

多読・速読のための 語彙力測定の構想

松下達彦（東京大学）

国際シンポジウム

「日本語教育における多読・速読の理論と実践
ー多読と速読で読みの流暢さを伸ばそう！ー」

（オンライン、2021年3月27日 ）



主なトピック

1. 読解力における語彙力の位置づけ
2. 語彙テストの得点とテキストレベルをどう結びつけるか
 - A) 語彙テストの結果から予測する既知語率
 - B) どの程度の既知語数（語彙量）で、どの程度の既知語率に達するか
 - C) どの程度の既知語数（語彙量）で、どの程度の読解レベルに達するか
 - D) どの程度の既知語数（語彙量）で、生教材が読めるか
3. テキストの語彙をコントロールする条件
4. テスト設計の構想
5. J V L C 試作品（サンプル）
6. 今後の課題



読解力における語彙力の位置づけ

一言で言えば… 語彙力以外の要素も重要だが、語彙力は重要

Q. 第二言語の読解力を構成する下位能力には何がありますか。

読解以外の能力で読解力を予測するとしたら、何をテストしますか。

Jeon & Yamashita (2014)

メタ分析：4000以上の論文から基準を満たした59論文を選んで再分析
(ただし、日本語を対象にしている論文は1本だけ)

符号化 decoding (音韻処理)、音韻認識、正書法 (スペリング) 知識、形態素の知識、理解語彙量、聴解力、文法力、第1言語の読解力、作業記憶、メタ認知



第二言語の読解力を構成する下位能力

Jeon & Yamashita (2014)

読解力の構成素	対象論文数	相関	95%信頼区間
文法力	18	.85	.58-.95
語彙知識 (広さ) (=語彙量=既知語数)	31	.79	.69-.86
聴解力	14	.77	.58-.88
形態素の知識	6	.61	.52-.69
符号化 decoding (音韻処理)	20	.56	.46-.65
正書法 (スペリング) 知識	5	.51	.43-.58
音韻認識	11	.48	.39-.57
第一言語の読解力	22	.47	.35-.58
作業記憶	10	.42	.29-.53
メタ認知	10	.32	.03-.55

やはり語彙力は重要

この辺も広義の語彙知識



語彙テストの得点とテキストレベルをどう結びつけるか (1)

学習者レベル



語彙テスト

— テキスト難度



語彙頻度プロファイラー

(J-LEX, jReadability, Reading Tutor など)

- ◆初めはある学習者Aさんに合うテキストを、Aさんの既知語数レベルか、それより少し低いレベルで考えればいいと思っていた。
- ◆例えば、あるテキストが7K (7000語) までの語で95%に達するなら、Aさんの語彙サイズが7000語より少し多いぐらいであれば読めると考えていた。
- ◆でも、どうもそうではないようだ・・・



語彙テストの得点とテキストレベルをどう結びつけるか (2)

◆ 既知語のテキストカバー率と読解レベルの関係の研究がある

- 95% : テキストがある程度理解できるレベル
 - 98% : 自力で読解を楽しめるレベル
 - 100% : 速読に使う素材
- } この点は後述

(Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010など)

◆ では、既知語のカバー率をどう測るか？

語彙量テスト？ ⇒ 既知語数2000語の人は上位2000語で書かれた文章が読める？

⇒ みなと同じように頻度上位から順に学習するわけではない

⇒ 2000語を知っている人が必ず上位2000語を全て知っているとは限らない

⇒ テキストの語彙レベルを調べるだけでは既知語率を予測できない

⇒ 例えば既知語数2000語の人が上位2000語で書かれた文章を読めるとは限らない



語彙テストの得点とテキストレベルを どう結びつけるか (3)

- ◆ 日本語全体（均衡コーパス）の頻度レベル別の平均テキストカバー率はわかっている
- ◆ 個別テキストの頻度レベル別テキストカバー率は語彙頻度プロファイラー（J-LEXなど）でわかる

ならば、頻度レベル別に語彙テストをして、「カバー率×正答率」で既知語のカバー率を予測できるのでは？

⇒やってみた



語彙テストの結果から予測する既知語率

(Matsushita, 2014)

(既知語率 = 既知語のテキストカバー率)

上位15,000語における学習者語彙サイズレベル別平均予測テキストカバー率

		1K	2K	6K	10K	15K
学習者 得点範囲	語彙レベル別 予測テキストカバー率 (A)	71.1	10.2	1.2	0.5	0.3
	平均既知語数: 13,526					
131-140	正答率 % (B)	99.5	99.5	98.9	91.1	82.1
	予測既知語カバー率 (%) n=19 (A*B)	70.8	10.1	1.1	0.5	0.2
	累積予測テキストカバー率 (%)	70.8	80.9	89.7	92.2	93.6
91-100	平均既知語数: 9,576					
	正答率 % (B)	96.4	78.2	87.3	61.8	37.3
	累積予測テキストカバー率 (%) n=11	68.5	76.5	83.6	85.3	86.1
51-60	平均既知語数: 5,355					
	正答率 % (B)	88.2	65.5	50.0	25.5	11.8
	累積予測テキストカバー率 (%) n=11	62.7	69.4	73.3	74.2	74.5
11-20	平均既知語数: 1,536					
	正答率 % (B)	60.0	16.4	7.3	2.7	2.7
	累積予測テキストカバー率 (%) n=11	42.7	44.3	45.2	45.4	45.4



どの程度の既知語数（語彙量）で、 どの程度の既知語率に達するか

← 累積テキストカバー率の予測

= 各テスト項目が代表するテキストカバー率 × 各項目のレベル別正答率

上位1万五千語のうち、○千語を知っている学習者は平均的な文章中△%の語を理解できる、と予測

(○語 ⇒ △%)

14K ⇒ 94%

10K ⇒ 86%

6K ⇒ 75%

2K ⇒ 45%

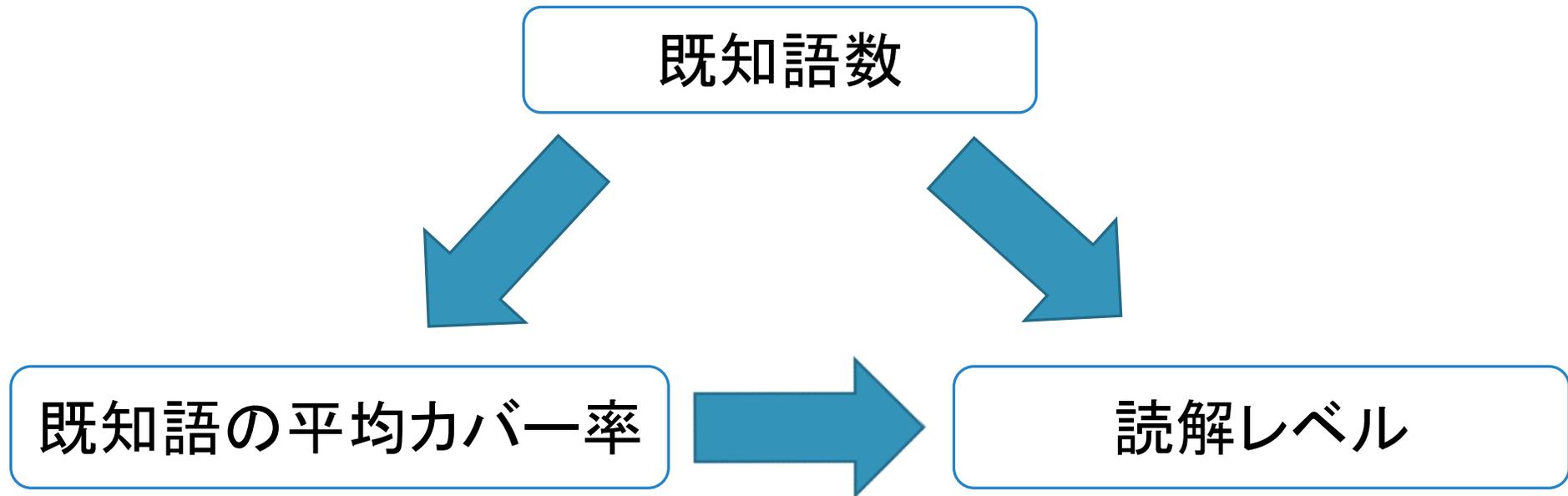


語彙テストの得点とテキストレベルを どう結びつけるか（４）

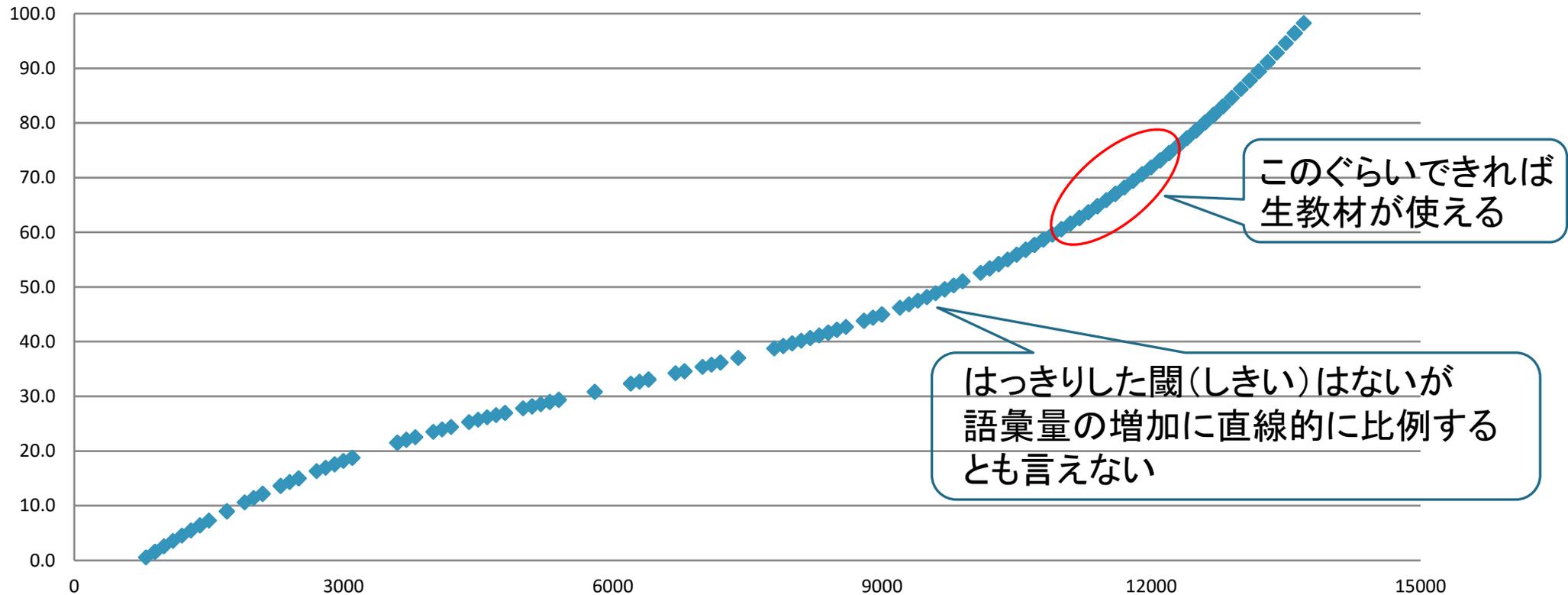
- ◆ 英語中心の先行研究で言われている既知語率（テキストカバー率）と読解レベルの関係は、本当に日本語にも当てはまるのか？
 - 日本語では・・・もう少し低い既知語率でよいのでは？
 - 96%程度に閾値？（小森ほか2004）
 - 意味的透明度の高い多数の漢字語の存在（Matsushita, 2011; 2012）
 - 文章が長いほど低頻度語彙の割合が高いはず ⇒ カバー率からの単純な予測は無理？
- ◆ 語彙テスト受験者に読解テストも受けてもらえばはっきりする
（当たり前！）
 1. 語彙テストに加えて読解テストを受けてもらう
 2. 読解テストの文章の語彙レベルを分析する（語彙頻度プロファイラー）
 3. 語彙頻度レベル別の正答率から既知語率を計算する
 4. 上の1と2を突き合わせる（回帰分析） ⇒ やって見た



語彙テストの得点とテキストレベルをどう結びつけるか (5)



回帰式による読解量の期待値 (%) (Matsushita, 2014)



上位15,000語中の推定既知語数



どれだけの既知語数（語彙量）（x）で どの程度の読解レベル（y）になるか

$$y = 0.00002x^3 - 0.0036x^2 + 0.3306x - 0.6366$$

既知語数 ⇒ 読解レベル

14,000語 ⇒ 75-100% (?)

12,000 ⇒ 65-91%

10,000 ⇒ 55-68%

8,000 ⇒ 約54%

6,000 ⇒ 約44%

4,000 ⇒ 約34%

2,000 ⇒ 約20%

これは平均的な生の日本語の場合

* 正確な予測はやはり難しい



どの程度の語彙量で生教材が読めるか

ID	Group	Score	266	237	204	187	195	222	187	158	142	156	145	161	131	152	129
			1K	2K	3K	4K	5K	6K	7K	8K	9K	10K	11K	12K	13K	14K	15K
701	G3	146	10	10	10	10	10	10	10	10	10	10	10	10	8	9	9
702	G3	140	10	10	10	9	10	10	10	9	9	9	9	10	7	8	10
703	G3	138	10	10	10	10	10	10	10	10	9	8	6	10	8	8	9
704	G3	130	10	10	10	9	8	10	10	7	7	8	8	9	7	9	8
705	G3	127	9	10	10	10	9	10	9	8	7	8	9	8	6	8	6
706	G3	125	10	10	8	8	9	10	9	8	8	10	7	8	5	8	7
707	G2	123	10	10	9	9	9	9	9	8	5	8	7	6	6	9	9
708	G2	118	10	10	10	9	10	10	7	7	6	8	6	7	6	5	7
709	G1	118	9	9	8	8	9	10	8	9	8	7	7	8	6	7	5
710	G1	118	10	9	7	8	9	9	9	8	7	7	6	7	8	7	7
711	G2	109	10	9	8	9	6	9	8	6	5	6	6	8	5	7	7
712	G2	100	10	9	7	7	8	9	8	6	6	6	3	6	5	6	4
713	G2	94	10	8	9	8	4	9	6	7	4	8	4	4	3	6	4
714	G2	93	10	10	8	5	6	8	4	6	3	5	6	7	6	5	4
715	G2	83	10	8	8	7	5	6	5	6	4	6		3	3	4	3
716	G2	70	10	8	5	5	7	7	6	2	3	3	3	4	4	2	1
717	G1	60	10	8	5	3	5	6	6	3	3	3	3	2	1	1	1
718	G2	59	10	7	6	2	5	5	5	2	5	1	2	4	1	3	1
719	G2	53	9	7	7	5	7	8	4	1	1	0	0	2	1	1	0
720	G2	52	10	9	3	2	3	4	2	1	1	1	2	3	3	4	4
721	G2	36	9	4	3	1	3	5	2	1	2	1	1	2	1	1	0
722	G1	23	9	2	2	1	2	2	2	0	0	0	0	0	0	2	1

5



どの程度の語彙量で生教材が読めるか

ID	Group	Score	266 237 204 187 195 222 187 158 142 156 145 161 131 152 129														
			1K	2K	3K	4K	5K	6K	7K	8K	9K	10K	11K	12K	13K	14K	15K
701	G3	146	10	10	10	10	10	10	10	10	10	10	10	10	8	9	9
702	G3	140	10	10	10	9	10	10	10	9	9	9	9	10	7	8	10
703	G3	138	10	10	10	10	10	10	10	10	9	8	6	10	8	8	9
704	G3	130	10	10	10	9	8	10	10	7	7	8	8	9	7	9	8
705	G3	127	9	10	10	10	9	10	9	8	7	8	9	8	6	8	6
706	G3	125	10	10	8	8	9	10	9	8	8	10	7	8	5	8	7
707	G2	123	10	10	9	9	9	9	9	8	5	8	7	6	6	9	9
708	G2	118	10	10	10	9	10	10	7	7	6	8	6	7	6	5	7
709	G1	118	9	9	8	8	9	10	8	9	8	7	7	8	6	7	5
710	G1	118	10	9	7	8	9	9	9	8	7	7	6	7	8	7	7
711	G2	109	10	9	8	9	6	9	8	6	5	6	6	8	5	7	7
712	G2	100	10	9	7	7	8	9	8	6	6	6	3	6	5	6	4
713	G2	94	10	8	9	8	4	9	6	7	4	8	4	4	3	6	4
714	G2	93	10	10	8	5	6	8	4	6	3	5	6	7	6	5	4
715	G2	83	10	8	8	7	5	6	5	6	4	6		3	3	4	3
716	G2	70	10	8	5	5	7	7	6	2	3	3	3	4	4	2	1
717	G1	60	10	8	5	3	5	6	6	3	3	3	3	2	1	1	1
718	G2	59	10	7	6	2	5	5	5	2	5	1	2	4	1	3	1
719	G2	53	9	7	7	5	7	8	4	1	1	0	0	2	1	1	0
720	G2	52	10	9	3	2	3	4	2	1	1	1	2	3	3	4	4
721	G2	36	9	4	3	1	3	5	2	1	2	1	1	2	1	1	0
722	G1	23	9	2	2	1	2	2	2	0	0	0	0	0	0	2	1

上級学習者の特徴:
7Kまでに失点が少ない

5



読解／聴解に必要な既知語率

既知語の割合 Text coverage (テキストカバー率) by known words

英語では・・・既知語率95%と98%が一つの目安

- 95%：テキストがある程度理解できるレベル
- 98%：自力で読解を楽しめるレベル
- 100%：速読に使う素材

(Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010など)

◆日本語では…多読なら95%前後でいいかも（テキストによるので既知語率は目安）
（未知語には意味的透明度の高い語も含む 例「大会」「作り変える」）

◆図版や言い換えなど、“助け”があれば必要な既知語率が下がる

◆物語文と論説文でも異なる。論説文の場合、キーワードの理解が重要

◆速読ではやはり100%既知が原則。

速読は新しい言語知識を得る場ではない ⇒Timed reading でFluencyの進歩を見る



テキストの語彙をコントロールする条件

- ◆ 上級の学生は既知語数は多いが、みなが同じ順に習得しているわけではない
- ◆ ふりがなをつける基準、語句リスト掲載語彙を決める基準はかなり保守的に。
- ◆ 7Kまで95%ぐらいできる学生ばかりだとしても、4Kぐらいの語を読めないことは十分にある
- ◆ L3で1500位、L4で2000位、L5で2500位、L6で3000位ぐらいの基準で。
 - ◆ 基本語でも用法によっては難しい
 - ◆ 複数の読みのある漢字は難しい

等々



テスト設計の構想（Ⅰ）

- ◆ まずは構想
- ◆ 目標：読み手の語彙レベルがどのレベルの読み物に合っているかを簡便に判定すること
- ◆ 条件：時間のかかるテストでない方がよい
 - ⇒ Yes/No タイプ（既知か未知かを自己診断するタイプ）のテスト
 - ある程度の知識は測れる（猪原ほか2021、など）
 - 自信度を加味？
 - Vocabulary Knowledge Scale (Paribakht & Wesche, 1993)
 - A modified version of VKS (Brown, 2008)



テスト設計の構想（2） 語知識の自信度

◆ Vocabulary Knowledge Scale (VKS) (Paribakht & Wesche, 1993)

- I. I have never seen this word before.
- II. I have seen this word before, but I do not know what it means.
- III. I have seen this word before, and I think it means _____. (synonym or translation)
- IV. I know this word. It means _____. (synonym or translation)
- V. I can use this word in a sentence: _____.

◆ A modified version of VKS (Brown, 2008) ⇒日本語訳して試す (Type 2)

- A) = I know what this word/phrase means and I can use it in a sentence.
- B) = I know what this word/phrase means, but I'm not sure how to use it.
- C) = I've seen this word/phrase before, but I don't know what it means.
- D) = I've never seen this word/phrase before.



テスト設計の構想（3）

ここまで述べてきた、一般的な語彙力と読解力の研究をそのまま各レベルに当てはめられるか？

- ◆多読が重要なのは初級～中級前半。初級と中上級では違うかも。
- ◆多読の場合、初級はより細かく判定したい（テキストも学習者も）
- ◆初級は教科書の影響が強い
 - ⇒教科書別テスト？ 各教科書に共通の語をテスト項目にする？
- ◆日本語の場合、文法も重要なものはほぼ初級に入っている
 - ⇒機能語（文法的形態素）もテスト項目に入れる？
- ◆ルビつきテキストを前提に考えれば、漢字レベルは気にしなくてよい
- ◆テキストタイプの違いやジャンルの違いは中級前半まではあまり気にしなくてよい



テスト設計の構想（４） 聴解のために

- ▶ 理解のメカニズムは、読解にも聴解にも共通する部分が多い
- ▶ 多聴もやるべき
- ▶ ただし、読解はリライトがしやすいが、聴解は難しい
- ▶ テレビ番組などで要求される語彙力は、読解より低めであることが多い
- ▶ とりあえず音声朗読付きの多読素材を使って多聴をやるのもいいかも
- ◆ ルビ付きの多読素材と多聴素材のレベル測定は、どちらもルビ付きの語彙テストをすればある程度測れる（漢字力を考慮しなくてもよい）
 - * かなを音韻に変換することが容易にできるレベルであることが前提
- ◆ 音声呈示のテストも作ることはできる



JVLC 試作品（サンプル）

Google Form を利用

[JVLC-Elem. \(Type 1\) Ver.1.0](#) （単純なチェックリスト）

[JVLC-Elem. \(Type 2\) Ver.1.0](#) （VKS: 自信度を加味）

↑

クリック

URL:

https://docs.google.com/forms/d/e/1FAIpQLSf4mUKaCEJDoRWT0mo9Y7UMzIp7Q7WhrDaJsgoIzuIUWxdHvA/viewform?vc=0&c=0&w=1&flr=0&usp=mail_form_link

https://docs.google.com/forms/d/e/1FAIpQLSeCzyvhPkV5EiBD0i3WOjb-kB90M7JLay7hVO_YkcyrUoaFyg/viewform?vc=0&c=0&w=1&flr=0&usp=mail_form_link



項目の サンプリング

『げんき』 Vol.1 Vol.2 (課順)

*語彙ランクは松下
(2011) 『日本語を読む
ための語彙データベー
ス』による

レベル	項目	会話・文法編 課数	留学生用 語彙ランク
1	いま	会L01	58
1	友だち	会L01	235
1	高校	会L01	871
1	文学	会L01	1205
1	本	会L02	166
1	靴	会L02	513
1	カーテン	会L02	1168
1	早い	会L03	179
1	日曜	会L03	527
1	前	会L04	64
2	(2) 週間	会L04	290
2	今月	会L04	636
2	土産	会L04	1123
2	楽しい	会L05	301
2	宿題	会L05	654
2	かっこいい	会L05	905
2	遊ぶ	会L06	287
2	角(かど)	会L06	540
2	男の人	会L07	97
2	鼻	会L07	384
3	やせる	会L07	1030
3	有名	会L08	364
3	空気	会L08	949
3	黒い	会L09	380
3	コンサート	会L09	1225
3	いちばん	会L10	171
3	泊まる	会L10	1041
3	売り場	会L10	1247
3	急行	会L10	1255
3	男の子	会L11	447

レベル	項目	会話・文法編 課数	留学生用 語彙ランク
4	正月	会L11	1147
4	神社	会L11	1154
4	おもちゃ	会L11	1160
4	なくす	会L12	548
4	用事	会L12	1197
4	注射	会L12	1217
4	厳しい	会L13	937
4	引き出し	会L13	1239
4	みなさん	会L14	101
4	指輪	会L14	1180
5	そば(にいます)	会L15	211
5	卒業する	会L15	999
5	地震	会L15	1207
5	起こす	会L16	870
5	台風	会L16	1139
5	選ぶ	会L17	789
5	火事	会L17	1234
5	(お湯を)沸かす	会L17	1271
5	考える	会L18	706
5	かたづける	会L18	1155
6	毎朝	会L19	643
6	召し上がる	会L19	1269
6	音	会L20	746
6	おや?	会L20	1226
6	包む	会L21	950
6	いじめる	会L21	1204
6	こと	会L22	690
6	間に合う	会L22	1141
6	拾う	会L22	1105
6	そろそろ	会L23	1012



項目の サンプリング

『げんき』 Vol.1 Vol.2 (語彙ランク順)

*語彙ランクは松下
(2011) 『日本語を読む
ための語彙データベ
ース』による

レベル	項目	会話・文法編 課数	留学生用 語彙ランク
1	いま	会L01	58
1	前	会L04	64
2	男の人	会L07	97
4	みなさん	会L14	101
1	本	会L02	166
3	いちばん	会L10	171
1	早い	会L03	179
5	そば(にいます)	会L15	211
1	友だち	会L01	235
2	遊ぶ	会L06	287
2	(2)週間	会L04	290
2	楽しい	会L05	301
3	有名	会L08	364
3	黒い	会L09	380
2	鼻	会L07	384
3	男の子	会L11	447
1	靴	会L02	513
1	日曜	会L03	527
2	角(かど)	会L06	540
4	なくす	会L12	548
2	今月	会L04	636
6	毎朝	会L19	643
2	宿題	会L05	654
6	こと	会L22	690
5	考える	会L18	706
6	音	会L20	746
5	選ぶ	会L17	789
5	起こす	会L16	870
1	高校	会L01	871
2	カッコいい	会L05	905

レベル	項目	会話・文法編 課数	留学生用 語彙ランク
4	厳しい	会L13	937
3	空気	会L08	949
6	包む	会L21	950
5	卒業する	会L15	999
6	そろそろ	会L23	1012
3	やせる	会L07	1030
3	泊まる	会L10	1041
6	拾う	会L22	1105
2	土産	会L04	1123
5	台風	会L16	1139
6	間に合う	会L22	1141
4	正月	会L11	1147
4	神社	会L11	1154
5	かたづける	会L18	1155
4	おもちゃ	会L11	1160
1	カーテン	会L02	1168
4	指輪	会L14	1180
4	用事	会L12	1197
6	いじめる	会L21	1204
1	文学	会L01	1205
5	地震	会L15	1207
4	注射	会L12	1217
3	コンサート	会L09	1225
6	おや?	会L20	1226
5	火事	会L17	1234
4	引き出し	会L13	1239
3	売り場	会L10	1247
3	急行	会L10	1255
6	召し上がる	会L19	1269
5	(お湯を)沸かす	会L17	1271



今後の課題

とりあえずGoogle Formで作ってみた

(本当は選択肢によって配点を変えられるツールがよい)

[Japanese Vocabulary Level Check for Elementary Learners \(JVLC-Elem.\) Vol.1](#)

あとは使ってみて、多読や速読のテキストのレベルとの関係を調べる

⇒各レベルの正答数とテキストの読解レベルとの関係を調べる

多読テキストは、読者によるテキスト難度の評定でよい

速読テキストは、簡単な内容理解問題の正答率をみる

=ほぼ満点が取れる人の語彙レベルをみる (多読テキストの評定より簡単か)

⇒テストそのものを検証、修正する (validation)



これからちゃんとしたものを作れるよう、
がんばります。

ご清聴、ありがとうございました。



参考文献(1)

- 猪原敬介・松尾千佳・古屋美樹・沓澤糸 (2021) 「Yes/No型と多肢選択型語彙テストの差異—読書指標との関連から—」 『心理学研究』 91(6), 367-377.
- 小森和子・三國純子・近藤安月子 (2004) 「文章理解を促進する語彙知識の量的側面—既知語率の閾値探索の試み—」 『日本語教育』 125, 83-92.
- Brown, D. (2008). Using a modified version of the Vocabulary Knowledge Scale to aid vocabulary development. *The Language Teacher* 32(12), 15-16.
- Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 160–212. <https://doi.org/10.1111/lang.12034>
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.



参考文献(2)

- Matsushita, T. (2011). Is the vocabulary learning burden of Japanese really heavier than that of English? The 17th Biennial Conference of the Japanese Studies Association of Australia. University of Melbourne, Melbourne, Australia, 6 July 2011.
- Matsushita, T. (2012). In What Order Should Learners Learn Japanese Vocabulary? A Corpus-based Approach. PhD dissertation, Victoria University of Wellington.
- Matsushita (2014). How is the relationship between vocabulary knowledge and reading comprehension? A case of Japanese. AILA World Congress 2014, Brisbane, 15 August 2014
- Paribakht, T. S. & Wesche, M.B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11(1), 9-29.

