

自然言語処理で話者の能力を測る

荒牧英治* ** 若宮翔子* 四方朱子* 木下彩栄***

*奈良先端科学技術大学院大学 研究推進機構

**科学技術振興機構 さきがけ

*** 京都大学大学院 医学系研究科

{aramaki,wakamiya,shikata}@is.naist.jp kinoshita.ayae.6v@kyoto-u.ac.jp

1. はじめに

これまで自然言語処理は、様々な処理階層でテキスト処理精度を高めてきた。この結果、新聞などの文法的な文章に関しては高い精度を実現し、機械翻訳などの応用技術も実用化に近い段階に近づいてきた。しかし、Twitter などのソーシャルメディア上の文章は会話文、方言などのくだけた表現や非文法的なテキストの解析精度は、まだ比較的低い精度にとどまる。つまり、文法的なテキストは高い精度で解析でき、非文法的なテキストについては比較的低い精度にとどまる段階である。この性質を利用すると、解析の難易度から、対象となるテキストがどの程度文法的かどうかを測定することができる可能性がある。

例えば、構文解析は長い係り受けが頻出すると解析が困難であるが、言い換えれば、長い係り受けが稀だからであり、それを多用する記述者の日本語能力に問題がある可能性がある。未知語についても同様で、未知語が多い場合には、話者の言語能力に問題がある可能性がある。

このように考えると、言語能力の測定は、形態素解析や構文解析と異なり、解析結果ではなく、解析の難しさを知るタスクとして解決できる可能性がある。処理の難しさを推定することは、正しい処理結果を出力する問題のサブセットとみなせるため、これまでの解析器を用いて自然に扱える。また、その社会的要請の高さから実用性も高く、今後の新しい言語処理の一分野となる可能性がある。

以上の背景のもと、我々は、テキスト産出者の言語能力を推定することを試みている。これまで、源氏物語やシェークスピアの作品など古典の著者推定など限られた目的のもと語彙量などいくつかの指標が用いられたことはあったが、品詞の割合など比較的単純な統計によるところが大きかった。本研究では、言語処理の難しさに相当する係り受け距離や語の標準性といった言語処理との整合性を考慮した指標を導入し、話者の言語能力を推定する。

ここで、言語能力を測定の妥当性を示すためには、能力の差がある研究協力者による共通したタスクが必要となる。本研究では、アルツハイマー患者識別タスク、アルツハイマー型認知症者(中程度の認知能力低下)と健常者群を話し言葉から識別するタスク、という応用性も高いタスクを用いる。アルツハイマー型認知症は認知症の約半数をしめ、症状が進行すれば、言語能力、特に語彙量、の低下が報告されてる。同時に、他の言語能力、構文能力や語彙の質では健常者との区別が難しいと報告されている [1]。本研究はこれを言語処理に基づく指標で再現することを目指す。

予備的な実験(n=17)の結果、アルツハイマー型認知症患者と健常者の間に有意な語彙量の差がある一方、語彙の標準性、係り受け距離に関して有意差は得られなかった。これは、先行研究の報告と合致しており、提案する測定法の可能性を支持する結果を得たので報告する。

2. 関連研究

2.1 認知症と言語

世界に先駆けて高齢化社会を迎える日本にとって、認知症は医療費が 10 兆円規模と算定され(平成 22 年 厚生労働省)、国家的問題となっている。現在、予防または進行抑制の確立に向け、大規模な臨床試験が世界各地で開始されつつあり、いかに認知症を早期発見するかが今後重要な課題となる。これまでも、認知症者の語りにおける語彙量からスクリーニングを行う研究はあった。英語母語話者についての先行研究 [1] によれば、健康な高齢者においては、語彙能力は維持される一方、認知症患者においては急激に低下する。構文能力については、認知症者であるかどうかにかかわらず、概ね 70 歳を堺に低下し、構文的観点から区別はできないとされている。何をもちて語彙能力や構文能力とするかは、いくつかバリエーションがあるが、以下のようにまとめられる。

2.2 言語能力

言語能力は、大きく構文能力と語彙能力の2つに大別される(場合によって、コミュニケーション能力といった語用論的な観点も加えて3分類とする場合もある)。

構文能力については大別すると2つの尺度がある。第一に、文法特徴を経験則によって捉えるもの、第二に、構文解析結果を用いるものがある。前者の代表的なものがD-levelである。D-Level [2]は、文の複雑さを8つのレベルに分割した尺度であり、それぞれのレベルは、経験則によって定められた文法的特徴の有無によって判定される [3]。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。

後者は、構文解析器の出力を利用し、そこから得られる統計量をもとに、構文の複雑さを判定する。Frazier等は、構文木の各ノードから根ノードまでの距離を構文の複雑さとみなした [4]。これはFrazier scoreと呼ばれる。Yngveは構文木の分岐数を文の複雑さとみなした [5]。これはYngve scoreと呼ばれる。これら2つのスコアは、句構造に依存している。句構造を用いないものとして、Linの依存構造上の係り受け距離を用いるものがある [6]。本研究では、Linの係り受け距離を用いる。この係り受け距離は、構文解析の難しさに相当していると考えられる。

語彙に関する能力は量、質、アクセスの速さ(流暢さ)からなるとされる [7]。語彙の量は、理解できる、または、使用できる語彙の量を指す。しかし、これを正確に測定することは難しく、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。本研究でも、TTR (Type Token Ratio)を用いる。

一方、語彙の質は、語彙に関する理解の深さや語彙の難しさに測られる。理解の深さに関しては、ある語について知っている上位語や下位語の数 [8]や、多義語の曖昧性解消能力などで判別できる [9]。これを近似的に自動計測するため、文章中に含まれる平均の語彙の難しさ [10]と語の一般性 [11]を用いた。語の一般性は、形態素解析や曖昧性解消など形態素に関連する処理の難しさに相当すると考えられる。

最後に、単語へのアクセスの速さは、命名課題によって調査される。本研究では、単位時間で発話した単語数を用いる。

3. コーパス

材料となる認知症者の語りは、病院を受診中の患者が、心理検査時に話す内容を収録し、書き起こしたものを用いた。

3.1 フィールド

実験協力者のリクルートは、京都大学医学部附属病院神経内科にて以下の基準で行った。

【包含基準】

- 認知症群 (AD; Alzheimer's Disease) : レベル軽度～中等度のアルツハイマー型認知症患者

表1: 実験協力者

	AD (MMSE21点以下)	nonAD (MMSE22点以上)
性別	男性: 1名; 女性: 7名	男性: 4名; 女性: 5名
年齢	77.2歳	76.6歳
MMSE Score	17.0	25.1

表2: コーパス例 (ADグループの語りの抜粋)

あー 体調は一別にもないです
物をよう忘れるだけで
ほんで足がちょっと悪いのでね
はい手術して足が悪いね
あの一 まあだいたい かがと かがとがつかないんですよ
今あの一 痛んでてね
でこー 今つま先だけで今ちょっと歩いてるからね
そ そうですね ほんでちょっとあのよっぽど痛いときは ま
一 松葉杖ついたりするんですけどね まー今日は娘が
自動車乗せてってくれたからね
もうつま先だけでね この病院内を歩くぐらいやったらね
どうもないから つま先だけで歩いてます

- 対照者群 (nonAD) : 認知症でないと確認でき、中枢神経系に異常を認めない、認知症群と年齢をマッチさせた群

【除外基準】

中枢神経系の代謝に影響を及ぼすような疾患を持つ者、病状などにより十分な同意能力を持たない者。非日本語母語話者。

3.2 認知能力評価

認知症患者かどうかの判定はMMSE (Mini Mental State Examination)に基づき行った。MMSEとは、1975年に考案された知能検査であり、アルツハイマー症などの疑いがある実験協力者のために作られた簡便な検査方法である。実験協力者に対し口頭による11問の質問形式(各質問に点数があり、30点満点で判定)である。点数の解釈は、27～30点が正常、22～26点が軽度認知症の疑いあり、21点以下が認知症の疑いが強いと判定される。

3.3 統計

実験協力者の統計を表1に示す。コーパスの例を表2に示す。

4. 言語尺度

本研究では次の5つの言語尺度を用いた。表3に用いた言語尺度を示す。

4.1 Type Token Ratio (TTR)

前述のTypeとTokenの比率 (Type/Token)を集計する。この値が大きいほど語彙量が多いと考えられる。

TTR=Type/Token

4.2 Token Per Second (TPS)

1 秒当たりの発話単語数 (TOKEN).

TPS=Token/発話時間(秒)

4.3 Lexical Education Level (LEL)

単語の難易度. 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[12]. 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.

LEL=中級以上の名詞数/名詞数

4.4 Average Word User (AWU)

平均の単語ユーザ数. 語のユーザ数は文献 [11] の値を用いた. この値は, 話者の語彙の標準性を示している.

$$AWU = \sum_{w \in W} WU(w)$$

ここで, W は文に含まれる単語集合, $WU(w)$ はある単語 w のユーザ数とする.

4.5 Average Dependency Distance (ADD)

平均の係り受け距離を計算した. この値は構文解析の難しさに相当する.

$$ADD = \sum_{p \in P} DD(p)$$

ここで, p は文に含まれる文節の集合, $DD(p)$ はあるフレーズ p とかかり先との間のフレーズ数(ギャップ)とする. 例えば, 図 1 の例では 5 個の係り受けがあり, そのうち 1 つのギャップがある. よって, $ADD=0.2$ (1/5) となる.

5. 実験

5.1 デザイン

MMSE によって, 発話した語数から認知症となった群を推定した. MMSE は 30 点満点中 21 点以下だと認知症の可能性がある. 評価は, t 検定によって行った.

5.2 結果

結果を表 3 に示す. 発話数と種類の比 (TTR) は, 患者群で有意に少なくなっており, 疾患が語彙の小ささとの関連を示している. 語彙へのアクセス速度 (一秒あたりの発話速度 (TPS)) や語彙レベル (LEL), 語彙の一般性 (AWU) といった他の語彙能力に差はない. 構文能力 (ADD) についても差はない.

5.3 考察

アルツハイマー症の判別は語彙量 (TTR) のみであった. これは先行研究で報告されている事実, 構文能力は年齢とともに低下し, 健常者と疾患群での差はないが, 語彙能力は健常者で保たれており,

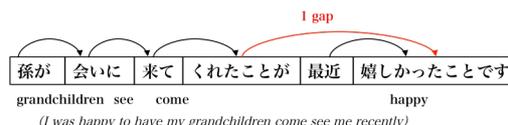


図 1: 係り受け距離.

表 3: 言語尺度.

尺度	説明
TTR	語彙量. 発話語数・発話種類数比 (Type Token Ratio; TTR).
TPS	語彙アクセス速度. 発話速度 (Token Per Second; TPS).
LEL	語彙の質. 語彙教育難易度 (Lexical Education Level; LEL).
AWU	語彙の標準性. 語彙平均使用者数 (Average Word User; AWU).
ADD	構文の質. 平均係り受け距離 (Average Dependency Distance; ADD).

表 4: 実験結果.

	AD	nonAD	p value
TTR	0.249 (0.054)	0.313 (0.052)	0.02
TPS	1.91 (0.99)	2.68 (0.37)	0.33
LEL	2.171 (0.157)	2.271 (0.135)	0.16
ADD	0.661 (0.20)	0.704 (0.182)	0.64
AWU	69736.1 (5877.8)	62430.4 (5578.2)	0.33

* 下線は AD と nonAD の間に有意な差があることを示す.

アルツハイマー発症認知症の特徴となる[1], と合致し, 本研究は, これを日本語にて再確認した.

なお, アルツハイマー症の識別には使えないものの, MMSE スコアと相関の高い指標はいくつかあり, TTR は相関係数 0.47, LEL は相関係数 -0.50, ADD は相関係数 0.33 を示している (図 2). アルツハイマー症かどうかの境界線は恣意的なものであり, アルツハイマー症の定義次第では, 他の指標も関係する疾患の識別に寄与する可能性がある. 今後, これらについて, 正確に言語能力を測れているのか, また, 疾患と関連した新たな言語能力があるのか, さらに, 言語能力から疾患概念を再定義できるのか, など多くの課題が残されている.

本研究には限界がある. 最も大きな限界の1つがサンプルサイズである. 例えば, アルツハイマー群には女性が多く含まれ (8 名中 7 人), これがバイアスとなっている可能性がある. 今後, より大規模な検証が必要となる.

もう一つの限界は一時点だけの測定である点である. 言語に関する指標は話題に影響されやすく,

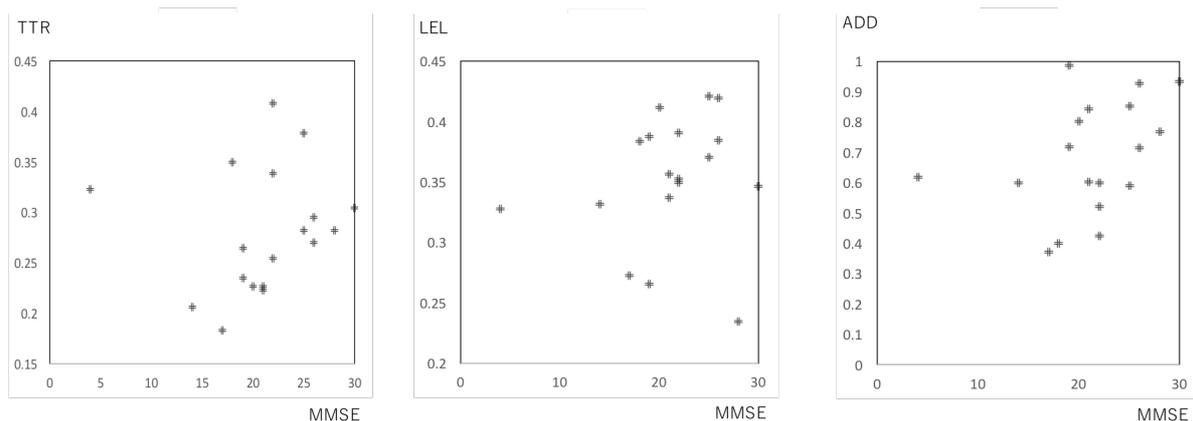


図 2: MMSE と各指標の相関. TTR(左). LEL(中). ADD(右).

また、個人差もある。同一人物を長期間観察し、同一人物内での変化をみるのが今後の課題である。

以上のような問題は、あるものの認知症者における話し言葉の非典型性を語彙量にて検出できた点は大きな知見である。

6. おわりに

本研究では、テキスト産出者の言語能力を推定することを試みた。アルツハイマー患者識別タスク、アルツハイマー型認知症者(中程度の認知能力低下)と健常者群を話し言葉から識別するタスクにおいて、予備的な実験(n=17)の結果、語彙量によって両者の区別が可能であった。これにより、言語処理によって、認知能力を測定できる可能性、および、アルツハイマー症の診断支援が実現できる可能性が示唆された。言語処理が言語能力を推定できることは今後の新しい研究領域の開拓へのさきがけとなると信じている。

参考文献

1. Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, Markesbery WR. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA : the journal of the American Medical Association*. 1996;275(7):528-32. PubMed PMID: 8606473.
2. Covington MA, He C, Brown C, Naçi L, Brown J. How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level scale. 2006.
3. Rosenberg S, Abbeduto L. Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics*. 1987;8(19-32).
4. Frazier L. Syntactic complexity. In: Dowty LK, Zwicky AM, editors. *Natural Language Parsing*: Cambridge University Press; 1985.
5. Yngve VH. A model and an hypothesis for language structure. *The American Philosophical Society* 1960. p. 444-66.
6. Lin D. On structural complexity. *COLING96* 1996.
7. Daller H, Milton J, Treffers-Daller J. *Modeling and assessing vocabulary knowledge*: Cambridge University Press; 2007.
8. Crossley S, Salsbury T, McNamara D. Predicting the proficiency level of language learners using lexical indices. *Language Testing*. 2012;29:243-63.
9. Verspoor M, Lowie W. Making sense of polysemous words. *Language Learning*. 2003;53:547-86.
10. Aramaki E, Shikata S, Miyabe M, Usuda Y, Asada K, Ayaya S, et al. Understanding the relationship between social cognition and word difficulty: a language based analysis of individuals with autism spectrum disorder. *Methods of Information in Medicine*, 2015. 2015.
11. Aramaki E, Maskawa S, Miyabe M, Morita M, Yasuda S, editors. *A Word in a Dictionary is used by Numerous Users*. International Joint Conference on Natural Language Processing (IJCNLP2013); 2013; Japan.
12. 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.