

聞き手の推定した語彙量と話し手の TTR との乖離に関する検討

岡久 太郎*† 柴田 大作‡ 伊藤 薫* 荒牧 英治*‡

* 奈良先端科学技術大学院大学 研究推進機構

† 京都大学大学院 人間・環境学研究科

‡ 奈良先端科学技術大学院大学 情報科学研究科

{taro-o, shibata.daisaku.rr8, kito, aramaki}@is.naist.jp

1 はじめに

近年、アルツハイマー型認知症（以降、アルツハイマー症と呼ぶ）などの精神疾患をスクリーニングする研究において、言語能力の測定はアルツハイマー症者と健常者を識別するための手法の一つとして注目されている [1, 2, 3]. 特に, TTR (Type-Token Ratio) はもっとも簡単に計測できる“語彙量”(vocabulary size) に対応した値として頻用されており, アルツハイマー症者は健常者に比べて TTR が低下するという報告がある [4], その一方で, 実際に TTR が低い語りをしているアルツハイマー症者であっても, 「認知症とは思えないほど語彙がしっかりしている」という関係者(家族など)からの声もしばしば現場からは聞かれる.

実際に, 我々が遭遇した事例は以下のようなものがあった:

「(私の父は) 認知症が進んでからも, ドイツ語の単語を沢山, しゃべっていた. 認知症患者の語彙量が減るというのは当てはまらない場合もあるのではないか?」

上の事例では, 使用している語の難易度から“語彙量”を考えている. このように, 介護者はアルツハイマー症当事者の“語彙量”は正常であると認識している場合が少なくない. このような事実は, “語彙量”という用語を専門外の人々が聞いて想起するものが, 単純に TTR と対応づけられないということを示唆している. 医療のような分野において, このような研究者と非研究者との間にある用語に対する認識のずれは重大なミスコミュニケーションを引き起こす可能性がある.

本研究では, 聞き手が推定した話し手の語彙量が, 実際の TTR を反映しているのかをクラウドソーシングによる調査によって検討した. 本稿の構成は以下の通りである. 2 節で今回の調査で材料として用いた高

齢者の作文コーパスについて詳細を述べ, 3 節では調査手法とその結果について述べる. そして, 5 節では, 調査結果の考察として考えられる 2 つの可能性について議論する.

2 材料

本稿では, 1 節で述べたように, アルツハイマー症の可能性のある話し手の語りを聞いた際に聞き手が推定する語彙量について検討しているため, 調査の呈示刺激として, 高齢者の作文コーパス¹を用いた. これは, 高齢者が「今までで一番嬉しかったこと」という課題に基づいて作文を行ったものを書き起こしたコーパスである.

今回は, 聞き手の語彙量推定に影響を与えうる要因として, TTR の他に Token 数を想定し, 全 30 個の作文データの中から TTR が大きい/小さい, Token が多い/少ないの 2×2 条件に当てはまる 8 データを選択した. 発話速度等の影響を統制するため, 音声読み上げソフトを用いて音声ファイルに変換した.² 調査で使用した 8 データの詳細を表 1 (時間は音声ファイルに変換した際のファイル長を示す) に, その一例を図 1 に示す. なお, TTR は異なり語数 (Type) を延べ語数 (Token) で除した値である. また, 表 1 の JEL は, 語ごとの語彙レベルから算出した全単語の平均値を表している. 語彙レベルは『日本語教育語彙表 ver1.0』³に収録されている 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) の値を用いた.

¹<http://sociocom.jp/jelico.html>

²音声ファイルへの変換を行う前に, 2 人の作業者がデータ中の個人情報 (人名や地名) の匿名化を行った.

³<http://jhlee.sakura.ne.jp/JEV.html>

表 1: 調査で使った作文データの詳細

ID	Token 数	Type 数	TTR	JEL	時間 (秒)
LL ₁	111	67	0.60	1.83	94
LL ₂	132	76	0.58	1.40	105
LH ₁	126	91	0.72	1.79	115
LH ₂	109	80	0.73	1.68	101
HL ₁	214	128	0.60	1.49	176
HL ₂	168	100	0.60	1.36	150
HH ₁	194	142	0.73	1.73	147
HH ₂	189	134	0.71	1.74	151

孫が中学校を選択するについてすばらしい決断をしたこと。
グレートジャッジメント
孫と両親は小学4年生になってから自分の将来を考え、
中学受験をすることにしました。
いくつかの進学したい学校を目標に、
学習塾の先生や家庭教師の指導を受けて苦手だった
教科も少しずつレベルアップしてきました。
6年生になってから進学希望の学校を1~2校に絞って
勉強していくべきところなかなか絞りきれず、混乱状態になりました。
この状態を何日までも続けていくことはできません。
孫、両親が何回も話し合いをしましたが希望校を絞ることはできませんでした。
ある日、孫から両親に話があり、その席で進学学校は自分で決めると言い、
両親はびっくりすると同時に孫の成長に驚きました。
そして孫は選んだ学校に無事進学でき、毎日元気に通学しています。

図 1: 作文データの例 (表 1 の LL₁)

3 手法

本研究では、Yahoo!クラウドソーシング⁴を用いて、表 1 の 8 つの音声発話を 100 名に聞かせ、使用された単語の量 (予測単語量)、話し手の知っている単語の量 (予測語彙量)、使用された単語の難易度 (予測単語難易度) の 3 つについて 4 段階で回答させた。呈示した設問の文章は以下の通りである。

- (1) 音声発話について、話し手が使った単語の量を推測し、以下から選択してください。(選択肢: 多い, やや多い, やや少ない, 少ない)
- (2) 音声発話について、語り手がどれだけの語彙を知っているかを推測し、以下から選択してください。(選択肢: 多い, やや多い, やや少ない, 少ない)
- (3) 音声発話について、話し手が使った単語の難しさを以下から選択してください。(選択肢: 難しい, やや難しい, やや簡単, 簡単)

なお、今回の調査に用いた音声発話は、使用単語の難

⁴<https://crowdsourcing.yahoo.co.jp/>



図 2: タスクの説明画面

易度については統制を行わなかったため、予測単語難易度の結果については扱わないものとする。

タスクは、2017年11月9日から11月10日にかけて実施し、100人のクラウドワーカーに対し、26の設問(うち2問はチェック質問)を目安時間20分として、5ポイント(5円相当)の報酬で依頼した。タスク画面を図2に示す。

4 結果

(1, 2)の設問に対する回答の上位2段階(「多い」「やや多い」)をHigh群, 下位2段階(「やや少ない」「少ない」)をLow群とし、実測Token数(多い/少ないの2水準), 実測TTR(大きい/小さいの2水準)の両指標との関係をクラメールの連関係数によって調べた。

各指標間のクラメールの連関係数の値を表2に示す。クラメールの連関係数は、2水準×2水準の場合、.10で小,.30で中,.50で大の効果量を表すと言われている[5]。これより、予測単語量, 予測語彙量はいずれも実測Token数, 実測TTRとの間には強い対応関係が認められないことが分かる。一方、設問の回答間においては、予測単語量と予測語彙量の間で連関係数が.49と大に近い効果量があった。

5 考察

調査によって、聞き手の推測する話し手の語彙量は実際のTTRの大小に影響を受けないことが示された。この結果が生じた理由として、以下の2つの可能性が考えられる。

表 2: 各指標間のクラメールの連関係数 (効果量 大: >.50; 中: >.30; 小: >.10)

	実測 Token 数	実測 TTR	予測単語量	予測語彙量
実測 Token 数	—	—	.003	0
実測 TTR	—	—	.084	.052
予測単語量	.003	.084	—	.49
予測語彙量	0	.052	.49	—

I. 語りを聞くだけではその TTR を正確に推定できない。

II. “語彙量” という用語がそもそも多義的であるため、聞き手の推定していたものが異なっていた。

以下では、これら 2 つの可能性をより詳細に見ていく。

5.1 聞き手による TTR 推定の困難さ

今回の調査結果において、聞き手の推測する語彙量は、実際の TTR とは対応していなかった。しかし、この結果から「聞き手は話し手の語彙量を推測する際に TTR を考慮していない」と結論付けることは早計である。

結果において示したように、予測単語量と予測語彙量の連関係数は .49 であり、これらの間には強い関係性が見られる。すなわち、聞き手は自身の感じた単語量に応じて、話し手の語彙量を推測した可能性がある。ただし、予測単語量が実測 Token 数と実測 TTR のいずれの指標にも対応していないため、予測された語彙量は実際の客観的指標とは乖離したものになっている。

今回の調査において、予測単語量は調査協力者に Token 数を推定させるために設定したものであったが、今回の結果を見る限り、彼らが本当に Token 数を推定していたかは定かでない。今回用いた材料は Token 数と発話時間が高い相関を示しており ($r = .97$)、音声の再生時間に基づいて単語量を推定すれば、ほぼ正しく単語量を予測することができた。そのため、調査協力者が単純に発話された単語の量を答えていたのであれば、実際の Token 数と予測単語量はもっと高い対応関係を示すはずである。

今回、単語量が Token 数と対応しなかった要因として、予測単語量を尋ねる設問が (1) であったことが考えられる。すなわち、「話し手の使った単語の量」という質問内容から、「話し手の使った単語の“種類”の量」を想定し、聞いた音声の時間長において使用される単語の平均的な Type 数と比較し、多いか少ないかを回答した可能性がある。この場合、調査協力者が回答し

たのは Token 数ではなく、聴取した音声の再生時間 TTR であると考えられ、予測単語量-実測 TTR 間の連関係数 (.084) が予測単語量-実測 Token 数間の連関係数 (.003) よりも大きいことが説明できる。すると、聞き手の語彙量推定が実測 TTR と乖離しているのは、聞き手が TTR と“語彙量”を別のものとして考えているためではなく、語りを聞くだけでは正しく TTR を把握できないためということになる。

以上の仮説を検証するためには、調査協力者に Token 数や TTR を正しく理解させた上で、より精緻な調査が必要になる。クラウドソーシングは、対面で行う調査とは異なり、調査協力者の管理が困難であるため、質問項目を出来る限り平易かつ簡素なものとする必要がある。今回の調査においては、「単語」の説明として、図 2 のような説明を加えたが、Type 数を数えるのか、Token 数を数えるのかといった指示は、質問が複雑になることを考え、行わなかった。⁵ 今後は対面式の調査においても今回と同様の結果が観察できるか追試を実施したい。

5.2 “語彙量” という用語の多義性

今回の調査結果の説明として考えられるもう一つの可能性は、“語彙量”，すなわち、話し手の使用できる語彙の豊富さという概念そのものが多義的であるため、調査協力者が回答した“語彙量”が個々人で異なっていたというものがある。

これまで本稿では、聞き手の推定した語彙量が TTR に対応するかという観点で議論を行ってきたが、そもそも“語彙量”が、ある文章内の異なり語数に基づくものであるという認識は、計量的研究においては一般的なものである [6] が、他の学問領域においては“語彙量”という概念は異なる理解がされることがある。一例として、第二言語学習の研究における“語彙量”は、ある学習者が使用している単語の難易度から、同程度

⁵このようにタスク内容を簡素化しても、クラウドソーシングにおいてノイズの混入を避けることは難しい。今回はチェック設問を導入し、予めノイズ対策を行ったものの、音声の再生回数を確認したところ、音声を聞かずに回答した参加者が複数人いたようである。

の難易度の単語は使用できるものと仮定し、そこから学習者が知っている単語の総量を推定し、「X 語レベル」という形式で表される。⁶

TTR は使用されている単語の難易度を考慮していないという点で、「X 語レベル」の形で表される“語彙量”とは本質的に異なるものである。そのため、今回の調査において、聞き手が使用されている単語難易度を考慮していた場合、そこから予測される話し手の語彙量は TTR に基づいて予測された語彙量とは大きく異なると考えられる。

また、仮に全ての聞き手が単語難易度のみから語彙量を推定していた場合であっても、推定される語彙量の大きさは個人間で異なる可能性が高い。なぜなら、単語難易度の決定には複数の要因が関わっているためである。第二言語学習の研究においても、単語難易度として扱われる指標は頻度や単語親密度等、手法毎に異なっており、単語難易度の決定方法が重要な問題となっている [7]。今回の調査において、聞き手が語彙量の推定に単語難易度を用いていた場合、その単語難易度は頻度や親密度のような複数の要因から算出されたものだと考えられる。そのため、聞き手ごとに、各要因の重み付け、それらの要因から算出される単語難易度が異なり、結果として最終的に推定される語彙量も大きく異なることが予測される。

このように、“語彙量”という概念は、その推定に単語難易度が考慮された場合、TTR のような客観的な指標とは乖離した結果になる。

6 おわりに

今回の調査結果から、聞き手が発話を聞いて直感的に予測する語彙量は、実際の TTR とは対応していないことが分かった。この理由として、5 節で指摘したように、聞き手が TTR を正しく予測できていない可能性と聞き手が想定している“語彙量”がそれぞれ異なっている可能性の 2 つが考えられた。当然、これらの可能性は排他的なものではなく、両立しうるものである。いずれにせよ、今回の結果は、アルツハイマー症者と普段会話をしている周囲の者が直感的に判断している語彙量と TTR のような客観的な指標との間の齟齬を再現するものとなった。

⁶ここでは単に“語彙量”と記述しているが、「使用されている単語の難易度から推定される」のは使用/発表語彙量と呼ばれるものである。これに対して、自発的に使用するかどうかには関わらず、見る、ないしは聞いて理解できる単語から推定される理解語彙量/受容語彙量と呼ばれる語彙量も第二言語学習では研究されている(詳細は [7, 8] 等を参照)。本稿では TTR に対応する語彙量が使用語彙量であることから、使用語彙量についてのみ言及する。

アルツハイマー症は治療が困難であるため、進行を遅らせるためにも早期発見が重要な課題となる。しかし、今回の結果が示すように、高齢者の周囲の人間は彼らの語彙量の変化を正確に把握することが難しいと考えられる。近年、先行研究において、アルツハイマー症の早期発見に寄与するべく、[9] のようなアプリケーションの開発が進められている。このような試みは、周囲の聞き手が正確に推定することが難しい語りの量的情報を手軽に、かつ利用者に分かりやすい形で提供することを可能にするものである。本研究の結果は、こういったアプリケーション開発の意義をより明確にすると同時に、我々研究者が測っている対象についてより自覚的になり、それを明確に説明する必要性があることを示している。今後、言語の計量的研究が誤解なく応用されることを祈念している。

参考文献

- [1] David A Snowdon, Susan J Kemper, James A Mortimer, Lydia H Greiner, David R Wekstein, and William R Markesbery. 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.
- [2] William Jarrold, Bart Peintner, David Wilkins, Dimitra Vergyi, Colleen Richey, Maria Luisa Gorno-Tempini, and Jennifer Ogar. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In *Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology*, pp. 27–36, 2014.
- [3] Kathleen C Fraser, Jed A Meltzer, and Frank Rudzicz. Linguistic features identify alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, Vol. 49, No. 2, pp. 407–422, 2016.
- [4] 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, 2016.
- [5] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. New Jersey: Lawrence Erlbaum Associates, 2nd edition, 1988.
- [6] 計量国語学会 (編集). 『計量国語学事典』. 東京: 朝倉書店, 2009.
- [7] Norbert Schmitt. *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan, 2010.
- [8] I.S.P. Nation. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press, 2013.
- [9] 柴田大作, 伊藤薫, 若宮翔子, 宮部真衣, 木下彩栄, 荒牧英治. 「自由発話による認知症スクリーニングを支援するアプリケーションの開発」. 第 37 回医療情報学連合大会, 2017.

謝辞

本研究の一部は、JSPS 科研費 JP16H06395, JP16H06399 の支援を受けたものです。