

---

論文

---

## 医学語彙テストの開発と評価

稻田朋晃<sup>1</sup>, 品川なぎさ<sup>2</sup>, 丸山岳彦<sup>3</sup>, 松下達彦<sup>4</sup>

近年、医療分野の留学生が増え、医療語彙教育の重要さが認識されつつあるが、学習者の習得状況を測定するためのテストは管見の限りない。そこで、筆者らは医学語彙テストを開発・実施し、その有用性を評価した。対象語の選択にあたっては、『イヤーノート 2020 内科・外科編』のテキストデータを、独自に開発した「医学語彙辞書」を用いて形態素解析を行い、頻度付きリストを作成した。高頻度層、中頻度層、低頻度層から 24 語ずつ、計 72 語を抽出し、対象語と平易な日本語で書かれた定義文をマッチングさせる形式のテストを作成した。医学部留学生 18 名に対してテストを実施し、結果を分析した。信頼性は  $\alpha = 0.919$  と極めて高かった。頻度と正答率の関係については、高頻度層の項目の正答率が低頻度層の項目よりも有意に高く、全体としては頻度順位と正答率の間に弱い相関が認められた。識別力の平均は 0.424 であった。識別力が 0.2 以下の項目が 9 項目あったが、一部の項目は定義文の分かりづらさが識別力の低さにつながったと推測された。また、満点の項目が 12 項目あったが、使用領域が広いために易しかったと考えられるもののほか、語の意味的透明性が高いために推測が働きやすかったと考えられるものもあった。定義文の作り方などに改善の余地があるものの、学習者の医学語彙能力を測るテストとしては一定の有用性を確認することができた。

キーワード：専門用語、医学、語彙テスト、コーパス、頻度

### 1. はじめに

近年、医療・福祉分野で就労を希望する外国人が増加している。「保健」分野を専攻する留学生の数はここ 10 年でほぼ倍増し<sup>1)</sup>、医学部留学生の数も増加している<sup>2)</sup>。留学生が大学で学ぶためには、高度な日本語の習得が必要となるが、医療分野では特に語彙学習が重要かつ負担が大きいと考えられる。英語の医学語彙の研究によれば、解剖学に関する文章の約 30% は専門用語で占められている<sup>3)</sup>。また、日本語の医学語彙研究では、医師国家試験に特徴的な語、つまり専門性の高い語の 70% 以上は、一般的な日本語教育では指導されないことが示唆されている<sup>4)</sup>。

医学生に日本語指導を行うにあたっては、学習者の

語彙力を把握しながら、医学用語を効率的に習得できる環境を提供する必要がある。医学用語の指導に関する試みには、品川ほか<sup>5)</sup>などがあるが、習得状況を測定するためのテストに関する研究は、管見の限りない。そこで、筆者らは医学語彙テストを開発・実施し、その有用性を評価することにした。

#### 1. 1 専門語彙テストに関する先行研究

専門語彙のテストに関する先行研究は極めて少ない。Lessard-Clouston<sup>6)</sup>は、神学語彙の広さと深さを測定するためのテストを開発し、12 名の神学生（母語話者 7 名、非母語話者 5 名）を対象に学期前後の 2 回にわたり実施した。その結果、母語話者と非母語話者の差は、語彙の広さの面では縮まったが、語彙の深さの面では逆に広がったことを報告している。しかし、テストの信頼性や妥当性は検証されていない。

専門用語学習の重要性は指摘されており<sup>7)</sup>、医学用語<sup>8)</sup>、薬学用語<sup>9)</sup>など各種の専門用語リストが開発されつつあるが、テストの開発や評価はいまだ発展途上といえる。

<sup>1)</sup>十文字学園女子大学国際交流センター講師

<sup>2)</sup>国際医療福祉大学総合教育センター講師

<sup>3)</sup>専修大学国際コミュニケーション学部教授

<sup>4)</sup>東京大学大学院総合文化研究科教授

専門語彙とは異なるが、同じく領域を限定した学術語彙においてはテストに関する研究がある。Schmittほか<sup>10)</sup>は、学術語彙のセクションを含む Vocabulary Levels Test の妥当性を検証している。また、田島ら<sup>11)</sup>は「日本語学術共通語彙リスト」に基づいたテストの開発と実施を行い、一定の弁別力と信頼性が確認できたことを報告している。

今後は専門語彙のテストも学術語彙同様に、開発と試行を重ねて知見を積み重ねていくことが期待される。

本研究の目的は、医学語彙テストを開発・実施し、テストが有用であるかどうかを評価することである。下位の研究課題として、(1)テストの信頼性は高いか、(2)対象語の出現頻度と、項目の難度にはどのような関係があるか、(3)それぞれの項目は十分な識別力を持っているか、を設定する。

## 2. 医学語彙テストの開発

### 2. 1 テストの概要

医学生が6年間で学ぶと考えられる医学用語を範囲とし、留学生の医学語彙力を診断するためのテストを計画した。受験者は、一般的な中級日本語の学習を終えた医学日本語の初学者から医師国家試験合格レベル

までを想定している。形式は、改訂版の Vocabulary Levels Test<sup>12)</sup>でも用いられている対象語とその定義をマッチングさせる方式である。マッチング方式は、項目同士が解答に影響し合う可能性が排除できないものの、多肢選択式などに比べて問題作成が容易であり実用性が高い。英訳等の対訳をマッチングさせる方法も考えられたが、医学用語の理解を測るにあたっては、表面的な意味だけでなく、器官の構造、疾患の機序など医学的内容の理解も構成概念に含めるべきであると考え、語の定義を選ばせることにした。

対象語は二字漢字語に限定した<sup>注1)</sup>。先行研究において医学用語における二字漢字語の造語力の高さが指摘されているからである。増田ら<sup>13)</sup>は、基本的な医学術語約7400語の約9割が漢字を含んでいることを指摘し、さらに、それら漢字を含む約9割の術語の構成要素の86%は二字漢字語であったことを報告している。

約60分で解答できるように、対象語の数（項目数）を72に設定した。後述するように、テストは9つのブロックに分かれており、1ブロック内に8語の対象語と10の定義文が提示され、受験者はそれぞれの語に合った定義文を選ぶ（定義文のうち2つは錯乱肢である）。図1に実際の問題用紙の1ブロック分を示す。

#### 問題1～8：語の意味を右のa～jから1つ選びなさい。

- |           |   |
|-----------|---|
| 1. 灌流 [ ] | a. 心臓が血を送り出すための動き。1分間に60～100回くらい行われる。             |
| 2. 透析 [ ] | b. 体の組織や臓器に液体を流すこと。                               |
| 3. 血清 [ ] | c. 血を置いておくと、分かれてできる透明な薄い黄色の液体。                    |
| 4. 拍動 [ ] | d. 細胞 <sup>(注)</sup> や組織を見やすくするために、薬品で色をつけること。    |
| 5. 赤沈 [ ] | e. 血管 <sup>(注)</sup> などの内側の大きさが小さくなること。           |
| 6. 溶血 [ ] | f. 血液の成分が、血管 <sup>(注)</sup> の壊れていない壁から出ること。       |
| 7. 漏出 [ ] | g. 成分の一部だけを通す膜 <sup>(注)</sup> を使って、血をきれいにする方法。    |
| 8. 狹窄 [ ] | h. 血の中にある薬を加えたとき、血の中の成分が下のほうに集まる速さ。               |
|           | i. 血などが普通とは反対の方向に流れること。                           |
|           | j. 赤血球 <sup>(注)</sup> の表面が壊れて、ヘモグロビンという成分が流れ出ること。 |

(注) 細胞：生物を構成する基本的な単位。多くの細胞は核を持つ。 血管：体の中で血が通る道。  
膜：物の表面をおおう薄いもの。 赤血球：血液の主な成分。酸素などを運ぶ。

図1：問題用紙の一部（1ブロック分）。実際のテストでは読みも問うているが、紙幅の都合上省略している。

## 2. 2 対象語の選択

本テストでは、医学書をコーパス化し、頻度情報を参考しながら対象語を選択する方法をとった。まず、医学書として『イヤーノート 2020 内科・外科編（以下、イヤーノート）』<sup>14)</sup>を選択した。本書を選んだ理由は、(1) 内科・外科の主要疾患に関する内容がコンパクトにまとめられた参考書であり、幅広い職種の医療従事者に活用されていること、(2) 医学部の学部レベルの内容であるため、留学生が日常的に触れる語が多いこと、(3) コーパス化して頻度情報を取得するのに十分な分量を持っていること、の 3 点である。

以下では、テストの対象語を選択した手順を示す。最初に出版社に交渉し、研究利用の目的に限定して『イヤーノート』本文テキストの利用許諾を得た。次に、形態素解析器として MeCab (Ver.0.996)、解析用辞書として IPADic (Ver.2.7.0) を用い、独自に開発した「医学語彙辞書」を組み込んで、形態素解析を実施した。この「医学語彙辞書」は、医学用語を正しく解析させることを目的として、「今日の診療プレミアム Vol.28」<sup>15)</sup>に含まれている『医学大辞典』の見出し語（計 76,625 語）と 4 冊の医学書（『今日の治療指針 2018 年版』『今日の診断指針第 7 版』『新臨床内科学第 9 版』『今日の小児治療指針第 16 版』）の索引に掲載された語（計 39,273 語）をリスト化し、計 108,942 語のエントリを持つユーザ辞書として整備したものである。

『イヤーノート』本文テキストを形態素解析した結果、延べ語数は 608,893 語、異なり語数は 25,811 語となった。このうち、「医学語彙辞書」によって同定されたのは、延べ語数で 112,570 語、異なり語数で 11,169 語であった。この 11,169 語の中から二字漢字語のみを抽出し、1,407 語を得た。二字漢字語による医療語彙全体の頻度カバー率を計算したところ 44%となつた。また、これらの二字漢字語を 1 つ以上含む語による頻度カバー率を計算したところ 51% であった。

表 1：頻度層ごとの頻度順位と頻度

	頻度順位	頻度
高頻度層	1 位～508 位	頻度 1597～16
中頻度層	509 位～939 位	頻度 15～4
低頻度層	940 位～1407 位	頻度 3～1

続いて、形態素解析時に得た頻度順位を用いて二字漢字語 1407 語を 3 つの頻度層に分割した（表 1）。一般的に語の出現頻度が高いほどテストにおける難度が低い傾向があり<sup>10)</sup>、頻度別に対象語を抽出したほうが幅広い難度の問題が揃えられると考えたからである。同一頻度の語が複数あったため、厳密に 3 等分することはできなかった。そこで、頻度を基準として、各層がおよそ同じ語数になるように 3 分割した。

次に、各層から対象語 24 語と錯乱肢用の 6 語を無作為に抽出した。このとき、「日本語教育語彙表」<sup>16)</sup>で初級～中級と判断された語は、一般語に近い語と判断し対象語から除外した（例：手術、言語など）。また、後述する定義文作成の際に、定義文に二字漢字語の両方の漢字が含まれてしまうもの、いわゆる（意味的）透明性<sup>17)</sup>が高い語は、難度が低くなりすぎると考え、差し替えた（例：心膜=心臓を包んでいる二重の膜）。

各層の語を 3 ブロックに分け、計 9 ブロックを作成した。テスト用紙上では各頻度層ブロックが連続しないように配置した（表 2 における A,B,C…I）。

## 2. 3 定義文の作成

選択肢に用いる定義文は、次の手順で作成した。

まず、筆者の 1 名が「今日の診療プレミアム Vol.28」内<sup>15)</sup>の『医学大辞典第 2 版』と「デジタル大辞泉」(<https://daijisen.jp/digital/>) を用いて、定義文の原案を作成した。本テストは対象語の理解を問うものであり、定義文の日本語が分からずに解答できないことは避けるべきである。そこで、定義文で使用する語を「日本語教育語彙表」の中級以下の語になるように言い換えを行った。言い換えが困難な語については、注釈をつけたうえで使用することとした（図 1 下部参照）。

次に、原案を医師 2 名（感染症専門医 1 名、循環器専門医 1 名）にチェックしてもらい、修正が必要と判断された部分を修正し、定義文を完成させた。

## 3. 結果と考察

2021 年 7 月中旬、関東地方の A 大学の教室において試験を実施した。受験者は、A 大学医学部の 2 年生留学生 18 名であり、国籍は、ベトナム 6 名、ミャンマー 5 名、カンボジア 3 名、モンゴル 2 名、インドネシア 2 名であった。全員、2019 年の冬に来日し、約

表2：全72項目のリスト。正答率が1.000の項目は識別力が計算できないため「---」と記している。

頻度層	プロック	語	正答率	識別力	頻度順位	頻度層	プロック	語	正答率	識別力	頻度順位
高	A	灌流	0.389	0.543	444	中	E	角化	0.833	0.733	866
高	A	透析	0.833	0.587	120	中	E	裂創	0.667	0.431	808
高	A	血清	0.833	0.411	45	中	E	大網	0.667	0.304	842
高	A	拍動	1.000	---	478	中	E	莢膜	0.056	0.189	558
高	A	赤沈	0.556	0.553	341	中	H	催吐	0.889	0.192	692
高	A	溶血	0.611	0.484	260	中	H	笑氣	0.778	0.592	643
高	A	漏出	0.778	0.671	291	中	H	嗅覚	0.944	0.500	937
高	A	狭窄	1.000	---	43	中	H	核型	0.833	0.587	765
高	D	捻転	0.667	0.639	460	中	H	余命	1.000	---	932
高	D	屈曲	0.833	0.075	430	中	H	傾眠	0.500	0.038	769
高	D	拡散	0.500	0.397	364	中	H	耳介	0.944	0.548	693
高	D	粘膜	1.000	---	49	中	H	中和	0.889	0.764	519
高	D	動悸	1.000	---	389	低	C	偏食	0.889	0.296	1366
高	D	異型	0.889	0.591	463	低	C	呑酸	0.056	0.308	998
高	D	虚脱	0.889	0.470	411	低	C	咀嚼	0.389	0.308	1023
高	D	還元	0.889	0.556	428	低	C	菌類	0.833	0.353	1210
高	G	中枢	0.944	-0.189	78	低	C	大彎	0.889	0.054	987
高	G	発作	0.667	0.339	38	低	C	骨塩	1.000	---	1234
高	G	感染	1.000	---	5	低	C	痕跡	0.444	0.368	1070
高	G	水痘	1.000	---	388	低	C	離床	0.833	0.368	1385
高	G	疾患	1.000	---	3	低	F	転座	0.444	0.598	996
高	G	患側	0.889	-0.154	394	低	F	単相	0.333	0.516	989
高	G	結石	1.000	---	249	低	F	投射	0.389	0.085	1115
高	G	縫合	0.889	0.764	424	低	F	上丘	0.111	0.396	1272
中	B	充填	0.444	0.411	735	低	F	体節	0.444	0.532	1304
中	B	圧痕	0.611	0.607	815	低	F	嗅球	0.889	0.764	1024
中	B	弾性	0.722	0.696	518	低	F	鎖肛	0.833	0.733	1071
中	B	固縮	0.500	0.201	879	低	F	泉門	0.556	0.509	1291
中	B	後屈	0.889	0.314	774	低	I	焼痂	0.889	0.348	976
中	B	徘徊	0.667	0.420	938	低	I	凍瘡	0.889	0.504	1330
中	B	牽引	0.611	0.496	773	低	I	強膜	0.778	0.500	1204
中	B	脱臼	0.778	0.461	906	低	I	羊膜	1.000	---	1139
中	E	疣贅	0.556	0.585	858	低	I	面疔	0.389	0.487	1135
中	E	水腫	0.889	0.348	742	低	I	囊腫	0.889	0.556	1344
中	E	腸癰	0.889	0.262	910	低	I	赤潮	0.389	0.487	1288
中	E	剖検	0.889	-0.067	685	低	I	勒帶	1.000	---	1282

15か月間一般的な日本語教育を受け中級レベルまで達した後、医学日本語を約4か月学習していた。

回答時間は、平均49.4分であった。無答率が16.3%であったが、無答が問題の後半に集中しているわけではないため試験時間不足によるものではないと考え、すべて誤答として扱った。

### 3. 1 テスト全体の分析

受験者ごとのテスト全体の正答率は、平均0.737（標準偏差0.146）であった。図2に度数分布を示す。0.8～0.9の受験者が最も多く、やや右寄りの分布になっているが、天井効果は見られない。項目ごとの正答率は、平均0.737（標準偏差0.243）であった。図3に度数分布を示す。正答率0.8以上の項目が40あり、項目の半数以上を占めている。

信頼性を示す指標であるクロンバッックの $\alpha$ 係数は、 $\alpha = 0.919$ であり、非常に高い信頼性が確認できた。

項目ごとの識別力を表す指標として、点双列相関係数を算出した（表2）<sup>注2)</sup>。点双列相関係数は、-1から1の間の値を取り、数字が大きいほど受験者を識別する力が高いことを示す。一般的に0.2以上が望ましいとされるが、0.2以上の項目が72項目中51項目、0.2未満が9項目、正答率が1.000のため算出できないものが12項目であった。算出できない項目を除いた平均値は0.424（標準偏差0.219）であった。

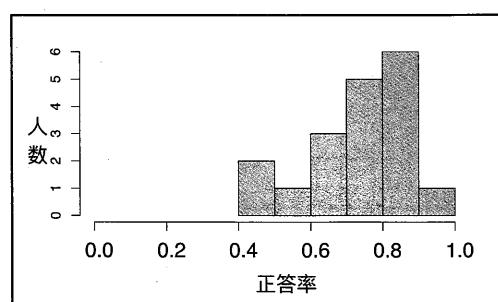


図2：受験者ごとのテスト全体の正答率

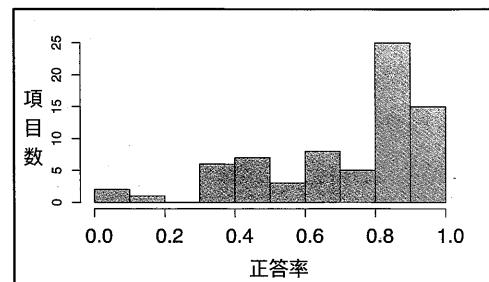


図3：項目ごとの正答率

### 3. 2 語の頻度と正答率の相関

頻度が高い語ほど正答率が高くなる傾向があるかを確認した。まず、語の頻度順位（1～1407位）と正答率の相関を求めたところ、Spearmanの順位相関係数は $\rho = -0.270$ であった。これは、ゆるやかではあるが頻度順位が高いほど正答率が高いことを示している。

次に、頻度層ごとの正答率を比較した。事前に、頻度層内の3つのブロック間で正答率に差がないかを確認した。一元配置の分散分析を実施した結果、高頻度層内のブロックA、D、G（ $p = 0.155$ 、本稿では有意水準を5%とする、 $\eta^2 = 0.162$ ）、中頻度層内のブロックB、E、H（ $p = 0.146$ 、 $\eta^2 = 0.168$ ）、低頻度層のブロックC、F、I（ $p = 0.164$ 、 $\eta^2 = 0.158$ ）すべてで正答率の差がないことが確認された。

続いて、頻度層ごとの正答率を比較した。高頻度層の平均正答率は0.836（標準偏差0.180）、中頻度層の平均正答率は0.727（標準偏差0.214）、低頻度層の平均正答率は0.648（標準偏差0.293）であり、頻度が高いほど正答率が高い傾向が見られた。3つの頻度層間の正答率をTukey-kramer法で検定した結果、高頻度層と低頻度層の間には有意差が認められた（ $p = 0.019$ 、 $d = 0.771$ ）が、高頻度層と中頻度層（ $p = 0.248$ 、 $d = 0.550$ ）、中頻度層と低頻度層（ $p = 0.477$ 、 $d = 0.307$ ）の間には有意な差は認められなかった。

以上のことから、全体として頻度順位と正答率の間に弱い相関があることが分かった。田島ら<sup>11)</sup>は、学術共通語彙における頻度順位と正答率の相関が-0.277と、一般語彙の-0.486よりも弱かったことを報告しているが、医学語彙においても同様の傾向があると言える。

### 3. 3 識別力が低かった項目

識別力が低い項目、特に正答率が中程度で識別力が低い項目は、テストの信頼性を低下させるため<sup>18)</sup>、テストの改善に向けて、原因を把握しておく必要がある。表3は、正答率が中程度（0.2～0.8）で点双列相関係数が0.2以下であった項目である。

先述の医師とともに正答と定義文をあらためてチェックしたところ、正答と定義文の結びつきが弱い可能性が指摘された。「投射」の定義文は、「目や耳などで生まれた感覚の情報を、脳のある場所で受けること。」であるが、「投射」の「投げる」という意味と「受ける」

の動作の方向が一致しないため、混乱を招いたのではないかと考えられる。また、「傾眠」の定義文は「意識の状態を示す言葉。名前を呼んだりすると目覚める。」であるが、「眠る」と反対の「目覚める」という言葉が使われており、状態の示し方が間接的であるため、混乱を招いたのではないかと考えられる。

### 3. 4 正答率が高かった項目

次に正答率が高かった項目を見る。表4は、正答率が1.000の項目である。表中、「感染」「疾患」は「日本語教育語彙表」で「上級前半」と判定される語、「粘膜」は「上級後半」と判定される語である。これらの語は純粋な医学用語というより一般語に近い使用領域を持つ語であるため、正答率が高かったと推測される。

「骨塩」は低頻度層の語であるが、対応する定義文が「骨の中に含まれるカルシウムなどの成分」であり、

「骨」という漢字が共通しているため選ばれやすかつた可能性がある。同様に、「羊膜」も低頻度層の語であるが、定義文が「妊娠した女性のお腹の中で、子供を包んでいる膜。膜の内側は水分でいっぱいになっている」である。これらのことから、定義文中に対象語の漢字が含まれていると正答率が高くなることが予想される。そこで、すべての項目(72項目)について、定義文の中に対象語の漢字が含まれている場合を1、それ以外を0として、正答率との相関を調べた。その結果、 $r=0.228$ であり、弱い正の相関が認められた。

### 3. 5 正答率が低かった項目

最後に正答率が比較的低かった項目を確認しておく(表5)。全体的に正答率の高い項目が多かったなかで、正答率が低くかつ識別力が高い項目は、成績上位群を識別するための貴重な項目と考えるからである。

「灌流」と「莢膜」以外は、すべて低頻度層の語である。頻度の低さが難しさに反映していると考えられる。「灌」「咀」「嚼」「疔」「莢」「呑」は常用漢字表外の漢字である。漢字の難度が高く、語の意味が推測しにくいため、正答を選べなかつた可能性がある。そこで、すべての項目(72項目)について、対象語に常用漢字外の漢字が含まれている場合を1、そうでない場合を0として、正答率との相関を調べた。その結果、 $r=-0.153$ であり、相関は認められなかった。

「上丘」は語の意味的透明性の低さが難度に影響し

表3：識別力(点双列相関)が0.2以下の項目

頻度層	語	正答率	識別力	頻度順位
低	投射	0.389	0.085	1115
中	傾眠	0.500	0.038	769

表4：正答率が1.000の項目

頻度層	語	正答率	頻度順位
高	拍動	1.000	478
高	狭窄	1.000	43
高	粘膜	1.000	49
高	動悸	1.000	389
高	感染	1.000	5
高	水痘	1.000	388
高	疾患	1.000	3
高	結石	1.000	249
中	余命	1.000	932
低	骨塩	1.000	1234
低	羊膜	1.000	1139
低	韌帶	1.000	1282

表5：正答率が0.4以下の項目

頻度層	語	正答率	識別力	頻度順位
高	灌流	0.389	0.543	444
低	咀嚼	0.389	0.308	1023
低	投射	0.389	0.085	1115
低	面疔	0.389	0.487	1135
低	赤潮	0.389	0.487	1288
低	单相	0.333	0.516	989
低	上丘	0.111	0.396	1272
中	莢膜	0.056	0.189	558
低	呑酸	0.056	0.308	998

ていると考えられる。この語の定義文は、「脳の一部分で、おもに視覚情報の処理に関わる」であるが、「上」も「丘」も脳や視覚と意味的なつながりを持たない漢字であるため、語義を知らない受験者にとって推測は極めて難しかったと考えられる。また、「赤潮」は食中毒に関連する語であり、他の診療科ではほとんど用い

られないため知らない受験者が多かった可能性がある。

#### 4.まとめと今後の課題

留学生の医学語彙力を測定するためのテストを開発・実施した。その結果、全体として高い信頼性が得られた ( $\alpha = 0.919$ )。頻度と難度の相関については、高頻度層と低頻度層の間に難度差が確認でき、全体として頻度順位と正答率の間に弱い相関が認められた。専門用語における頻度と難度の相関は一般語ほど高くないようだが、テストの難度を調整するために頻度情報を使うことは有効であると思われる。識別力の平均は 0.424 と十分であった。医学語彙力を診断するためのテストとして一定の有用性を持っているといえる。

一方で、改善のために検討すべき点も明らかになった。今回は医学日本語の学習歴が 4か月と比較的短い学習者への実施であったが、満点で識別力が算出できない項目が 12 項目あった。今後、学習歴がより短い、あるいは、より長い受験者への実施を通して、これら低難度項目の扱いを検討していく必要がある。また、識別力が 0.2 以下の項目のうち、定義文が分かりづらく正答が選びにくかったと思われる項目は、定義文の再検討が必要であろう。

本テストは医学用語を対象としているが、テストの作成方法は、工学、経済学など他の専門分野にも援用できると考えられる。例えば、専門語彙のテストを多肢選択式で作成すると、専門家が選択肢を作成する負担がかなり大きいが、本テストの方法であれば、専門家のチェックが必要であるとはいえ、日本語教員が主導で作成することができる。また、記述式のテストと違い採点の労力も少なく、実用性も高い。

本研究の限界を二点述べる。一点目は、語の使用領域が考慮できていないことである。三枝ほか<sup>19</sup>は、診療科ごとに特徴的な語が異なること、および複数の診療科にまたがる「医学共通語彙」の存在を示唆している。診療科ごとの使用頻度の違いが語の難度に影響する可能性があるため、これらの使用領域を統制したテストの開発も必要であろう。もう一点は、妥当性や波及効果を検討できていないことである。Schmitt<sup>20</sup>は、受験後の受験者にインタビューし、本当に語の意味を知っていて解答したのかを確認することによって妥当

性を確かめることを推奨している。本テストにも同様の妥当性検証が必要であろう。また、今後、本テストの得点と医学授業の成績に相関があるかどうかや、本テストに向けて準備することが医学日本語学習全体にどのような影響を与えるかといった波及効果についてのデータも集めたい。

**付記:** 本研究は JSPS 科研費 18H00679 の助成を受けている。

#### 注

注 1 医学用語には、二字漢字語+二字漢字語（例：動脈硬化、慢性肝炎）、二字漢字語+接辞的要素（例：咽頭痛）、カタカナ語（例：ステロイド）、ひらがなを含んだ語（例：かぜ症候群）などさまざまな語構成の語が存在する。

注 2 受験者の能力と独立して項目の困難度や弁別力を分析する理論に項目応答理論やラッシュ分析があるが、これらの方法で誤差なく測定するためには、少なくとも 100 サンプル程度が必要と言われているため、本稿では古典的テスト理論の枠組みを使って分析を行う。

#### 参考文献

- 1) 日本学生支援機構：外国人留学生在籍状況調査,  
<https://www.studyinjapan.go.jp/ja/statistics/zaiseki/>
- 2) 池田俊也・天野隆弘：国際医療福祉大学医学部の開学について、国際医療福祉大学学会誌 22 卷, 2 号, pp.1-5 (2017)
- 3) Chung, T.M., Nation, P. : Technical vocabulary in specialised texts, Reading in a Foreign Language, Vol.15, No.2, pp.103–116 (2003)
- 4) 山元一晃・稻田朋晃・品川なぎさ：医師国家試験の名詞語彙の対数尤度比に基づく分析と教材開発の可能性、日本語／日本語教育研究 9 号, pp.245-260 (2018)
- 5) 品川なぎさ・稻田朋晃：医学部留学生を対象とした医学日本語授業の実践—医学教員との連携による日本語教育の取り組み—、専門日本語教育研究, 21 号, pp.61-68 (2019)
- 6) Lessard-Clouston, M. : Breadth and Depth: Specialized Vocabulary Learning in Theology among Native and Non-native English Speakers, The Canadian Modern Language Review, Vol.63, No.2, pp.175–198 (2006)

- 7) Nation, P. : Learning Vocabulary in Another Language, Cambridge, Cambridge University Press (2001)
- 8) Hsu, W. : Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list, *Language Teaching Research*, Vol.17, No.4, pp.454–484 (2013)
- 9) Fraser, S. : Providing ESP Learners with the Vocabulary They Need: Corpora and the Creation of Specialized Word Lists, *広島外国語教育研究*, 10 号, pp.127-143 (2007)
- 10) Schmitt, N., Schmitt, D. & Clapham, C. : Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test, *Language Testing*, Vol.18, No.1, pp.55–88 (2001)
- 11) 田島ますみ・佐藤尚子・橋本美香・松下達彦・笹尾洋介：日本語学術共通語彙テストの開発, 中央学院大学人間・自然論叢, 45 号, pp.19-31 (2018)
- 12) Webb, S., Sasao, Y., Ballance, O. : The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT, *ITL - International Journal of Applied Linguistics*, Vol.16, No.1, pp.33–69 (2017)
- 13) 増田光司・中川健司・佐藤千史：二漢字語を見出し語とした医学術語学習辞典作成の試み, 専門日本語教育研究, 6 号, pp.49–54 (2004)
- 14) 岡庭豊編：イヤーノート 2020 内科・外科編, メディックメディア (2019)
- 15) 今日の診療プレミアム Vol.28 DVD-ROM for Windows, 医学書院 (2018)
- 16) 日本語教育語彙表：<https://readability.net/jew/>
- 17) 本多由美子：二字漢語における語の透明性, 計量国語学, 31 卷, 1 号, pp.1-19 (2017)
- 18) 日本語教育学会編：日本語テストハンドブック, 大修館書店, (1991)
- 19) 三枝令子・丸山岳彦・松下達彦・品川なぎさ・稻田朋晃・山元一晃・石川和信・小林元・遠藤織枝・庵功雄：医学用語の収集と分類, *日本語教育*, 176 号, pp.33-47 (2020)
- 20) Schmitt, N. : The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge, *Language Testing*, Vol.16, No.2, pp.189–216 (1999)

## Development and Evaluation of Medical Vocabulary Tests

INADA, Tomoaki<sup>1</sup> SHINAGAWA, Nagisa<sup>2</sup>  
MARUYAMA, Takehiko<sup>3</sup> MATSUSHITA, Tatsuhiko<sup>4</sup>

<sup>1</sup> International Center, Jumonji University

<sup>2</sup> Center for Liberal Arts, International University of Health and Welfare

<sup>3</sup> School of International Communication, Senshu University

<sup>4</sup> Graduate School of Arts and Sciences, University of Tokyo

In recent years, the number of medical international students has been increasing. However, there are few tests to measure learners' medical vocabulary skills. Therefore, we developed and conducted a medical vocabulary test and evaluated its usefulness. Morphological analysis was conducted on the text data of the "year note 2020" using the "Medical Vocabulary Dictionary" developed by the authors. A total of 72 words (24 words each from the high, medium, and low-frequency layers) were selected and a test was created in the matching format. Eighteen international medical students took the test. The reliability was 0.919. The facility value in the high-frequency layer was significantly higher than in the low-frequency layer and overall there was a weak correlation between word frequency and facility value. The mean discrimination power was 0.424 and there were 9 items with low discrimination power. In some items, it was assumed that the unintelligible definition led to the low discrimination power. In addition, there were 12 items with perfect scores, some of which were considered to be easy because of the wide range of use or the high semantic transparency. We confirmed a certain level of usefulness as a test for measuring medical vocabulary knowledge.

**keywords:** technical term, medical science, vocabulary test, corpus, word frequency