専門用語を対象とした語彙数推定テストの開発とその信頼性の評価

図書館情報学分野を事例として

朱心茹 † 浅石卓真 †‡ 河村俊太郎 †‡‡

† 東京工業大学 †‡ 南山大学 †‡‡ 東京大学

†zhu.x.ac@m.titech.ac.jp

本稿では，専門用語を対象とした語彙数推定テストの開発とその信頼性の評価について報告する。本専門語彙数推定テストは，数十程度の専門用語について，これを知っているかどうかを回答させることで回答者が持つ当該分野の専門語彙数を推定し，それを通して回答者の当該分野における知識量を測るものである。本テストをウェブアプリケーションとして開発し，司書課程を履修している大学生 28 名を対象に本テストの信頼性の評価を行った。その結果から，本テストには一定の信頼性があることが分かった。

1 はじめに

ある個人が特定の分野に関して習得している語彙数 [1] は，その個人の当該分野における知識量を反映するものである。そのため，学習者が持つ専門語彙数とその変化を把握することは，学習者の知識量と学習を通した知識形成の過程を記録することにつながる。これは，学習者による学習成果の自己確認や，教授者による講義計画の策定にも有用であると考えられる。

本研究では，専門分野における知識量を効果的に測定する手段として，一定の信頼性と妥当性を持った専門語彙数推定テストを提案する。この専門語彙数推定テストは，数十程度の専門用語について，これを知っているかどうかを回答させることで，回答者が持つ専門語彙数を推定し，これを通して回答者が習得している当該分野の知識量を測定するものである。本稿では，図書館情報学分野を事例として，専門用語語彙数推定テストの開発とその信頼性の評価について報告する。

図書館情報学の知識量を測定する試みとして，2010 年代に日本図書館情報学会が主体となり実施された図書館情報学検定試験（以下，検定試験）がある。検定試験では，図書館情報学のコア領域から 50 問が出題された。検定試験の結果からは，同試験が図書館情報学の知識量を測るという目的と概ね整合的であったことが報告されている [2]。しかし，このような大規模な試験を安定して実施するには多大な時間と労力，資金がかかるため，現在では実施されていない。本研究で提案する専門語彙数推定テストを用いることで，より簡便に図書館情報学の知識量を測定することができる可能性がある。

テストには，測定結果が一貫しているかどうかを示す信頼性と，測定しようとしているものを確実に測定しているかどうかを示す妥当性の両方が備わっている必要がある。本稿では専門語彙数推定テストの信頼性に関する評価結果を報告し，妥当性に関しては今後の研究で評価する。

本稿の構成は以下の通りである。第 2 章で，先行研究に基づき語彙数推定の方法を概観した上で，本研究で行った専門語彙数推定テストの開発について説明する。第 3 章で，専門語彙数推定テストの信頼性の評価について報告する。第 4 章で，本稿のまとめと本研究の展望を述べる。

2 専門語彙数推定テストの開発

2.1 語彙数測定の方法

個人の語彙数を測定する方法として，語彙数調査と語彙数推定がある。

語彙数調査には全数調査とサンプリング調査がある。全数調査は，予め用意されたリストの全単語について，知っているかどうかを調査するものである。全数調査では正確な結果を得ることができるが，時間と労力がかかるため，大規模な調査は困難である。サンプリング調査は，リストから単語を何らかの方法で抽出し，調査を行う方法である。全数調査と比べて短時間で行うことができるが，抽出された単語集合に偏りが生じる可能性

があり，測定精度が保証されない。

このような全数調査とサンプリング調査の問題に対して，単語親密度を利用して語彙数推定を行う方法が考案されている [3]。単語親密度は単語の主観的特性値のひとつで，単語の「なじみ深さ」を数値化したものである。単語親密度が大きいほど人々になじみがあり，よく「知られている」単語であるとみなされる。単語親密度を利用する推定方法では，単語をリストからランダムにではなく，単語親密度の分布を考慮して抽出し，単語集合を作成する。その後，テスト項目となる単語を知っているかどうかを回答してもらい，知っていると回答された単語の中で単語親密度が最も低いものより単語親密度が高い単語は全て知っていると仮定して語彙数を推定する。

この方法では，50 語程度の比較的少数の単語集合から語彙数を推定することができることが知られている。また，予めリストの全単語に単語親密度を付与しておくことで，様々なテストを作成できる [4]。本研究では，日本語全体を対象とした単語親密度に基づく語彙数推定テスト[4]を参考に，専門語彙数推定テストを開発する。

2.2 専門語彙数推定アルゴリズムの設計

単語親密度に基づく語彙数推定では，対象となる全単語に予め単語親密度を付与する必要がある。筆者らはこれまでの研究において，『図書館情報学用語辞典 第 5 版』の見出し語 1677 語に単語親密度を付与した [5]。また，この中から図書館情報学の最も重要な専門用語として，日本図書館協会，学文社，樹村房，ミネルヴァ書房から出版された司書課程の教科書シリーズ（計 35 冊）のうち，5 冊以上で索引語として使われている 249 語を抽出し，重要語リストを作成した。

語彙数を推定するアルゴリズムは，天野ほか (2005) と藤田・小林 (2022) の手法を組み合わせた。基本的には後述するアルゴリズム 1 を用いて推定を試みるが，失敗した場合（全て「知っている」か「知らない」と回答するなど，極端な回答があった場合に推定が失敗することがある）にはアルゴリズム 2 にフォールバックすることとした。以下に 2 つのアルゴリズムについて説明する。

2.2.1 アルゴリズム 1[4]

1 回答者に専門用語を提示し，「知っている」かどうかを回答してもらう。
2 各語に対応する単語親密度の値を説明変数 x，回答者が各語を知っているかどうかを目的変数 y とし，x と y のデータに対してロジスティック回帰分析を行う。ここで，回答者が「知っている」語は $y = 1$，「知らない」語は $y = 0$ とする。
3 上記で得られた回帰モデルを用いて，「知っている」確率が 50%の語の単語親密度を計算する。
4 専門用語リスト（本稿の場合は重要語リスト）の中で，上記で得られた単語親密度より単語親密度が大きい語の数を求め，推定語彙数とする。

2.2.2 アルゴリズム 2[3]

1 回答者に専門用語を提示し，「知っている」かどうかを回答してもらう。
2 単語親密度順に「知らない」語が 2 つ以上連続する語の単語親密度と，「知っている」語が 2 つ以上連続する語の単語親密度との中間を単語親密度の境界とする。
3 専門用語リスト（本稿の場合は重要語リスト）の中で，上記境界より単語親密度が大きい語の数を求め，推定語彙数とする。

2.3 ウェブアプリケーションの開発

上記アルゴリズムを実装し，専門語彙数推定テストを簡便に作成・回答することができるウェブアプリケーションとして，TermMator[6] を開発した。TermMator はウェブサーバ上で公開されており，十分な速度で動作することが確認されている。以下に TermMator の機能を説明する。

2.3.1 回答者画面

回答者は，回答したい用語集（テストセット）を選択し，各専門用語を知っているかどうかを回答することができる（図 1）。選択肢は「簡潔に説明できる」「見たり聞いたりしたことがある」「全く知らない」の 3 つである。全ての項目に回答することで，回答者は自身の推定認識語彙数を確認することができる。

後述する管理画面では，「簡潔に説明できる」または「見たり聞いたりしたことがある」を「知っている」とみなした時と，「簡潔に説明できる」のみを「知っている」とみなした時の推定語彙数をそれぞれ確認することができる。本稿では前者を認識語，後者を理解語とする。

図 1: 回答画面

2.3.2 管理者画面

管理画面にログイン後，管理者は専門用語（本稿の場合は重要語リストに含まれる 249 語）の登録，テストセットの登録，回答結果の確認とダウンロードなどを行うことができる（図 2）。また，実証実験参加者の管理などを行うこともできる。

図 2: 回答結果確認画面

3 専門語彙数推定テストの評価

3.1 評価実験の実施

専門語彙数推定テストの信頼性を評価するため，評価実験を実施した。評価実験には司書課程を履修している南山大学の学生 28 名が参加した。本実験は，南山大学の「人を対象とする研究」倫理審査で承認を得て行った（承認番号 22-007）。

実験にあたっては，前述した 249 語の図書館情報学重要語リストをもとに，50 語からなるテストセットを 3 つ作成した。このテストセットの構築手順は以下の通りである。

1 重要語リストに含まれる専門用語を単語親密度順に並べる。
2 並べられた専門用語それぞれについて，直前の専門用語との単語親密度の差を計算する。
3 重要語リストから，手順 2 で計算した差が最も小さい専門用語を除外する。
4 手順 3 を，重要語リストが 150 語になるまで繰り返す。
5 残った専門用語に 1 から 150 までの通し番号を付与した後，通し番号が $3i+1$，$3i+2$，$3i+3$ $(0 \leq i \leq 49)$ のものに分割し，50 語のテストセットを 3 つ作成する。

実験参加者には，基本情報に回答してもらった上で，3 つのテストセットに回答してもらった。基本情報には，「大学名」「学部名」「学科名」「学年」「性別」「履修済み／履修中の講義」が含まれる。3 つのテストセットに含まれる 50 語に対する回答からそれぞれ語彙数推定を行った他，150 語に対する回答全体からも推定を行った。

次節では評価実験の結果を報告する。

3.2 評価実験の結果

まず，参加者の回答結果と語彙数推定の結果を見る。参加者の 3 つのテストセットにおける認識語と理解語の回答結果と推定語彙数を図 3 と図 4 に示す。2 つの箱ひげ図を結ぶ線は，回答結果とそれに対応する推定結果を示している。語彙数のサンプリング調査とは異なり，テストセットにおける認識語数や理解語数が必ずしも多くなくとも，ある程度難しい用語を知っていれば推定語彙数が増えるなど，単語親密度に応じて語彙数推定がなされていることが見てとれる。

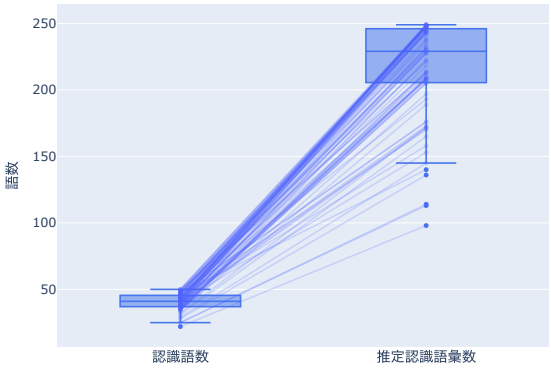

図 3: 認識語の回答結果と推定結果

次に，回答結果の分布を見る（図 5）。単語親密度が高い専門用語に対しては「簡潔に説明できる」という回答が多く，単語親密度が低い専門用語に

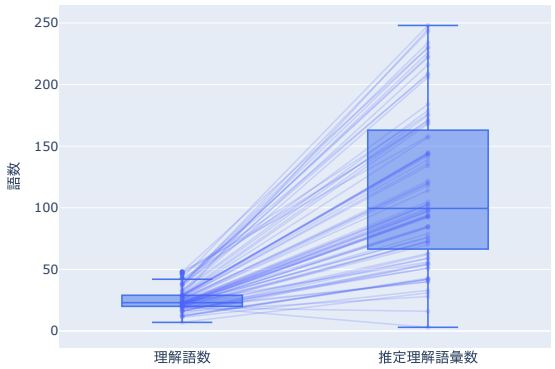

図 4: 理解語の回答結果と推定結果

対しては「全く知らない」という回答が多いことが分かる。一方で，単語親密度が低くても「簡潔に説明できる」という回答が多い「MLA 連携」や「デューイ十進分類法」，「集中目録作業」などの用語があり，これには参加者の履修済み／履修中の講義が影響している可能性がある。

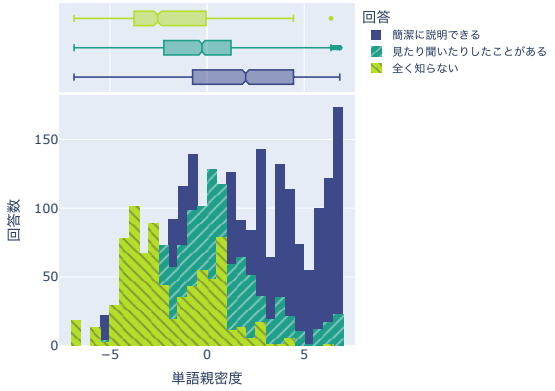

図 5: 回答結果の分布

最後に，専門語彙数推定テストの信頼性を評価するため，(1) 50 語からの推定結果の平均と 150 語からの推定結果の相関係数と (2) 50 語からの推定結果の級内相関係数（ICC）を求めた結果を表 1 に示す。前者は 2 回の測定値の相関の強さに基づく信頼性係数である。後者は複数回測定したときの測定値の一致度に基づく信頼性係数であり，分散分析で算出される平均平方和の期待値を利用して算出する。ICC には 6 つの種類があるが，今回は複数回のテストで測定された参加者それぞれの推定語彙数の一致度を測りたいため，$ICC(2,1)$ を採用した [7]。

係数をそれぞれの解釈基準に照らし合わせると，専門語彙数推定テストは，認識語彙数推定では一定の信頼性が，理解語彙数推定では高い信頼性があると言える。理解語数推定の方が信頼性が高いのは，理解語数の方がテストセットによらず安定している，理解語数の方が少ないためアルゴリズム 1 による推定が成功しやすい，などの理由が考えられるが，具体的な検証は今後の課題とする。

表 1: 信頼性係数

	相関係数	$ICC(2,1)$
認識語彙数推定	0.68	0.55
理解語彙数推定	0.88	0.72

4　まとめと今後の展望

本稿では，専門語彙数推定テストの開発とその信頼性の評価について報告した。本テストは一定の信頼性を備えているため，知識量を測定する簡便なテストとして有用である可能性が高い。

今後は，他の知識量を測定するテストと推定語彙数の関係や，回答者の学年や科目履修状況と推定語彙数の関係を分析することで，本テストの妥当性について評価を進める。また，語彙数の全数調査と推定語彙数の関係を分析することで，本テストの精度についても評価を行う予定である。

謝辞

本稿の執筆にあたりご助言をいただいた東京大学の影浦峡教授に御礼申し上げます。

注

1) 一般的には「語彙量」であるが，本研究では日本語教育や言語心理学分野での慣例にしたがい，「語彙数」を用いる。

2) 浅石卓真ほか「図書館情報学検定試験の結果分析」『第 62 回日本図書館情報学会研究大会発表論文集』2014, p. 13–16.

3) 天野成昭ほか「単語親密度を利用した語彙数推定：インターネットによる大規模調査」『日本認知科学会第 22 回大会』2005, p. 58–59.

4) 藤田早苗・小林哲生「令和版単語親密度に基づく大規模語彙数推定調査～Web 公開版の利用ログ分析～」『第 36 回人工知能学会全国大会論文集』2022, p. 1–4.

5) 浅石卓真ほか「図書館情報学用語を対象とした単語親密度の推定」『2022 年度日本図書館情報学会春期研究集会』2022, p. 59–62.

6) https://termmator.zhuxinru.com/ (last accessed 2022-09-29)

7) Shrout, Patrick E.; and Fleiss, Joseph L. “Intraclass Correlations: Uses in Assessing Rater Reliability,” *Psychological Bulletin*, vol. 86, no. 2, 1979, p. 420–428.