SP2^{*14}

本検証実験で用いるテキストデータは以下の 3 種類である。各テキストの一部を表 2 に示す。

- (1) 書き起こしデータ(以下、「書き起こし」と表記する)
- (2) Julius を用いた認識結果(以下, Julius と表記する),
- (3) AmiVoice SP2 を用いた認識結果(以下, AmiVoice と表記する)

6. 実験結果と考察

本章では、それぞれのテキストに対し、スコアリングモ

表 3 各スコアの平均値

	書き起こし		Julius		AmiVoice	
	平均	S.D.	平均	S.D.	平均	S.D.
Type	129.9	34.3	152.1	50.0	148.4	39.2
Token	400.1	120.3	317.8	129.9	284.7	114.2
TTR	0.3	0.0	0.5	0.1	0.5	0.1
FPU	28.7	7.0	26.4	6.0	26.1	5.8
JEL	2.9	0.2	2.9	0.3	3.0	0.3
FNC	2.0	0.4	1.7	0.3	1.8	0.4
PLT	2.9	0.3	2.9	0.5	2.9	0.4
NER	1.0	0.0	1.0	0.0	1.0	0.0

ジュールによる言語能力スコアの算出結果を比較する。なお、本稿では理論的に認識誤りに頑健な指標である TTR スコアをもとに、音声認識システムを用いた言語能力測定の可能性を議論する。TTR スコア以外の言語能力スコアについては、今回は参考値として提示することとし、今後、認識結果と併せた詳細な分析を行うことにより、測定可能性を検証する。

6.1 平均スコアとデータ間の相関

Type(異なり語数), Token(延べ語数)および 3 章で述べた 6 指標の平均値を表 3 に示す。表 3 より、Type については書き起こしよりも、Julius, AmiVoice の方が多い傾向がみられる。一方、Token については、書き起こしよりも Julius, AmiVoice の方が少ない傾向がみられた。TTR については、Julius, AmiVoice が、書き起こしよりも若干高い値となった。その他の言語能力スコアについては、大きな違いはみられなかった。

各スコアに関する、テキストデータ間のピアソンの相関係数を表 4 にそれぞれ示す。表 4 より、Type, Token についてはいずれの組み合わせでも 0.9 前後の強い正の相関 ($p < 0.05$) が確認できた。一方、TTR スコアについては、書き起こしと Julius, 書き起こしと AmiVoice の相関係数はいずれも 0.2 未満であった。

表 4 における相関係数は、50 回分の発話データ全体での相関を調べたものである。発話者ごとの TTR スコアの

*14 <http://sp.advanced-media.co.jp/>

表 4 テキストデータ間の相関係数

	書き起こし / Julius	書き起こし / AmiVoice	Julius / AmiVoice
Type	0.901*	0.967*	0.904*
Token	0.899*	0.916*	0.938*
TTR	0.184	0.177	0.721*
FPU	0.493*	0.586*	0.361*
JEL	0.307*	0.577*	0.368*
FNC	0.587*	0.426*	0.381*
PLT	0.234	0.242	0.196
NER	-0.268	0.080	0.082

* : $p < 0.05$

相関係数が 0.4 以上であったものを太字で示している。

表 5 話者別に見た TTR スコアに関するテキストデータ間の相関係数

	書き起こし / Julius	書き起こし / AmiVoice
発話者 A	0.690*	0.830*
発話者 B	0.049	0.533
発話者 C	0.492	0.379
発話者 D	0.528	0.876*
発話者 E	-0.156	0.588

* : $p < 0.05$

相関係数が 0.4 以上であったものを太字で示している。

相関係数を表 5 に示す。表 5 より、個人差はあるものの、特に AmiVoice を用いた認識結果については、正の相関がある傾向がみられた。

6.2 発話音声データに基づく言語能力測定の可能性

6.2.1 TTR の測定可能性

6.1 節で示したように、検証の結果、Type, Token については実際の発話内容（書き起こし）と音声認識システムの認識結果（Julius, AmiVoice）との間に強い正の相関がみられたが、50 回分の発話データ全体で検証すると、TTR スコアの相関は確認できなかった。一方、発話者ごとに TTR スコアの相関係数（表 5）をみると、5%水準で有意なものは一部のみであるが、特に AmiVoice を用いた認識結果については、5 名中 4 名は相関係数 0.4 以上（うち 2 名は 5%水準で有意）であり、正の相関が示された。

以上の結果から、音声認識システムを用いた言語能力の測定可能性について議論する。6.1 節で示したように、単純に TTR スコアのみを比較した場合、本来の TTR スコア（表 3 では「書き起こし」が相当）よりも高くなる可能性がある。したがって、単純にある閾値を下回ったかどうか、といった観点から、TTR スコアを認知症などの初期判断（4.1 節 用途 2）に用いることは難しい。ただし、本研究で、音声認識により各指標がどの程度バイアスを受けられるかが明らかになった。今後、これを補正することで、さらに精度の高い推定を実現できる可能性があると思われる。

一方、前述したように、同じ発話者に限定すれば、TTR スコアは本来の TTR スコアと正の相関がある可能性が示唆される。これまでの調査では、長期的に認知症者の言語能力スコアの変化を見ていくと、徐々に減少していく傾向が確認されている [8]。このような傾向を鑑みると、継続的に被測定者の TTR スコアを測定・記録し、その変化を見るという利用法であれば、認知症の疑いがあるかどうかの初期判断（4.1 節 用途 2）を行うことができる可能性がある。同様に、被測定者自身が現状把握・言語能力の維持・改善目的で継続的に言語能力スコアを計測・比較するといった利用方法（4.1 節 用途 1）も可能であると考えられる。

6.2.2 その他の指標の測定可能性

本節では、TTR 以外の指標の測定可能性について述べる。

5.1 節でも述べたが、認知症の進行と関連のある指標は TTR および JEL である [7]。また、我々が行った予備調査の結果、発達障害の判断においては、JEL, FPU（頻度・使用者数比）、NER（具体性）が関連する可能性があることが示唆されている。発達障害の初期判断においては、TTR 以外の指標の自動測定が望まれる。

TTR 以外の指標は、各語に紐づいた特有の値（例えば、FPU の場合は頻度・使用者数の比）や、品詞（NER の場合、固有名詞の数）に基づいてスコアが算出される。つまり、発話内容が正しく認識されなければ、正しいスコア算出が難しい指標であるといえる。今後音声認識の精度が向上し、正確な認識結果が得られるようになれば、TTR 以外の 5 指標についても自動測定が可能になると考えられる。

7. おわりに

本研究では、認知症や発達障害などにおける初期判断や自己把握・状況改善での利用を想定した、発話者の音声から言語能力を測定するシステム「言秤」を提案した。(1) 音声認識システムの組み込み、および(2) テキストデータから定量的に言語能力を測定する指標の採用を行うことで、従来人手で行っていたテキスト化および言語能力指標の算出を自動化し、コストの軽減と手軽な測定を実現した。ただし、テキスト化に音声認識システムを用いることで、正確な言語能力スコアの算出が困難となる可能性がある。

そこで、言語能力測定に関する「被測定者自身による自己把握・状況改善（用途 1）」および「被測定者以外による初期判断（用途 2）」という観点から、言語能力スコア算出における音声認識システムの利用可能性について検証を行った。検証実験の結果、以下の点を明らかにした。

(1) Type（異なり語数）、Token（延べ語数）は、発話内容と音声認識結果で強い正の相関がある。また、TTR スコア（Type・Token の比率）は、複数発話者の混在するデータにおいて相関はみられなかったが、同じ発話者の発話データであれば、正の相関がある傾向がみられた。

- (2) Type, Token, TTR スコアのいずれも, 相関関係はみられるものの, 実際の値には発話データとの差異がみられた.
- (3) 上記の (1), (2) より, 閾値との比較のような, 単純な言語能力スコアの対比による初期判断 (用途 2) は難しいが, 被測定者の言語能力スコアを継続的に測定し, その変化を観察することによる初期判断 (用途 1) や言語能力の現状把握・維持・改善 (用途 2) ができる可能性がある.

今回の検証においては, 評価用音声データとして, 音声認識することを想定していない録音音声データを用いた. ただし, 発話者に対し, 録音していることを伝えているため, 録音されることを意識している可能性はある. 今後, 録音を全く意識していない発話データでの検証を行い, 同様の傾向が得られるかを確認する. また, 今回用いた音声データには, 高齢者の音声データは含まれていないため, それを用いた検証も行う必要があると考えている. 今回の検証においては, TTR スコア以外の言語能力スコアを参考値として提示したが, 今後, 認識結果と併せた詳細な分析により, それらの測定可能性を検証する.

謝辞 評価に用いた音声データの収録・書き起こしにあたり, 国立国語研究所の加藤祥氏に多大なるご協力をいただいた. 音声認識システム AmiVoice SP2 は, アドバンスト・メディア社にご提供いただいた. ここに深く感謝の意を表する. 本研究は, JST 戦略的創造研究推進事業の助成による.

参考文献

- [1] 自閉症・発達障害児のための療育 55 段階プログラム【四谷学院】 | 自閉症・発達障害とは, <http://yotsuyagakuin-ryoiku.com/jiheisyou/>
- [2] Hampshire, A., Highfield, R.R., Parkin, B.L., et al.: Fractionating human intelligence, *Neuron*, Vol.76, No.6, pp.1225-1237 (2012).
- [3] Kemper, S., Marquis, J. and Thompson, M.: Longitudinal change in language production: effects of aging and dementia on grammatical complexity and propositional content, *Psychology and Aging*, Vol.16, No.4, pp.600-614 (2001).
- [4] Kubo, M., Kiyohara, Y., Kato, I., et al.: Trends in the incidence, mortality, and survival rate of cardiovascular disease in a Japanese community: the Hisayama study, *Stroke*, Vol.34, No.10, pp.2349-2354 (2003).
- [5] Snowdon, D.A., Kemper, S.J., Mortimer, J.A., et al.: Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study, *JAMA*, Vol.275, No.7, pp.528-532 (1996).
- [6] Dalgleish, T. and Werner-Seidler, A.: Disruptions in autobiographical memory processing in depression and the emergence of memory therapeutics, *Trends in Cognitive Sciences*, Vol.18, No.11, pp.596-604 (2014).
- [7] 荒牧英治, 久保圭, 四方朱子: 老いと <ことば> : ブログ・テキストから測る老化, *情報処理学会研究報告*, Vol.2014-DBS-159, No.23, pp.1-6 (2014).
- [8] 四方朱子, 荒牧英治: 言語能力検査としての言語処理: 長期間のブログ執筆を継続した認知症の 1 例, *言語処理学会第 20 回年次大会*, pp.1126-1129 (2014).
- [9] 呉田陽一, 伏見貴夫, 佐久間尚子: 言語能力の加齢変化, 第 9 回東京都老年学会誌, pp.200-205 (2002).
- [10] 厚生労働省研究班, 都市部における認知症有病率と認知症の生活機能障害への対応, http://www.tsukuba-psychiatry.com/wp-content/uploads/2013/06/H24Report_Part1.pdf
- [11] Ikeda, T., Ando, S., Satoh, K., et al.: Automatic Interpretation System Integrating Free-style Sentence Translation and Parallel Text Based Translation, *Proceedings of the ACL-02 Workshop on Speech-to-speech Translation: Algorithms and Systems*, Vol.7, pp.85-92 (2002).
- [12] 笹島宗彦, 井本和範, 下森大志ほか: 発話意図理解と回答誘導による異言語間会話支援ツール「グローバルコミュニケーター」, *インタラクシオン 2005 予稿集*, pp.119-126 (2005).
- [13] 花沢健, 荒川隆行, 岡部浩司ほか: 携帯電話試作機上で動作する旅行会話向け音声認識システム, *情報処理学会第 71 回全国大会*, 第 2 分冊, pp.39-40 (2009).
- [14] 加藤恒夫: 音声認識技術の実用化への取り組み: 2. 携帯電話における分散型音声認識システムの実用化, *情報処理学会誌*, Vol.51, No.11, pp.1394-1400 (2010).
- [15] 下郡信宏, 坪井創吾: 音声認識で生成した英語字幕による英語理解向上の測定実験, *情報処理学会論文誌*, Vol.51, No.9, pp.1951-1959 (2010).
- [16] 別所克人, 松永昭一, 大附克年ほか: 話題構造抽出に基づく会議音声インデクシングシステム, *電子情報通信学会論文誌. D, 情報・システム*, Vol.91, No.9, pp.2256-2267 (2008).
- [17] Kintsch, W. and Keenan, J.: Reading rate and retention as a function of the number of the propositions in the base structure of sentences, *Cognitive Psychology*, Vol.5, No.3, pp.257-274 (1973).
- [18] Turner, A. and Greene, E.: The Construction and Use of a Propositional Text Base, *Technical report 63*, Institute for the Study of Intellectual Behavior, pp.1-87 (1977).
- [19] Aramaki, E., Maskawa, S., Miyabe, M., et al.: A Word in a Dictionary is used by Numerous Users. In *Proceedings of International Joint Conference on Natural Language Processing (IJCNLP2013)*, pp.874-877 (2013).
- [20] 砂川有里子: 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*, Vol.24, pp.164-169 (2012).
- [21] 松吉俊, 佐藤理史, 宇津呂武仁: 日本語機能表現辞書の編纂, *自然言語処理*, Vol.14, No.5, p.123-146 (2007).
- [22] 松吉俊, 佐藤理史: 文体と難易度を制御可能な日本語機能表現の言い換え, *自然言語処理*, Vol.15, No.2, pp.75-99 (2008).
- [23] Daisuke, K. and Kurohashi, S.: A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis, In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL2006)*, pp.176-183 (2006).
- [24] 大武美保子: 認知症予防回復支援サービスの開発と忘却の科学 会話における思考の状態遷移モデルと会話相互作用量計測法の開発, *人工知能学会論文誌*, Vol.25, No.5, pp.662-669 (2010).
- [25] 保田祥, 田中弥生, 荒牧英治: 繰り返しにおける独話の変化, *社会言語科学会第 31 回大会発表論文集*, pp.190-193 (2013).
- [26] 河原達也, 李晃伸: 連続音声認識ソフトウェア Julius, *人工知能学会誌*, Vol.20, No.1, pp.41-49 (2005).