

日本語教育に活かすための看護師国家試験におけるカタカナ語の様相の分析

山元一晃(国際医療福祉大学)

1. はじめに

日本国内において看護を学ぶ留学生や、経済連携協定(EPA)により看護師を目指す看護師候補者などは、看護師として実務に就く前に看護師国家試験を受験し合格する必要がある。看護師国家試験は、看護師として必須とされる知識を問うものであると考えられるが、問題文や選択肢などを流暢に読めなければ、理解が難しくなることが予想される。Nation(2002)が、先行研究を引用しながら語彙の知識が読解に深く関係していることを指摘しているように、流暢な読みには語彙の知識が重要になる。語彙においては、カタカナ語が難しいことがよく指摘される。たとえば、陣内(2008)の調査によれば、日本語学習者の77.8%が、「カタカナ語が分からなくて困ったこと」が「ある」「よくある」「ときどきある」と回答しており、中国語話者に限ると約90%となっている。このことを踏まえると、看護を学ぶ過程や、国家試験の対策においてもカタカナ語に困難を覚える学習者が多いことが予想される。しかし、これまで、語彙を網羅的に扱った岩田(2012)を除き、国家試験のカタカナ語の様相について分析した研究はみられない。そこで、本発表では、テキストカバー率や対数尤度比(LLR)に基づく特徴度といった指標を参考にしながら、看護師国家試験で、1) どのようなカタカナ語が用いられており、2) それらは、一般的なテキストと比べてどの程度特徴的なのか、3) それらはどの程度日本語教育で扱われるのか、を明らかにする。

2. 方法

まず101回から108回までの看護師国家試験8回分をテキスト化し、形態素解析機MeCab 0.996および形態素解析用辞書UniDic 2.3.0を用いて短単位に形態素解析を行った。短単位を用いたのは、BCCWJ(現代日本語書き言葉均衡コーパス)の短単位語彙表と比較し特徴的なカタカナ語を抽出するため、また、将来的に、基礎的な単位から、それに関連する複合語を抽出するためである。

次に、語彙素がカタカナの短単位の頻度を集計し、BCCWJの短単位の頻度と対照させLLRを算出した。LLRは2つのコーパスのどちらか一方に特徴的な表現について高値になるため、BCCWJにおいて相対的な頻度が上回った場合、LLRに-1を乗ずる補正を行い特徴度とした。LLRは表1の分割表に基づき、以下の式により算出した。(Kilgarriff 2001, 寺嶋 2009)

$$LLR = 2(a \log(a) + b \log(b) + c \log(c) + d \log(d) - (a+b) \log(a+b) - (a+c) \log(a+c) - (b+d) \log(b+d) - (c+d) \log(c+d) + n \log(n))$$

表1 統計量を求めるための分割表(寺嶋 2009)

	対象コーパス	参照コーパス	計
見出し語 W の頻度	a	b	a+b
見出し語 W 以外の頻度	c	d	c+d
計	a+b	b+d	a+b+c+d=n

なお、LLRは寺嶋(2009)やDunning(1993)などで、コーパスサイズの影響を受けにくいことが指摘されている。そのため、本研究でも、特徴度の指標として、LLRを用いた。LLRの算出については、Kilgarriff(2001)に依った。本発表においては、特徴度が15.13($p < .0001$)を上回る場合を特徴的とした。Rayson et al.(2004)によれば、期待頻度を1以上とした場合、有意水準を0.001%にすることが安全だとされているためである。なお、8回分の国家試験には、頻度が低いものの、特徴度が高いものが多く含まれている。

さらに、形態素解析によって抽出された語を『日本語教育語彙表 Ver 1.0』(Sunakawa et al. 2012)と対照させ、日本語教育で扱われるかを判定した。

なお、本発表でのカタカナ語とは、語彙素がカタカナのみで表記されているものをさし、単位など原文ではアルファベット表記されているものを含む一方、複数の文字種からなる語は含まない。このことについては、今後、検討していく必要がある。

3. 結果と考察

本節では、2節に述べた方法による結果を示し、テキストカバー率等の観点から考察を行う。

形態素解析を行った結果、延べで133,804語、異なりで7,289語の語が抽出された。そのうちカタカナ語は、延べで4,386語(5.4%)、異なりで918語(12.6%)であった。延べ語数の割合に比して、多様なカタカナ語が使われることがうかがえる。

Nation (2002) は、テキストを理解し、知らない語の意味を推測しながら読むためには、知らない語が2%を超えないことが望ましいとしている。また流暢な読みのためには、99%—100%のカバー率が必要であろうことを指摘している。看護師として必須の知識を問う試験であることを考えれば、看護師国家試験においては、問題文や選択肢を流暢に読めなければいけないだろう。そのことを考慮に入れると、99%超のテキストカバー率を満たしていることが望ましいと考えられる。

カタカナ語以外を全て理解していたとした場合、既に96.7%のテキストカバー率を満たしていることになる。そこで、98%および99%のテキストカバー率を満たすためには、頻度が高い順にさらに何語のカタカナ語を覚えればよいかを調べた。その結果、98%のカバー率を満たすためには、32語を覚えればよいことが分かった。一方で、99%を満たすためにはさらに167語を覚える必要がある。

カタカナ語以外を全て覚えており、かつ、『日本語教育語彙表』の初級前半から順番に全てのカタカナ語を覚えていった場合のテキストカバー率を検討した結果を表2に示す。なお、各レベルに相当するカタカナ語の延べ語数と異なり語数も表2に示してある。

表2 『日本語教育語彙表』に掲載されているカタカナ語の語数とテキストカバー率の関係

	異なり語数	延べ語数	累積頻度	テキストカバー率
初級前半	18	106	129524	96.8%
初級後半	44	370	129894	97.1%
中級前半	96	709	130603	97.6%
中級後半	116	932	131535	98.3%
上級前半	44	135	131670	98.4%
上級後半	13	45	131715	98.4%

異なりで36.1%、延べで52.4%の語が『日本語教育語彙表』にあり、中級後半までの語を覚えれば、98%のテキストカバー率を満たすことができる。一方で、99%のテキストカバー率を満たすためには、上級後半までの語では足りず、さらに多くの語を覚える必要がある。具体的には、延べで751語に相当する異なり語を覚えなければならない。日本語教育語彙表にないカタカナ語を、頻度の高い順から覚えるとすると、102語を追加で覚える必要がある。

高頻度のカタカナ語から順に覚えていく場合と比較すると、『日本語教育語彙表』に掲載されている語を全て覚えてから学習していく場合の方が、追加で覚えるべき語が少ない。『日本語教育語彙表』に掲載されている語であれば、日本語教育や日常において使う可能性も高く、有用性が高いといえる。カタカナ語を苦手とする看護留学生に対しては、まず、『日本語教育語彙表』に掲載されているようなカタカナ語の習得を促すことが重要だと考えられる。

また、7,289語の異なり語のうち、看護師国家試験に有意に特徴的だとされる語(ただし、未知語を除く)は、1509語あった。そのうちカタカナ語は、207語(13.7%)あった。全異なり語のうち、カタカナ語の占める割合は、12.6%であったことを考えると、看護師国家試験に特徴的な語には、相対的にカタカナ語が多いといえる。

特徴的なカタカナ語のうち、『日本語教育語彙表』に掲載されている語は49語のみであった。158語(76.3%)が『日本語教育語彙表』になく、日本語教育を受けるだけでは国家試験に出題されるカタカナ語は習得が難しいことが示唆される。以下に、国家試験に有意に特徴的な語を示す。『日本語教育語彙表』にない語については下線を付した。

デシリットル, キャンサー, ケア, ミリリットル, カテーテル, バイタル, ミリグラム, ミリメートル, シンドロ

ーム, インシュリン, ドレーン, アセスメント, キログラム, リハビリテーション, アルブミン, ホルモン, センチメートル, ビタミン, ドレナージ, トイレ, クレアチニン, ストーマ, スケール, ベッド, カリウム, チアノーゼ, サイン, リンパ, ビリルビン, ステロイド, アシドーシス, ファウラー, イレウス, チューブ, ナトリウム, ギプス, ヘモグロビン, オブ, コンファレンス, ナンバー, ニチモク, トリアージ, オキシコドン, ワルファリン, マスク, アレルギー, パーキンソン, ステーション, エストロゲン, カルシウム, プロトロンビン, グロブリン, ノルアドレナリン, アドレナリン, アルツハイマー, フェンタニル, プロラクチン, レスパイト, カニューラ, ハヴィガースト, サービス, レボドパ, モルヒネ, プレドニゾロン, グリセリン, ナース, ボンベ, コード, セルフ, フード, ヒスタミン, アナフィラキシー, アトロピン, ランク, アルドステロン, ガーゼ, レノー, テストステロン, マッサージ, カニューレ, ピアサポーター, フロセミド, シーネ, パーセント, アセチルコリン, ワクチン, テオフィリン, イオン, シリンジ, ルクス, インシデント, ヘパリン, クラミジア, クッシング, フィブリノーゲン, ジャパン, ノーマリゼーション, ヘルペスウイルス, クレアチン, コレステロール, アミラーゼ, インフルエンザ, セロトニン, クランプ, イート, パニック, エリスロポエチン, バビンスキー, ベンチュリー, マズロー, プロゲステロン, ヒラナメ, トレンデレンブルク, ブジー, ウイルス, ネフローゼ, レム, パートナー, クリニカル, リエゾン, アルコール, エリテマトーデス, ステータ, ノロウイルス, リングル, ジストロフィ, ストレス, ヘンレ, ボディメカニクス, ロンベルグ, パッド, ステート, ポータブル, バソプレッシン, パイオハザード, レスキュー, クロルヘキシジン, ジギタリス, ホーム, アラーム, インドメタシン, リウマチ, グループ, ポンプ, カルシトニン, エタノール, エージズム, ヘルパー, ストレングス, ループス, トリソミー, スクリーニング, ボルグ, サンサ, ジゴキシ, レニン, セクレチン, テタニー, トロンボプラスチン, ペーシング, マンニトール, ヘルニア, アスピリン, ストックキング, リスク, キナーゼ, ドーパミン, アカシジア, スルホニル, ペースメーカー, オピオイド, メニエール, クリーゼ, ホルムアルデヒド, スコア, ヘンレ, ミルキング, ヘルス, ヘルペス, コントロール, アンギオテンシン, ソマトスタチン, フィンク, シャワー, タンポナーデ, オピニオン, シャント, アポクリン, コルチ, デュシエンヌ, エン, マットレス, ディレクティブ, ダンピング, フェース, リットル, オキシトシン, クロイツフェルト, シムス, トリプシン, ブルキニエ, アドボカシー, コーション, コンサルテーション, サイトメガロウイルス, マウス, チーム

病名や物質の名称の他、「マズロー」などの固有名詞なども含まれる。この結果を留学生等への支援に活かすためには、3年間から4年間にわたる看護の専門教育において習得する可能性のある語がどれかを峻別するなどしていく必要がある。山元ほか(2020)は医師国家試験に用いられている全ての特徴的な語について、専門用語辞典を参照して、専門教育において学習するかどうかを検討している。その結果、12.5%に相当する272語が専門教育で学習されない可能性が高い語であったという。本発表での分析についても同様の検討をしていく必要がある。また、それぞれの語について、習得の難易度を検討する必要もある。

なお、その他にカタカナ語のうち未知語として解析された語は延べで57語(全体の0.04%、カタカナ語の1.2%)、異なりで44語(全体の0.6%、カタカナ語の4.7%)あったことを付け加えたい。延べで0.04%であり、国家試験の理解には影響を及ぼす可能性は低い。

4. まとめと今後の課題

本発表では、看護師国家試験において用いられているカタカナ語について検討した。その結果、以下のことが分かった。

- 1) 頻度の高い語から学習していった場合、カタカナ語以外に加えて、98%のテキストカバー率を満たすためには32語、99%のテキストカバー率を満たすためには、167語を追加で覚える必要がある。
- 2) 『日本語教育語彙表』に掲載されている語を優先的に学習した場合、カタカナ語以外に加えて、中級後半までの語彙を学習することで、98%のテキストカバー率を満たすことができる。一方で、上級後半までの語を全て学習したとしても99%のテキストカバー率を満たすことはできず、102語を追加で覚える必要がある。
- 3) 有意に特徴的なカタカナ語は、207語あり、そのうち、158語(76.3%)が『日本語教育語彙表』にない語であった。

上記の3点を考慮に入れると、どのような順序で語彙学習を進めているとしてもカタカナ語を学習することは避けられないことが明らかとなった。しかし、看護留学生やEPAによる看護師候補者らが国家試験へ向けて勉強する際に、カタカナ語がどの程度障壁になっているのかは明らかとなっていない。カタカナ語の学習を積極的に取り入れていくべきなの

か、また、他の語彙と同時に学習していくべきなのかについては、今後の検討が必要である。課題としては、以下の点があげられる。

- 1) 看護留学生等にとって、カタカナ語の専門語と、それ以外の専門語について、難易度が異なるのかを明らかにすること。
- 2) 専門教育で扱われるカタカナ語と、そうでないカタカナ語を峻別すること。
- 3) 複合語を含む複数の語種からなる語についての扱いを検討すること。

また、本発表の分析においては、以下の課題が残る。

- 1) アルファベットで表記されている語でも UniDic に含まれている語であれば、カタカナ語として抽出されていること。具体的には、単位や Cancer などの語。

謝辞 本研究は JSPS 科研費 JP19K00744 の助成を受けたものです。

参考文献

- Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1), 61-74.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge, Cambridge University Press.
- 岩田一成. (2014). 看護師国家試験対策と「やさしい日本語」. *日本語教育*, 158, 36-48.
- 陣内正敬. (2008). 日本語学習者のカタカナ語意識とカタカナ語教育. *言語と文化*, 11, 47-60.
- Kilgarriff, Adam. (2001). Comparing corpora. *International Journal of Corpus Linguistics*, 6(1), 97-133.
- Rayson, P., Berridge, D., & Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. *7es Journées Internationales d'Analyse Statistique Des Données Textuelles*, 2, 926-936.
- Sunakawa, Y., Lee, J., & Takahara, M. (2012). The Construction of a Database to Support the Compilation of Japanese Learners' Dictionaries. *Acta Linguistica Asiatica*, 2(2), 97- 115. <https://doi.org/10.4312/ala.2.2.97-115>
- 寺嶋弘道 (2009). 日本語教育語彙を選定するための統計的指標-尤度比検定, カイ 2 乗検定, イェーツの補正公式の特徴. *ポリグロシヤ*, 17, 71-83.
- 山元一晃・稲田朋晃・品川なぎさ. (2020). 日本語教育で扱うべき語の選定のための医学用語と一般語のはざまの語彙の分析. *日本語教育*, 157, 80- 87.