

令和 2 年 6 月 29 日現在

機関番号：34442

研究種目：若手研究

研究期間：2018～2019

課題番号：18K12479

研究課題名(和文) The development of a Difficulty-Based Word List for Japanese EFL learners.

研究課題名(英文) The development of a Difficulty-Based Word List for Japanese EFL learners.

研究代表者

McLean Stuart (McLean, Stuart)

大阪女学院大学・国際・英語学部・専任講師

研究者番号：10624794

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：大学生863名は語彙プログラムの一環としてフォームリコールを行った。1学年かけて平均2千項目のフォームリコールを行いNGSLの難易度を推定した。語彙難易度とNGSLのランクとの相関は.45、COCAの語彙難易度とランクの相関は0.38であった。第二段階として無料のオンライン語彙レベルテストを作成した。これは、SubTlex語彙リストの最初の5千レマから5百語を使用し、テスト形式とテストプラットフォームを試験的に使用することを目的とした。第三段階では、無料オンライン語彙レベルテストでターゲットとする5千レマからランダムに語彙を選ぶよう変更を加えることで、時間が経つにつれレマの難易度が明らかになる。

研究成果の学術的意義や社会的意義

本プロジェクトのデータ収集は3段階で行われた。まず、どの語彙を学習する必要があるかを確認するために日本人大学生862名に28週間、週に最低30項目のフォームリコール(日英翻訳)を完了させ、NGSLの難易度推定を行った結果、語彙難易度とNGSLのランクとの相関関係は.45、COCAでは.38であった。次に、無料オンライン語彙テストを作成した。SubTlex語彙リストの最初の5千レマから5百語を使用し、テスト形式とテストプラットフォームの試験的な使用を目的とした。最後に、無料オンライン語彙レベルテストに語彙をランダムに選ぶよう変更を加えて、時間の経過とともに5千レマの難易度が得られるようになった。

研究成果の概要(英文)：Data was collected from 862 Japanese university students. The students completed form-recall (Japanese to English translation) as part of their vocabulary program. Over 28 weeks on average students completed 2817 form recall items resulting in difficulty estimates of the New General Service List. The correlation between the difficulty of the words and rank in the New General Service List (NGSL) was .45. The correlation between the difficulty of the words and rank in the Corpus of Contemporary American English (COCA) was .38. The second stage was to produce a free online vocabulary levels test. The goal of this stage was to use the test and pilot the test format and testing platform with 500 words from the first 5000 lemmas of a SubTlex word list. The third stage is to alter the free online vocabulary levels test so it randomly selects words from the target 5000 lemmas. This means with time stable difficulty estimates for the target 5000 lemmas are yielded.

研究分野：第二言語習得

キーワード：単語

1 . 研究開始当初の背景

A misfit between word frequency and word difficulty - Frequency-based word lists are used to develop materials, for example, vocabulary lists, textbooks, speed-reading texts, and extensive reading materials. They are also used in vocabulary tests, such as the Vocabulary Size Test (Nation & Beglar, 2017), the Listening Vocabulary Levels Test (McLean, Kramer & Beglar, 2015) and the New Vocabulary Levels Test (McLean & Kramer, 2015; McLean & Kramer, 2016), as well as in vocabulary research (Laufer & McLean, 2016; McLean, 2017; McLean, Kramer, & Stewart, 2015; Stewart, McLean, & Kramer, 2017; Stoeckel, Bennett, & McLean, 2016; Stoeckel, Stewart, McLean, Ishii, & Kramer, in press) and reading research (McLean and Rouault, 2017).

Word frequency is partially used as a proxy for word difficulty when word lists are made. However, the correlation between word frequency and word difficulty is weak. Table 1 shows that high-frequency words sometimes can be more difficult than low-frequency words. McLean, Nation, Pinchbeck, Kramer, and Dale (2017) and McLean, Pinchbeck, and Kramer (2017) showed that, among Japanese university students, the relationship between the frequency and difficulty of 149 words ranged between $r = 0.16$ and $r = 0.71$. Commonly known corpus-based word lists, such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), correlate poorly with word difficulty (BNC: $r = 0.339$; COCA: $r = 0.468$). This is in part because the Japanese language contains many English loanwords from the low-frequency bands. As a result of the strong influence of loanwords or cognates (Cobb, 2000; Elgort, 2013; Laufer & McLean, 2016) and a nation's educational system on the relative difficulty of words, a difficulty-based list should be L1 specific.

2 . 研究の目的

This KAKEN grant consists of three projects. The first project was to establish the difficulty of individual words within the New General Service List (NGSL). This word list was chosen as it is freely available to teachers, learners, and researchers.

The second project was to create a freely available online vocabulary levels test. This would be used to ensure that the online testing system was stable, and the item stems and answer banks were appropriate. The third and final goal was to then collect responses from items for the first 5000 words of the SUBTLEX Brysbaert & New (2009) lemma-based corpus. Then, from this data, the difficulty of each item will be estimated.

3 . 研究の方法

Project 1. Data was collected for this project in three stages. The first stage was to collect data from 862 Japanese university students as part of their university-wide vocabulary program. The students completed form-recall (Japanese to English translation) as part of their vocabulary program. Over 28 weeks each student was required to complete at least 30 form-recall items a week. Many students completed many more than 30-required items. On average each participant completed ??? items. The difficulty value for each word was calculated as a percentage of the participants who correctly answered a form-recall item (L1 to L2 translation) for the target word. The form-recall item presented the learner with the first letter of the target word, and after 15 seconds the learner was presented with the second letter. Learners had 30 seconds during which to answer each item.

Project 2 The second stage was to produce a free online vocabulary levels test (vocabularytest.org). The online test used a sampling rate of between 1% and 10% and allowed test creators to select meaning-recall or meaning-recognition items. As a result, it was necessary to create 500 item stems that could be used in form-recall and meaning-recall items. For the form-recall items valid, but none target answers were identified. These valid but none valid answers were added to the online test's database. If a learner answered with the valid but a none target word, for example, *cry* when the target word is *weep*, learners were given another 30 seconds during which to answer, and try to produce the word *weep*.

For the meaning-recall items, it was necessary to identify all the possible valid answers. This was first done with the help of online dictionaries. To help identify valid meaning-recall responses the online testing site provides test administrators with a list of the incorrect responses to each target item in frequency order.

Project 3 Based on the online vocabulary level test produced in project two, project 3 sets out to produce an ever-improving estimation of the difficulties of the first 5,000 lemmas of the SUBTLEX corpus Brysbaert & New (2009). In order to do this, it was necessary to write a further 4500 item stems for meaning-recall and form-recall items. Both sets of items, meaning-recall, and meaning-recognition, went through multiple-sets of production, piloting, and editing. First, a native English-Speaking with a background in vocabulary testing provided the number of examples of meaning-recall and form-recall items, and their respective valid alternative answers, and valid but incorrect answers. Valid alternative answers and valid but incorrect answers were established by looking at a number of online dictionaries. Then, a number of undergraduate English majors and one postgraduate English education major wrote over 5,000 meaning-recall and form-recall items, and respective, valid alternative form-recall answers and valid but incorrect meaning-recall answers. Following this, the one postgraduate English education major proofread all items, and valid alternative form-recall answers and valid but incorrect meaning-recall answers. Then, the one postgraduate English education major piloted all items (over 10,000 items) and checked their answers. For each question, they thought of at least 3 answers. Then, in cases where the native Japanese speaking postgraduate English major three answers were not listed in the valid alternative form-recall answers and valid but incorrect meaning-recall answers the led research was consulted. These steps were extensive and extremely timely, but necessary in order to ensure that the over 10,000 items and answer banks, will not exclude valid answers and not include invalid answers.

Data collected from learners, some paid participants, is used to producing every improving and stable estimates of increasing numbers of the first 10,000 lemmas of English. This data is still undergoing collection and analysis.

4 . 研究成果

The Spearman's correlation between frequency and difficulty for the first three 1,000-words of English was -0.44. The difficulties values for the first 426 lemmas of the NGSL are shown below. Future publications will include publicly available lists of the first 2800 lemmas of the NGSL and the first 5000 lemma of the SUBTLEX.

Flemma	% correct	Rank	Flemma	% correct	Rank	Flemma	% correct	Rank
THE	62.96%	1	ONLY	81.44%	72	OLD	92.21%	143
BE	95.00%	2	WANT	77.08%	73	OFF	6.67%	144
AND	100.00%	3	LOOK	93.81%	74	ANOTHER	63.64%	145
OF	83.67%	4	NEW	100.00%	75	DIFFERENT	41.33%	146
TO	97.73%	5	GIVE	91.21%	76	HIGH	88.16%	147
IN	97.73%	6	FIRST	76.40%	77	NEXT	93.59%	148
HAVE	97.62%	7	WAY	40.22%	78	INCLUDE	18.42%	149
YOU	97.22%	8	THING	53.19%	79	LATE	55.41%	150
IT	95.24%	9	FIND	80.22%	80	WHY	100.00%	151
HE	100.00%	10	ANY	73.63%	81	LIVE	98.68%	152
FOR	97.67%	11	OVER	60.22%	82	END	98.70%	153
THEY	87.50%	12	RIGHT	67.39%	83	WORLD	89.61%	154
NOT	97.56%	13	AFTER	76.09%	84	WEEK	92.00%	155
THAT	85.71%	14	DAY	98.90%	85	MUST	58.44%	156
WE	97.50%	15	WHERE	95.65%	86	WHILE	46.75%	157
ON	100.00%	16	MOST	23.91%	87	NEVER	67.57%	158
WITH	97.62%	17	SHOULD	52.22%	88	STUDY	92.21%	159
DO	100.00%	18	NEED	96.77%	89	KIND	94.81%	160
THIS	91.11%	19	MUCH	33.70%	90	REPORT	52.78%	161
AS	83.33%	20	HOW	96.77%	91	PLAY	33.78%	162
SHE	100.00%	21	BACK	44.44%	92	HOUSE	92.11%	163
AT	97.50%	22	MEAN	85.71%	93	GROUP	51.85%	164
BUT	97.37%	23	MAY	86.52%	94	MIGHT	28.95%	165
FROM	91.18%	24	SUCH	20.00%	95	YES	98.85%	166
BY	94.59%	25	HERE	84.27%	96	HOME	89.33%	167
WILL	79.49%	26	REALLY	66.30%	97	LET	84.42%	168
OR	96.30%	27	EVEN	59.77%	98	CASE	36.36%	169
SAY	100.00%	28	COMPANY	51.14%	99	SYSTEM	11.54%	170
GO	100.00%	29	MANY	84.44%	100	AGAIN	56.58%	171
SO	100.00%	30	CHILD	86.05%	101	WOMAN	93.15%	172
ALL	95.45%	31	TELL	34.88%	102	HEAR	81.08%	173
ABOUT	83.33%	32	LAST	6.90%	103	FAMILY	93.24%	174
IF	100.00%	33	CALL	91.46%	104	BOOK	100.00%	175
ONE	100.00%	34	DOWN	72.29%	105	SEEM	24.00%	176
WOULD	46.43%	35	BEFORE	54.55%	106	AROUND	67.12%	177
KNOW	100.00%	36	MAN	94.32%	107	DURING	15.58%	178
THERE	94.74%	37	THROUGH	33.33%	108	KEEP	78.08%	179
WHICH	90.00%	38	SHOW	61.36%	109	BIG	97.14%	180
CAN	100.00%	39	LIFE	94.32%	110	FOLLOW	37.84%	181
GET	95.83%	40	BETWEEN	45.88%	111	EVERY	54.17%	182
THINK	95.24%	41	FEEL	92.77%	112	QUESTION	87.84%	183
LIKE	95.83%	42	PLACE	54.02%	113	UNDER	97.22%	184
MORE	95.45%	43	CHANGE	89.53%	114	IMPORTANT	65.28%	185
WHO	93.04%	44	LONG	95.12%	115	ALWAYS	69.86%	186
WHEN	100.00%	45	TOO	75.61%	116	FRIEND	87.67%	187
WHAT	93.28%	46	PAUSE	10.13%	117	HOWEVER	76.71%	188
MAKE	49.57%	47	STILL	8.64%	118	SET	59.72%	189
TIME	92.92%	48	WRITE	32.50%	119	HAND	100.00%	190
SEE	95.54%	49	PROBLEM	49.40%	120	PROVIDE	13.33%	191
UP	96.43%	50	TALK	80.49%	121	SMALL	80.00%	192
PEOPLE	74.11%	51	TRY	87.80%	122	TURN	62.34%	193
SOME	88.99%	52	SOMETHING	78.48%	123	STATE	4.11%	194
OUT	96.36%	53	UNCLEAR	2.60%	124	BEGIN	48.61%	195
GOOD	92.66%	54	SAME	68.35%	125	RUN	98.59%	196
OTHER	82.08%	55	GREAT	74.03%	126	SINCE	76.06%	197
YEAR	94.29%	56	NUMBER	66.67%	127	EARLY	54.29%	198
WELL	90.74%	57	LEAVE	58.97%	128	MONEY	85.92%	199
BECAUSE	70.00%	58	LITTLE	66.23%	129	FEW	75.00%	200
VERY	99.06%	59	BOTH	70.00%	130	BRING	68.49%	201
JUST	92.45%	60	MEET	94.94%	131	MARKET	60.87%	202
NO	100.00%	61	HELP	100.00%	132	INFORMATION	54.79%	203
TAKE	60.78%	62	OWN	73.08%	133	AREA	73.97%	204
COME	92.08%	63	ASK	75.32%	134	MOVE	90.14%	205
COULD	29.00%	64	PART	60.76%	135	BUSINESS	9.72%	206
USE	97.06%	65	COUNTRY	74.03%	136	SERVICE	1.39%	207
WORK	94.95%	66	PUT	96.15%	137	GOVERNMENT	11.27%	208
THEN	40.00%	67	POINT	48.72%	138	FACT	52.11%	209
NOW	98.00%	68	START	94.67%	139	ISSUE	20.27%	210
ALSO	31.63%	69	SCHOOL	50.00%	140	THANK	82.19%	211
THAN	37.00%	70	EACH	75.68%	141	LARGE	72.60%	212
INTO	81.19%	71	BECOME	60.00%	142	RESULT	27.78%	213
ORDER	79.17%	214	SIDE	23.53%	285	STOP	62.90%	356
READ	97.18%	215	TRAIN	89.86%	286	SEVERAL	6.25%	357
MONTH	66.20%	216	SOON	57.35%	287	PERIOD	6.45%	358
INCREASE	5.71%	217	LOW	85.29%	288	CLASS	76.56%	359
NAME	95.52%	218	PRICE	60.87%	289	MATTER	17.46%	360
LOVE	100.00%	219	PUBLIC	3.13%	290	FOOD	96.83%	361
WORD	97.26%	220	OFTEN	69.01%	291	SOCIAL	52.46%	362

WITHOUT	70.83%	221	RATE	5.80%	292	REQUIRE	4.92%	363
OPEN	69.12%	222	POSSIBLE	68.66%	293	POLITICAL	14.75%	364
PAY	98.61%	223	LEAST	26.87%	294	WIN	98.31%	365
OFFER	11.59%	224	PARENT	52.24%	295	DECIDE	45.16%	366
BUILD	21.13%	225	CONSIDER	38.46%	296	STAFF	48.33%	367
HOLD	14.49%	226	EFFECT	31.82%	297	FIGURE	5.08%	368
HAPPEN	56.34%	227	RATHER	2.99%	298	REAL	79.37%	369
AGAINST	24.29%	228	CONTROL	35.29%	299	FUTURE	83.05%	370
AWAY	47.89%	229	STORY	52.31%	300	POLICY	16.67%	371
JOB	94.37%	230	LOCAL	52.31%	301	ANSWER	67.74%	372
BUY	96.97%	231	ANYTHING	53.97%	302	LAUGH	32.26%	373
THOUGH	21.74%	232	TOGETHER	31.25%	303	AMONG	7.46%	374
TODAY	100.00%	233	VALUE	47.76%	304	REMAIN	14.06%	375
EXAMPLE	60.29%	234	HARD	75.81%	305	AGO	71.88%	376
BELIEVE	53.62%	235	STAND	91.04%	306	TYPE	21.31%	377
PLAN	60.87%	236	VISIT	89.55%	307	SHOP	66.15%	378
SECOND	92.65%	237	WATCH	66.67%	308	SECURITY	18.33%	379
PROGRAM	9.86%	238	COLOR	76.06%	309	RECEIVE	29.03%	380
STUDENT	95.71%	239	PARTY	49.28%	310	MINUTE	49.15%	381
FORM	36.23%	240	CONTINUE	45.59%	311	NOTE	25.86%	382
YOUNG	83.82%	241	EVER	58.21%	312	FUND	18.97%	383
LEAD	51.43%	242	EYE	92.54%	313	TOP	80.95%	384
FACE	92.75%	243	BASE	37.10%	314	GAME	63.79%	385
CLOSE	30.88%	244	CONCERN	7.46%	315	INVOLVE	15.00%	386
ROOM	89.86%	245	LETTER	88.24%	316	ACCOUNT	15.52%	387
HOPE	63.24%	246	CENTER	67.14%	317	HALF	67.80%	388
COST	80.60%	247	LOSE	55.38%	318	HISTORY	89.47%	389
HEAD	87.30%	248	YET	97.01%	319	CREATE	36.67%	390
UNDERSTAND	88.24%	249	ALMOST	55.22%	320	BREAK	61.29%	391
HOUR	86.11%	250	DEVELOPMENT	16.18%	321	MOMENT	53.57%	392
FAR	70.59%	251	ALREADY	33.82%	322	INDIVIDUAL	13.56%	393
SPEND	22.73%	252	TEST	94.44%	323	ACROSS	49.18%	394
CAR	97.01%	253	PROBABLY	19.12%	324	EITHER	36.07%	395
ACTUALLY	27.14%	254	SALE	19.12%	325	MUSIC	95.24%	396
LEVEL	20.90%	255	SUGGEST	24.24%	326	FURTHER	6.25%	397
CITY	92.06%	256	NOTHING	81.16%	327	REACH	30.00%	398
PRESENT	82.54%	257	WHOLE	16.42%	328	CLEAR	61.67%	399
LESS	39.39%	258	CARE	16.67%	329	RULE	55.00%	400
IDEA	82.09%	259	DEAL	21.21%	330	COMPUTER	13.56%	401
REASON	65.22%	260	LANGUAGE	54.93%	331	WAIT	88.52%	402
LEARN	71.64%	261	SEND	77.05%	332	SOUND	91.80%	403
UNTIL	52.86%	262	FALL	71.21%	333	TEAM	80.65%	404
MEMBER	81.82%	263	EXPECT	18.46%	334	ALONG	34.92%	405
PROCESS	15.94%	264	RETURN	45.16%	335	RESEARCH	16.67%	406
PERSON	27.27%	265	WATER	95.31%	336	APPEAR	25.00%	407
EXPERIENCE	37.88%	266	ALLOW	3.08%	337	DRIVE	93.33%	408
NIGHT	83.33%	267	PER	33.33%	338	ACTIVITY	24.19%	409
SUPPORT	5.97%	268	CAUSE	37.50%	339	BLACK	88.89%	410
SURE	37.14%	269	POWER	66.15%	340	PRODUCE	28.57%	411
SORT	7.58%	270	SIT	98.51%	341	FREE	88.89%	412
QUITE	17.46%	271	WALK	72.73%	342	GENERAL	32.20%	413
BAD	88.06%	272	MOTHER	96.97%	343	BODY	80.33%	414
ONCE	23.81%	273	SUBJECT	23.08%	344	PLEASE	71.67%	415
ENOUGH	56.72%	274	DEVELOP	47.62%	345	TOWARD	22.03%	416
ALTHOUGH	11.94%	275	STAY	23.08%	346	SENSE	24.14%	417
WITHIN	8.96%	276	RECORD	29.23%	347	PERHAPS	5.17%	418
AGE	65.71%	277	MIND	60.94%	348	EVERYTHING	59.65%	419
TERM	13.24%	278	REMEMBER	68.25%	349	ADD	78.18%	420
WHETHER	17.91%	279	PAST	68.25%	350	LAW	59.65%	421
ABLE	54.55%	280	OFFICE	40.00%	351	SELL	82.76%	422
SHARE	47.76%	281	FORCE	15.63%	352	EASY	89.09%	423
LINE	85.94%	282	GROW	49.18%	353	FULL	25.42%	424
PRODUCT	22.73%	283	TOWN	87.50%	354	FILM	70.97%	425
SPEAK	14.93%	284	LIGHT	85.48%	355	MODEL	54.55%	426

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 1件/うちオープンアクセス 1件）

1. 著者名 Nakata, T., Tada, T., McLean, S., & Kim, Y.	4. 巻 Onlinefirst
2. 論文標題 Effects of Distributed Retrieval Practice Over a Semester: Cumulative Tests as a Way to Facilitate Second Language Vocabulary Learning	5. 発行年 2020年
3. 雑誌名 TESOL Quarterly	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） doi.org/10.1002/tesq.596	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 McLean, S., Stewart, J., & Batty, A.	4. 巻 Onlinefirst
2. 論文標題 Predicting L2 reading proficiency with modalities of vocabulary knowledge A bootstrapping approach.	5. 発行年 2020年
3. 雑誌名 Language Testing	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） doi.org/10.1177/0265532219898380	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Stoeckel, T., McLean, S., & Nation, P.	4. 巻 Onlinefirst
2. 論文標題 LIMITATIONS OF SIZE AND LEVELS TESTS OF WRITTEN RECEPTIVE VOCABULARY KNOWLEDGE	5. 発行年 2020年
3. 雑誌名 Studies in Second Language Acquisition	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Stoeckel, T., Stewart, J., McLean, S., Ishii, T., Kramer, B., & Matsumoto, Y.	4. 巻 87
2. 論文標題 The relationship of four variants of the Vocabulary Size Test to a criterion measure of meaning recall vocabulary knowledge	5. 発行年 2019年
3. 雑誌名 System	6. 最初と最後の頁 102-161
掲載論文のDOI（デジタルオブジェクト識別子） org/10.1016/j.system.2019.102161	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件 / うち国際学会 5件）

1. 発表者名 McLean, S.
2. 発表標題 Correlations of modalities of vocabulary knowledge to L2 Reading Proficiency: A bootstrapping approach.
3. 学会等名 14th Annual Cam TESOL Conference on English Language Teaching- (Phnompenh,Cambodia) (国際学会)
4. 発表年 2019年

1. 発表者名 McLean, S.
2. 発表標題 Evidence of the effectiveness and efficiency of extensive reading at developing Japanese university students' reading rates.
3. 学会等名 KOTESOL-(Seoul, Korea) (国際学会)
4. 発表年 2018年

1. 発表者名 McLean. S.
2. 発表標題 Predicting L2 reading proficiency with modalities of vocabulary knowledge A bootstrapping approach
3. 学会等名 Thai TESOL (国際学会)
4. 発表年 2020年

1. 発表者名 McLean. S.
2. 発表標題 Online self-marking typing, speaking, listening or reading vocabulary levels tests
3. 学会等名 KOTESOL-(Seoul, Korea) (国際学会)
4. 発表年 2019年

1. 発表者名 Rouault, G., & McLean, S.
2. 発表標題 Evidence of the effectiveness and efficiency of extensive reading at developing Japanese university students' reading rates
3. 学会等名 American Association for Applied Linguistics (AAAL)-(Chicago, America) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----