

高次脳機能障害スクリーニング検査支援ツールの開発

山本 苑美† 小西 慶和‡ 中島 陽子† 本間 宏利† 山本 和英*

† 釧路高専 情報工学科

‡ 釧路高専 電子情報システム工学専攻

* 長岡技術科学大学 電気系

yoko, honma@info.kushiro-ct.ac.jp

1 はじめに

知覚, 記憶, 学習, 思考, 判断などの認知過程, および, 行為の感情を含めた精神機能の総称は高次脳機能とよばれる. 脳血管障害や交通事故などによって脳が損傷したために, これらの認知機能に障害が起きた状態を高次脳機能障害という. 日本の高次脳機能障害者数については, さまざまな統計推定研究がされており, 2001年度から5年間行われた調査では, 全国の高次脳機能障害者数は約27万人と推定された [4]. また, 2008年に東京都で実施された調査によると, 全国の高次脳機能障害者数は約50万人と推測された [5]. 同じく, 2008年に福岡県で行なわれた調査では, リハビリテーションにより社会復帰をめざす中等度障害の高次脳機能障害者数は, 約7万人と推定されている [6].

高次脳機能障害を発症した患者の多くは, 治療と同時に急性期から, 病院やリハビリ施設にてリハビリテーションによる機能回復を目指す. 発症時の障害度やリハビリテーションによる回復度の判定に使用されるスクリーニング検査として, MMSE (Mini-Mental State Examination) [2] と HDS-R (Hasegawa Dementia Scale-Revised) [1] が非常に有名であり, 現在, 日本のほとんどの病院やリハビリ施設で, これらのどちらかのツールが採用されている. MMSE や HDS-R は, 神経内科にてのアルツハイマー病や認知症の判定, さらに, 入院や手術という特殊な状況下に置かれることを機に認知障害が発症するケースも多いことから, 外科, 整形外科, 内科など多くの診療科で採用されている [3].

MMSE, HDS-R は双方とも見当識, 記憶力, 計算力, 言語的能力等の検査項目から構成されており, 多くの類似した設問も含まれている. たとえば, 「三単語復唱」という検査項目では, 被験者に相互に無関係な3つの単語名を憶えさせ, 他の検査項目を挟んだ一

定時間後に, 先ほど記憶した3つの単語名をそのまま復唱させて, 記憶能力を診断する. この検査項目は, MMSE, HDS-R の双方に含まれている.

本研究では, これらのスクリーニング検査の共通項目となっている「三単語復唱」について, その検査実態に関してのいくつかの問題点を指摘し, それらを改善すべく, 自然言語処理技術を活用したスクリーニング検査支援ツールの提案を行う.

2 現状の問題点

MMSE や HDS-R のスクリーニング検査は, 外科や整形外科などの診療科では, 入院時や手術後に1, 2度だけ実施されることが通例である. 一方, リハビリテーション科や神経内科では, リハビリテーションや投薬コントロールの効果判定のため, これらの検査を定期的に繰り返して施行している.

このため, 「三単語復唱」検査では, 被験者が3つの単語を完全に暗記してしまうケースも多く, 記憶能力を正しく診断しているとはいえない状況が生じている. また, これは検査環境の問題ではあるが, 同じ時間帯に1つ居室にて複数人が検査を受診する場面も多く, この場合は, 他者の回答音声が入り交わらうため, 被験者本人の記憶との判断が正しくできない場合もある. 実際に, 現在も多くの病院で, MMSE では「桜・猫・電車」, HDS-R では「みかん・汽車・27」の三単語が長い期間, 変わらずに使用されている.

このような状況に疑問を持つ医者や療法士は実際に少なくない. 中には, 被験者である患者自身から疑問の声も聞くこともある. しかしながら, 「三単語復唱」検査で療法士が用いる3つの単語は, MMSE, HDS-R の双方で厳密に指定されており, これらを病院や療法士の側で勝手に変更することは, 検査結果の信頼性に問題が生じかねないため, 一般的に認められていない.

この検査で用いられている3つの単語は、以下の規則によって生成されていることが知られている。

- 性別・年代に関係なくよく知られていること。
- 短い音数であること。
- 単語間の連想性・相似性が低いこと。

よって、これらの規則を厳守しつつ、様々な種類、パターンの「三単語」を自動生成することで、これらの問題点の解決を試みる。

3 解決方法

「三単語復唱」検査で用いられる3つの単語の生成規則を遵守しつつ、それらの生成を実現させるため、下記のような方法を適用した。

3.1 よく知られている単語の抽出

スクリーニング検査の被験者は、一般的に高齢者が多く、「三単語復唱」で用いる単語には新造語や流行語は適切ではない。これに対し、童話や童謡は、対象が幼児、児童である故に、そこに出現する単語の多くは、あらゆる年代の人間に親しみ深く、非常に認知度が高い。そこで、本研究では童話や童謡歌詞を題材とし、それらに出現頻度の高い名詞を抽出して、「三単語」の候補とすることにした。

本研究では、日本の童話210編、童謡90曲¹をコーパスとして用い、そこから、出現頻度の高い名詞を自動抽出した。抽出された名詞を日本語語彙大系²の一般名詞意味属性体系に準拠させて「食べ物」、「乗り物」、「植物」、「動物」、「建築物」、「日用品」、「スポーツ」、「空想物」の8つのカテゴリに分類して格納する。これら8つのカテゴリは、互いに意味属性体系上の包含関係が生じないように定めた。この作業の終了後、「三単語」に適さないと思われる名詞を手動で削除し、さらに、各カテゴリに、適切と思われる名詞を適量ずつ追加した。

3.2 短い音数の単語の選択

上記処理で抽出された単語から、音数が5音以上の単語を全て排除する。これは、これまでの「三単語復唱」検査で使用されてきた単語の音数のほとんどが、4以下であったことを根拠としている。

¹<http://www.douwa-douyou.jp>.

²<http://www.kecl.ntt.co.jp/icl/lirg/resources/GoiTaikei/>.

3.3 単語間の連想性が低いこと

単語間の連想性とは、2つの単語A, Bに対して、A(B)が与えられたときのB(A)へのイメージの結びつきやすさ、をいう。実際に、MMSEで採用されている「桜・猫・電車」の三単語を例にすると、桜⇔猫、猫⇔電車、桜⇔電車のそれぞれの連想性は非常に低いことがわかる。しかし、「猫・こたつ・みかん」の三単語の組み合わせ例では、それぞれの単語が、異なる性質のカテゴリに分類される名詞であるにもかかわらず、非常に連想が容易である、つまり、記憶が容易である組み合わせになっている。

この問題の対策を以下に示す。

(1) 三単語候補の抽出

8個の単語カテゴリから任意に3個を選び、さらに、その3個のカテゴリから、それぞれ、任意に1つずつの単語を抽出して構成された3つの単語組を「三単語」候補とする。

(2) 連想語辞書による棄却

抽出された3つの単語に対して、以下の2つの検査を施行し、両方をクリアしなければ、その単語組は連想性が低くないとして「三単語」候補から棄却される。連想語辞書としてrenso³を使用する。

(TEST 1) 3つの単語を「A」、「B」、「C」とする。

- 単語「A」の連想語に「B」も「C」も含まれない。
 - 単語「B」の連想語に「A」も「C」も含まれない。
 - 単語「C」の連想語に「A」も「B」も含まれない。
- 上記の全てが満たされている場合のみ、クリアとする。

(TEST 2) 3つの単語を「A」、「B」、「C」とする。

- 「A」と「B」の連想語に共通要素が存在しない。
 - 「B」と「C」の連想語に共通要素が存在しない。
 - 「A」と「C」の連想語に共通要素が存在しない。
- 上記の全てが満たされている場合のみ、クリアとする。

TEST 1について、三単語:「猫・こたつ・みかん」を例に説明する。「こたつ」という単語を連想語辞書で検索したとき、そこに現れる連想語群中に「みかん」という単語が含まれていたとする。これは、「こたつ」という名詞から、直接「みかん」を連想しやすいという性質を含んでいることを示している。このように、3つの単語の中のどれか2つの組み合わせで、連想性が高いと判定された場合には、その3つの単語は「三単語」候補から破棄する。

TEST 2について、三単語:「猫・雪・みかん」を例に説明する。「雪」と「みかん」の単語の連想語群の両方に「冬」という単語が含まれていたとする。これは「雪」と「みかん」は「冬」という他の語を介して

³<http://www.comedywaltz.com/search/renso/>.

連想しやすい、結びつきやすいという性質を含んでいることを示している。このように、3つの単語の中のそれぞれの連想語群に、他のものと1つでも重複するものがあれば、その3つの単語は「三単語」候補から破棄する。

(3) 共起率判定による破棄

連想性の強さの判定は、連想語辞書のみでは不十分な場合がある。たとえば、三単語:「ハチ・柿・臼」の例を考える。これらの名詞を個々に調べても、単語間の意味的關係や連想性は低いであろう。しかし、童話「さるかに合戦」を知っている者ならば、この三単語は非常に記憶しやすく感じるだろう。

この対策として、共起率による連想性の判定を用いる。共起率算出の対象とするコーパスに、単語抽出でも利用した日本の童話 210 編と童謡 90 曲を用いる。2つの単語 A, B の少なくともどちらか一方を含むテキスト数を $n(A \cup B)$ 、A, B の両方を含むテキスト数を $n(A \cap B)$ としたとき、これらの単語間の共起率 ($0 \leq Cor(A, B) \leq 1$) は (1) 式にて算出される。

$$Cor(A, B) = \frac{n(A \cap B)}{n(A \cup B)} \quad (1)$$

なお、共起率算出時に用いた形態素解析ツールには MeCab⁴を使用した。三単語候補として選ばれた3つの単語のそれぞれの2組に対して共起率を算出し、それらの値に、1つでも0.0以外のものがあれば、童話や童謡の内容から連想される可能性があるものとみなして、「三単語」候補から破棄する。

(4) 類似度判定による破棄

「みかん」は植物でもあり、食べ物でもある。同様に「馬」は動物でもあり、乗り物、あるいは、食べ物でもある。たとえば、三単語:「自動車・馬・刺身」を例に考える。これらの3つの単語は、それぞれ乗り物カテゴリ、動物カテゴリ、食べ物カテゴリから選択された単語であるが「馬」を乗り物とみなす人にとっては、乗り物カテゴリから複数の意味的に似た単語が選択されているように感じるかもしれない。また「馬」を食べ物とみなす人にとっては、食べ物カテゴリから複数の単語が選択されているように見えるだろう。これらの例のように、連想性は各人の受け取り方に依存する場合もある。

この対策として、類似度による破棄判定を行う。三単語候補として選ばれた3つの単語のそれぞれの2組に対して、類似度を導出する。類似度計算は日本語彙大系の一般名詞意味属性体系分類をベースに行う。2つの単語 A, B の類義語辞書での根からの深さを

をそれぞれ d_A, d_B 、2つの単語の共通の上位語の根からの深さを d_C としたとき、これらの単語間の類似度 ($0 \leq Sim(A, B) \leq 1$) は (2) 式にて算出される。

$$Sim(A, B) = \frac{d_C \times 2}{d_A + d_B} \quad (2)$$

ここで、「自動車」と「馬」の類似度を導出する場合、「馬」は乗り物、動物、食べ物など、複数の意味属性に存在する。このとき、「自動車」と乗り物の「馬」、「自動車」と動物の「馬」、「自動車」と食べ物の「馬」の全ての類似度を求め、その中のもっとも類似度の高い値を「自動車」と「馬」との類似度として用いる。

3つの単語のそれぞれの2組の類似度が、1つでも0.4以上であれば、それらは被験者の受け取り方によって、連想性が高くなる可能性があることから、「三単語」の候補から破棄する。

(5) 第一音節（最終音節）が同音の組みの破棄

第一音節または最終音節が同音の単語は発音の相似性が高く、被験者に記憶されやすいことから、「三単語」にはふさわしいとはいえない。たとえば、三単語:「みかん・ミズ・ミシン」や「やかん、図鑑、ようかん」のように第一音節または最終音節が同一である場合は、それらの単語間の連想性が高くなくても、非常に記憶しやすくなってしまふ。

この対策として、三単語候補として選ばれた3つの単語は、全て、始音、終音が異なるものとし、それ以外は破棄する。

(6) 全単語のモーラ数が等しければ破棄

3つの単語のモーラ数が全て等しい場合、発声リズムが単調になり、相似性も高くなると思われるため、このような組み合わせは破棄する。

4 評価結果

本システムで、300組の三単語組を出力した。その一例を表1に示す。

これらのスクリーニング検査への実用適正を現役の言語聴覚士に判定していただいた結果、約96.7%が実際の検査にて実用可能であるとの高い評価を得た。

出力結果に不適切なものが含まれていた原因としては、共起率を導出する際に使用される形態素解析ツール MeCab の精度に限界があることや、類義語辞典の名詞意味属性体系分類と人間が潜在的に持つ分類の感覚に差があることで、類似度判定の精度に限界があったことが考えられる。たとえば、三単語「椅子・学校・すすき」の例では、「学校」と「椅子」は、連想しやす

⁴<http://mecab.sourceforge.net/>.

表 1: 「三単語」出力例

「三単語」出力例		
アサガオ	いど	傘
あじさい	駅	馬車
あじさい	砂糖	床屋
いど	きなこ	虎
学校	狸	杖
気球	椿	寺
きつね	天狗	ヨット
さる	床屋	馬車
車	天狗	百合
財布	寺	猫
アヤメ	天狗	マッチ

く不適切な出力である。我々の手法では、「学校」と「椅子」は、たとえば「机」という語を共通の連想語に持つことから、破棄されると期待していたが、今回の連想語辞書では識別できなかった。また、今回は「魔女」や「妖精」という単語が童話での出現頻度が高いという根拠から、空想物カテゴリーに追加したが、これらの単語は洋風的な空想物であることから、高齢者に理解されにくいとの指摘を受けた。よって、単語エントリーの条件をさらに厳しく吟味する必要がある。

5 おわりに

本研究では、高次脳機能障害の障害度判定のためのスクリーニング検査支援ツールの開発を行った。本システムは三単語復唱テストに必要な、意味属性的に独立であり、かつ、連想性の低い3つの単語を自動的に生成するものである。この手法で得られた中の約96.7%が、実際の検査にて十分実用に耐えられる精度であるとの高い評価を得た。さらに、今回の研究では、候補として抽出された三単語の中のいずれかの2単語間の類似度が0.4以上の場合、被験者に意味的に近い単語同士として認識される危険を防ぐために、その三単語を破棄していた。このことにより、実用性の精度は向上が高くなる反面、三単語に抽出されるカテゴリーや単語に、ある程度の偏りが生じることが見られた。今後は、これらの課題を克服して、より精度の高い結果出力を目指したい。

謝辞

本研究を進めるにあたり、ご協力を頂いた釧路孝仁会記念病院リハビリテーション科、佐野昌子言語聴覚士、ならびに、多くのセラピストの方々に深く感謝致

します。また、本研究は高専-長岡技科大共同研究の助成(2014)を受けて行なわれました。

参考文献

- [1] Y. Imai and K. Hasegawa. The Revised Hasegawa's Dementia Scale (HDS-R) - Evaluation of its usefulness as a screening test for dementia. *Journal of Hong Kong College of Psychiatrists*, Vol. 4, No. 2, 1994.
- [2] K. I. Shulman and A. Feinstein. *Quick Cognitive Screening for Clinicians: Clock-drawing and Other Brief Tests*. CRC Press, 2003.
- [3] R. Tsukamoto, T. Akisaki, and M. Kuranaga. Hasegawa Dementia Scale - Revised, for screening of early alzheimer's disease in the elderly with type 2 diabetes. *Geriatrics & Gerontology International*, Vol. 9, No. 2, pp. 213-215, 2009.
- [4] 中島八十一, 寺島彰. 高次脳機能障害ハンドブック. 医学書院, 2006.
- [5] 渡邊修, 山口武兼, 橋本圭司, 猪口雄二, 菅原誠. 東京都における高次脳機能障害者総数の推計. *日本リハビリテーション医学会誌*, Vol. 46, No. 2, pp. 118-125, 2009.
- [6] 蜂須賀研二, 加藤徳明, 岩永勝, 岡崎哲也. 日本の高次脳機能障害者の発症数. *高次脳機能研究*, Vol. 31, No. 2, pp. 143-150, 2011.