

語の出現頻度と難易度に着目した IT パスポート試験の分析

松山宏樹^{*1*2}, 長谷川忍^{*2}

^{*1} 秀明大学

^{*2} 北陸先端科学技術大学院大学

An Analysis of the IT Passport Examination Focusing on Word Frequency and Difficulty

Hiroki Matsuyama^{*1}, Shinobu Hasegawa^{*2},

^{*1} Shumei University

^{*2} Japan Advanced Institute of Science and Technology

IT passport examination is a qualification for proving basic knowledge about information technology. The number of international applicants for this exam is increasing from different nationalities. Thus, the ability to comprehend the kanji, katakana, and alphabet varies from learner to learner. This paper analyzes the frequency of words used in the IT passport examination and the difficulty level using the Japanese language learners' dictionary.

キーワード: IT パスポート試験, 外国人留学生, 語の出現頻度, 語の難易度

1. はじめに

IT パスポート試験⁽¹⁾の受験者数は増加傾向にあり, その中には外国人留学生の受験者も含まれる. その背景には, 2008 年に発表された「留学生 30 万人計画」がある⁽²⁾. これによって日本語教育機関をはじめ, 専修学校(専門課程)や大学(学部), 大学院に多くの外国人留学生が在籍し学業や研究を行うようになった.

令和 3 年度の国地域別の留学生は, 中国, ベトナム, ネパール, 韓国, インドネシアという順番である⁽³⁾. 非漢字圏の留学生も増えていることから, 学校側としてはこれまで以上に教える工夫が必要になってくる. 筆者が所属する秀明大学においても, 様々な国の外国人留学生が多く在籍している. 筆者も IT パスポート試験の内容を扱った授業を担当しており, その中で外国人留学生にとって効果的な学習支援について日々模索してきた. そこで本研究では, IT パスポート試験の合格を目指す外国人留学生への学習支援方法について提

案する.

IT パスポート試験の特徴としては, 設問や選択肢に専門用語が多く使われるため, 非常に難解な日本語であるという点が挙げられる. また, 出題範囲が広く, ストラテジ, マネジメント, テクノロジ系の 3 つの分野から出題される. ストラテジ系の具体例では企業活動や法務など, マネジメント系の具体例ではシステム開発技術やプロジェクトマネジメントなどが問われる. これらの分野で用いられる専門用語は漢字に留まらず, カタカナ語, アルファベット表記とその組み合わせが頻出する. 外国人留学生にとって, そのような専門性が高く且つ多様な表記の日本語文の意味理解が難しいのか, それとも専門分野の知識が不足しているのか, またはその両方なのかによって, 資格取得に向けた学習支援のアプローチを変えるべきである. 日本語文の意味理解が難しい場合には, 留学生にとって平易な語を併記することによって, 意味理解を支援することが出来ると考えられる⁽⁴⁾. また, 漢字の用語の意味理解

はしやすいが、カタカナ語の意味理解が難しいという学生や、逆にカタカナ語の意味理解はしやすいが、漢字の用語の意味理解が難しいという学生もいることが予想される。例えば、漢字圏の学生であれば、非漢字圏の学生よりも、漢字の用語の意味理解が得意な学生が多いと予想できる。英語が得意な学生であれば、カタカナ語から英語表記が想像できれば、意味理解がしやすいと予想できる。そういった個々の特性に合わせた学習支援アプローチを行う仕組みも今後は大事になってくると考えている。

本稿ではこうした留学生に向けた IT パスポートの学習支援を検討する前段階として、過去 5 回分の IT パスポート試験の設問と選択肢に対し、ストラテジ、マネジメント、テクノロジー系の各分野において、どういった名詞及び名詞複合語が頻繁に使われているのかを分析した結果を報告する。また、それらの語がどの程度の難易度の語であるかを既存の日本語難易度表を用いて分析した結果を議論する。

2. 先行研究

IT パスポート試験に対する研究は幅広く行われている。鈴木は IT パスポート試験の問題のストラテジ分野とマネジメント分野に対して、学習管理システム (LMS) を学習プロセスの様々な場面において活用することで、学習プロセスを蓄積・可視化し、学習活動に役立てようとしている⁽⁵⁾。また、河村・加藤も IT パスポート試験対策向けの e-Learning の教材を開発している⁽⁶⁾。こちらは、資格取得向け e-Learning における学習者のインセンティブに関する考察である。この研究の結論としては、資格取得向けの e-Learning といえどもきめ細かい指導が必要なこと、また、学生へのインセンティブとして e-Learning での学習状況を授業の平常点に加味することといったことが述べられている。更に、林坂も IT パスポート試験のための学習管理システムの開発を行なっている⁽⁷⁾。こちらの特徴としては、メモ機能を実装していることや、IT パスポート試験でどのような用語が頻繁に出題されているかを表示する機能を実装していることである。また、自身の到達度や正解率の表示といった機能も実装されて

いる。これらの先行研究には、本研究を今後 e-Learning システムとして発展させていくためにとっても有益な情報である。本研究との差異としては、これらの先行研究が、利用者として日本語母語話者の学生を想定していることである。本研究では、日本語学習者である外国人留学生を対象としている。外国人留学生にとって、どのような学習支援機能が実装されれば、継続して学習を進めていくことができるのかを考察し、システムを開発する。

外国人留学生を対象とした学習支援に関する研究も行われている。劉・内田は日本語を学習する外国人向けの日本語テキスト難易度判定手法に関する研究を行なっている⁽⁸⁾。これは日本語能力試験の問題集を難易度推定用のデータベースとして使用している点で参考になる。また、推定したテキストの難易度は、日本語能力試験の受験級の形で提示する点も、留学生にとってわかりやすいと考えられるため参考になる。川村・北村も、日本語学習者のための文章の難易度判定システムを構築している⁽⁹⁾。これは入力文の語彙の難易度と一文の長さに着目して、文章全体の難易度や語彙の難易度を表示するシステムである。更に北村は、日本語読解学習支援システム「リーディング・チュウ太」を開発し、ウェブ上で公開している⁽¹⁰⁾。現在のリーディング・チュウ太のレベル分けは、「N1」や「N2N3」といった日本語能力試験のレベルで表示されており、日本語学習者である外国人留学生からするとわかりやすい仕組みになっている。日本語能力試験 1 級以上の単語に関しては、「級外」というレベルが設けられている。これらの先行研究は外国人留学生をはじめとする日本語学習者を対象としていることで、本研究の参考になる部分が多い。本研究との差異としては、IT パスポート試験の学習支援に特化したシステムではないという点である。

難解語を平易化する手法に関する研究としては、呉が複数の言語資源を用いて、難解語の平易化を行なっている⁽⁴⁾。また、梶原・西原らは日本語の語彙平易化のために、評価用データセットや辞書、実装や評価を支援するツールキットといった言語資源を整備している⁽¹¹⁾。この研究では、日本語教育語彙表⁽¹²⁾を使用してい

る。本研究においても、語の難易度を付与するために日本語教育語彙表を用いることとする。これらの平易化に関する知見は、本研究にとって非常に有益な知見である。本研究においても、今後 IT パスポート試験で使われる語に対し、より平易な語を併記することで文章の意味理解を促進したいと考えている。

3. 語の出現頻度と難易度の分析

3.1 分析対象

本稿では、過去 5 回分の IT パスポート試験の設問と選択肢を分析対象とした。対象とした 5 回は、令和 3 年春、令和 2 年秋、令和元年秋、平成 31 年春、平成 30 年秋の試験問題である。尚、令和 2 年春の試験問題は公開されていない。データは多くの IT パスポート受験者が学習に使用するウェブサイトである IT パスポート試験ドットコム⁽¹³⁾のテキストデータを使用した。尚、設問や選択肢中の画像データ内の文字情報は今回は対象外としている。

ストラテジ、マネジメント、テクノロジー系の各分野の設問と選択肢のテキストデータを収集し、それぞれ JUMAN⁽¹⁴⁾を用いて形態素解析を行なった。各分野の試験ごとの問題数は表 1 の通りである。

表 1 各分野の問題数

年	期	ストラ テジ	マネジ メント	テクノロジー
令和 3 年	春	1-35 問 (35 問)	36-55 問 (20 問)	56-100 問 (45 問)
令和 2 年	秋	1-35 問 (35 問)	36-55 問 (20 問)	56-100 問 (45 問)
令和元年	秋	1-35 問 (35 問)	36-55 問 (20 問)	56-100 問 (45 問)
平成 31 年	春	1-35 問 (35 問)	36-56 問 (21 問)	57-100 問 (44 問)
平成 30 年	秋	1-35 問 (35 問)	36-56 問 (21 問)	57-100 問 (44 問)

収集したテキストデータから抽出対象とする語は、名詞及び名詞複合語（名詞の連続があった場合にそれ

らを結合してできた語）である。尚、この時に接頭辞は直後の名詞と結合させた。接尾辞に関しては、「性」、「化」、「率」、「者」、「氏」、「家」、「数」のみを直前の名詞と結合させた。理由は、これらの接尾辞は IT パスポート試験でよく使われる用語を形成する語であるからである。具体例としては、「機密性」、「標準化」、「稼働率」、「利用者」、「X 氏」、「専門家」、「アクセス数」といった用語が挙げられる。また、名詞の中で"こと"、"もの"などの語である「形式名詞」と、"ため"、"場合"などの語である「副詞的名詞」に関しては名詞として扱わないこととした。

3.2 語の出現頻度の分析結果

各分野及び全分野の設問及び選択肢の、名詞及び名詞複合語の語彙数は表 2 の通りである。また、設問及び選択肢の合計の数も記載している。但し、設問と選択肢で同じ語がある場合は 1 つの語として数えているため、合計は設問と選択肢の語数の和より少なくなっている。

ストラテジに関しては、100 問中 35 問だけであるにも関わらず（表 1）、設問、選択肢、合計にて最も多くの語彙が抽出されている。これによってストラテジの問題においては、設問文及び選択肢文が長い傾向にあると予想できる。また、テクノロジーに関しては、設問では 950 に対し、選択肢では 448 という結果になっている。設問と比べて選択肢の文は短い傾向にあり、この分野においては特に設問文の理解が大事になると考えられる。

表 2 各分野及び全分野の設問及び選択肢の語彙数

対象	ストラ テジ	マネジ メント	テクノ ロジ	全分野
設問	1,001	541	950	2,088
選択肢	1,198	633	448	1,915
合計	1,838	932	1,203	3,242

表 3 は各分野の語彙種別の語彙数を示した表である。語彙種は、漢字のみで形成された語（漢字）、カタカナのみで形成された語（カタカナ）、アルファベットのみ

で形成された語(アルファベット), これらと記号や数字, ひらがなを含め2つ以上が組み合わせられた語(MIX)の4種類とした。MIXの具体例としては、「IoT デバイス」や「E-R 図」といった語である。丸括弧内の数字は割合を示している。少数第4位で四捨五入をした値で表示をしている。

3分野とも漢字の出現頻度が高いことが見て取れる。ストラテジとマネジメントでは約70%, テクノロジでは約60%が漢字の語彙である。これは非漢字圏の学生にとっては文の意味理解が難しくなる要因となろう。カタカナ, アルファベット, MIXは割合こそ少ないものの, これらは専門用語の可能性が高い。

表4は各分野の出現順位別の頻出語と出現回数を記した表である。上位10単語をそれぞれ記した。丸括弧内の数字は出現回数である。これらの語のほとんどが, 専門用語ではなく日本語母語話者には馴染みのある一般用語である。これらは頻出語であることから, ITパスポートの学習を開始するにあたり, 最初に覚えるべき語彙である。しかしながら, これらの語彙の中には外国人留学生にとっては難易度の高いものが含まれている。例えば, 日本語読解学習支援システムであるリーディング・チュウ太⁽¹⁰⁾で「業務」という語をレベル判定すると, 日本語能力試験1級(N1)レベルと判定される。また, ITパスポート試験で求められる専門用語の出現回数は, これらの一般用語と比べると低い。例えば, 出現回数が4回だった専門用語として, ストラテジでは「RPA」や「経常利益」, マネジメントでは「インシデント管理」や「リバーズエンジニアリング」, テクノロジでは「公開鍵暗号方式」や「IoTシステム」などの語が挙げられる。これらの語は, 出現頻度の高い一般用語と組み合わせられて設問文や選択肢文に使用される。

表3 各分野の語彙種別の語彙数と割合

語彙種	ストラテジ	マネジメント	テクノロジー
漢字	1,243 (0.676)	634 (0.680)	714 (0.594)
カタカナ	222 (0.121)	122 (0.131)	163 (0.135)

アルファベット	98 (0.053)	24 (0.026)	77 (0.064)
MIX	275 (0.150)	152 (0.163)	249 (0.207)

表4 各分野の出現順位別の頻出語と出現回数

出現順位	ストラテジ	マネジメント	テクノロジー
1	企業(47)	開発(42)	記述(49)
2	分析(31)	実施(37)	利用(44)
3	業務(30)	作成(34)	データ(37)
4	利用(29)	作業(33)	PC(30)
5	自社(24)	プロジェクト (26)	情報(24)
6	実現(23)	ソフトウェア (24)	使用(22)
7	商品(23)	プログラム (22)	電子メール(22)
8	データ(22)	記述(21)	説明(21)
9	サービス(22)	システム(20)	入力(20)
10	システム(22)	定義(18)	暗号化(20)

3.3 語の難易度の分析結果

語の難易度を分析するために, 本研究では, 日本語学習辞書支援グループ(2015)「日本語教育語彙表 Ver 1.0」(<http://jhlee.sakura.ne.jp/JEV/>)⁽¹²⁾を利用した。これは, 17,920語に対し6段階の難易度をつけた語彙表である。今回分析対象としている過去5回分のITパスポート試験の語彙3,242語を, この語彙表を用いて分析した。

表5を確認すると, 上級前半の語彙には「分析」や「顧客」といったITパスポート試験において頻出する語がある。また, 日本語教育語彙表には存在しなかった語を「未対応」とした。未対応の語彙が2,187語あり, 全体の約67%にのぼる。これらは専門用語である語も多く存在すると考えられる。

表6を確認すると, 日本語の語彙表を使用しているため, 漢字の語彙には約46%が語彙表から難易度の付与がされている。しかし, カタカナの語彙には約26%

だけである。

表 7 は各分野の難易度別の語彙数を表している。ストラテジ、マネジメント、テクノロジーの分析対象の語数はそれぞれ、1,838 語、932 語、1,203 語である (表 2)。また、表内の丸括弧内の数字は割合を示している。少数第 4 位で四捨五入をした値で表示をしている。中級後半がどの分野においても約 20% であり、比較的多くの語彙が使われていることが確認できる。難易度別の割合は全ての分野でほぼ同じような値になっている。

表 8 は出現順位別の IT パスポートの語とその難易度を表したものである。これはストラテジ、マネジメント、テクノロジーの 3 つの分野を合わせた語彙の出現順位である。難易度が中級後半や中級前半の語が多いことが確認できる。これらは出現頻度が高いことから、これらの語彙の意味が理解できていないと文の意味を理解することが難しい。専門用語を習得することもさることながら、これらの出現頻度の高い語の理解をすることが重要である。

表 9 は、各語彙種の出現順位別の難易度が未対応の語である。これらには名詞複合語が多い。名詞複合語の例として、「利用者」という語は「利用 (名詞)」と「者 (接尾辞)」が結同している。また「プロジェクトマネジメント」では「プロジェクト (名詞)」と「マネジメント (名詞)」が結合している。また、表 9 の中には、「利用者」や「PC」、「スマートフォン」といった現代においては専門性が高いとは言えない語も含まれている。未対応の語彙に関しては、このような一般的な語もあることを考慮して、今後、難易度の推定をしていくことが重要になる。

表 5 IT パスポートの語彙の難易度別の語彙数

難易度	数	IT パスポートの語の具体例
初級前半	26	時間, 会社, 意味
初級後半	40	インターネット, 自分, 問題
中級前半	184	利用, 情報, 説明
中級後半	516	記述, データ, 開発
上級前半	274	分析, 顧客, 定義
上級後半	15	委託, 法令, 蓄積
未対応	2,187	PC, 利用者, Web サイト

表 6 各語彙種の難易度別の語彙数

難易度	漢字	カタカナ	アルファベット	MIX
初級前半	19	6	0	1
初級後半	31	9	0	0
中級前半	153	29	0	2
中級後半	464	44	0	8
上級前半	254	19	0	1
上級後半	13	2	0	0
未対応	1,079	307	184	617

表 7 各分野の難易度別の語彙数と割合

難易度	ストラテジ	マネジメント	テクノロジー
初級前半	17 (0.009)	6 (0.006)	11 (0.009)
初級後半	29 (0.016)	14 (0.015)	13 (0.011)
中級前半	135 (0.073)	65 (0.070)	88 (0.073)
中級後半	354 (0.193)	207 (0.222)	239 (0.199)
上級前半	176 (0.096)	103 (0.111)	115 (0.096)
上級後半	6 (0.003)	6 (0.006)	7 (0.006)
未対応	1,121 (0.610)	531 (0.570)	730 (0.607)

表 8 出現順位別の IT パスポートの語と難易度

出現順位	出現回数	IT パスポートの語	難易度
1	86	記述	中級後半
2	81	利用	中級前半
3	65	データ	中級後半
4	60	開発	中級後半
5	59	企業	中級後半
6	58	システム	中級後半
7	52	PC	未対応

8	51	情報	中級前半
8	51	実施	中級後半
10	49	説明	中級前半
10	49	作業	中級後半

表 9 各語彙種の出現順位別の未対応語

出現 順位	漢字	カタカナ	アルファ ベット	MIX
1	利用者	サーバ	PC	A 社
2	自社	スマート フォン	ISMS	Web サ イト
3	字句	スケジュー ール	SLA	情報シ ステム
4	暗号化	プロジェ クトマネ ジメント	OSS	電子メ ール
5	策定	ブログ	IoT	IoT デ バイス
6	内部統制	メンバ	AI	B 社
7	成果物	サービ スデスク	SFA	情報セ キュリ ティ
8	認証	ブラウザ	WBS	無 線 LAN
9	効率化	データベ ース	CPU	システ ム監査
10	売上高	メールア ドレス	CRM	IoT サ ーバ

4. まとめと今後の課題

ここまで過去 5 回分の IT パスポート試験の設問と選択肢に対する語の出現頻度と難易度の分析を行なった。語彙種を漢字、カタカナ、アルファベット、MIX の 4 つに分類し、それぞれ出現頻度の分析を行った。結果は漢字の出現頻度が高く、ストラテジとマネジメントでは約 70%、テクノロジーでは約 60%が漢字の語彙であった。難易度の分析では、日本語教育語彙表⁽¹²⁾を用いて、対象語彙に対し難易度の付与を行なった。結果として、難易度の付与ができない未対応の語彙が

2,187 語あり、全体の約 67%であった。漢字の語彙に関しては約 46%が語彙表から難易度の付与ができたが、カタカナの語彙は約 26%だけであった。未対応の語彙では、「暗号化」や「プロジェクトマネジメント」といった専門性の高い頻出語の存在を確認することができた。

今後の課題としては、解析対象とする IT パスポートの設問と選択肢を増やすことや、語の難易度の指標を留学生にとってよりわかりやすくすることが考えられる。例えば、留学生にとっては、語の難易度に対して「中級後半」という記載よりも、「日本語能力試験 2 級 (N2)」という記載の方が馴染みがあるだろう。また、専門用語に関しては、日本語能力試験 1 級の単語帳でも記載はされていないだろうし、日本語教育語彙表にも記載がない。しかしながら、記載がない語（未対応の語）の全てが専門用語であるというわけではない。まずは未対応の語彙に対し、一般用語か専門用語かを判定する必要がある。判定に使うデータとして、「IT パスポート試験」シラバス(Ver.6.0)⁽¹⁵⁾を検討している。シラバスの用語例欄に記載されている語彙は、IT パスポート試験における専門用語と判断する。また、IT パスポート試験の問題における選択肢に着目し、選択肢が句や文ではなく語であった場合には、それらは専門用語だと判断することもできていると考えている。今後 e-Learning システムとして実装する際には、専門用語に対しても難易度レベルの推定をしていけるような仕組みを実装することが望ましいだろう。これに関しては、専門用語に対するユーザの正解率を参照することで推定ができると考えている。例えば、多くのユーザが理解していない（正解していない）専門用語は難易度の高い専門用語であり、逆に多くのユーザが理解している（正解している）専門用語は難易度の低い専門用語であると考えられる。

平易語の併記に関しては、難解な語に対して平易な語を関連付けた辞書を構築する必要がある。平易語としては、単純に日本語でより易しい語に言い換えることがまずは考えられる。例えば、「企業」という語よりも「会社」という語の方が易しいと感じる人が多いだろう。日本語教育語彙表では、「企業」の難易度は中級

後半であり、「会社」は初級前半である。次に考えられるのは、漢字やカタカナ語に対し英語を併記することである。外国人留学生の中には英語が得意な学生もいることや、特にカタカナが苦手な学生もいる。例えば、「strategy」と書かれたら意味が理解できるが、「ストラテジ」と書かれると strategy と結び付けができず、意味が理解できないというような状況である。また、読み方が難解な漢字に対してふりがなを付けることも、効果的な学習支援に繋がるだろう。

こういった平易語の併記は学習序盤では大きな効果を発揮すると考えているが、あくまで学習を補助する仕組みにすぎない。学習が進み到達度が上がるにつれて、こういった補助を段階的に減らしていくことで、平易語がなくても IT パスポートで問われている内容が理解できるようになると考えている。

参 考 文 献

- (1) 独立行政法人情報処理推進機構: IT パスポート試験, <https://www3.jitec.ipa.go.jp/JitesCbt/index.html>, (2022年8月24日確認)
- (2) 文部科学省, 外務省, 法務省, 厚生労働省, 経済産業省, 国土交通省: “「留学生30万人計画」骨子”, 資料6, pp.1-3 (2008)
- (3) 独立行政法人 日本学生支援機構: “2021 (令和3) 年度外国人留学生在籍状況調査結果”, pp.1-22 (2022)
- (4) 呉浩東: “複数の言語資源を用いた難解語の平易化モデル”, 情報学研究, 第8号, pp.68-72 (2019)
- (5) 鈴木大助: “IT パスポート試験対策授業における学習管理システムの活用”, 情報処理学会研究報告, Vol.2018-CE-143 No.21, pp.1-5 (2018)
- (6) 河村一樹, 加藤大: “IT パスポート試験対策向け e-Learning における学習者のインセンティブに関する考察”, e-Learning 教育研究, 9巻, pp.44-52 (2014)
- (7) 林坂弘一郎: “IT パスポート試験のための学習管理システムの開発”, 神戸学院大学経営学論集, 第17巻, 第1号, pp.61-78 (2020)
- (8) 劉志宇, 内田理: “日本語を学習する外国人を対象とした日本語テキスト難易度推定手法”, 情報処理学会研究報告, Vol.2012-IFAT-105 No.11, Vol.2012-NL-205 No.11, pp.1-5 (2012)
- (9) 川村よし子, 北村達也: “日本語学習者のための文章の難易度判定システムの構築と運用実験”, Journal CAJLE, Vol. 14, pp.18-30 (2013)
- (10) 北村達也: “日本語読解学習支援システム「リーディング・チュウ太」”, 甲南大学紀要. 知能情報学編, 第6巻, 第2号 pp.243-253 (2013)
- (11) 梶原智之, 西原大貴, 小平知範, 小町守: “日本語の語彙平易化のための言語資源の整備”, 自然言語処理, Vol. 27, No.4, pp.801-824 (2020)
- (12) 日本語学習辞書支援グループ: 日本語教育語彙表 Ver 1.0, <http://jhlee.sakura.ne.jp/JEV/>, (2015) (2022年8月24日確認)
- (13) IT パスポート試験ドットコム, <https://www.itpassportsiken.com/>, (2022年8月24日確認)
- (14) 京都大学大学院情報学研究科 黒橋・楮・村脇研究室: 日本語形態素解析システム JUMAN Ver.7.01, (2014)
- (15) 独立行政法人情報処理推進機構: “IT パスポート試験シラバス”, Ver.6.0, pp.1-60 (2021)