

# WebKIC の利用者データによる Kanji in Context 語彙の検討 —学習者の感じた習得難易度と、単語親密度を中心に—

秋 澤 委太郎

## 【要旨】

Kanji in Context 学習用の Web アプリケーション WebKIC は、2017 年のアップデートで、利用者それぞれの意向に応じて出題内容をカスタマイズすることが可能になり、学習履歴も保存できるようになった。保存されるのは、利用者が単語に「スター」を付けたかどうか、そして、クイズモードにおける各単語の学習状況である。この機能を利用した 2017-18 年度ならびに 2018-19 年度の学生から収集したデータのうち、スター関連のものを総合的に検討することで、利用者がどの単語の習得を困難に感じているかが判明し、それを天野・近藤 (1999) による単語親密度と考えあわせることによって、Kanji in Context 所収単語の傾向が見出された。もって学生指導や今後の改訂の参考に供するものである。

## 【キーワード】

WebKIC、『Kanji in Context』、WebKIC 利用者データ、単語の習得難易度、単語親密度

## 1 本稿の目的

WebKIC とは、2014 年より筆者が開発、運用している漢字教育ならびに学習用 Web アプリケーションである。本センター作成の『[Kanji in Context \[改訂新版\]](#)』（以後 KIC と略記）に準拠しており、教員が漢字テスト問題を作成するため、そして学生が KIC を自習するために活用している<sup>1</sup>。

2017 年におけるアップデートで、利用者は自分の学習項目をカスタマイズでき、学習履歴を記録できるようになった。具体的には、「スター」機能により自分が望む単語だけを選んで学習することができる。また「チェックマーク」システムによって学習履歴が記録される<sup>2</sup>。

利用者個人が学習をより効率化するため、そして自分の到達度を確認するためにこれらの機能が役立つのは当然であるが、全ての利用者のデータは本センターが管理するサーバーに一括して保存されているため、教育目的でそのデータを活用することも技術的に容易になっている。本稿は、そのデータが何を示唆するかを検討するものである。

## 2 WebKIC が収集しているデータ

WebKIC が利用者から収集しているデータは以下の通りである。

1. 利用者がどの単語にスターをつけたか
2. 利用者の日本語学習歴と第一言語
3. 利用者の単語ごとの学習状況（フラッシュカードでどの単語を何回テストしたか、等）

本稿はこのうち1について検討する。また、データの検討対象の属性を2によって概観する。

## 3 「スター」の意味

WebKIC には、リファレンスモードとクイズモードが搭載されている。リファレンスモードは KIC のコンテンツをインタラクティブに閲覧できる辞書のような位置付けである。クイズモードは、「確認テスト」「フラッシュカード」「文脈クイズ」の3種類が用意されている。

これらのいずれのモードにおいても、表示されている単語には「スター」を付けることができる（図1）。

どのような場合に単語にスターを付けるかは開発者の側では特に限定を設けていないが、ユーザーズ・マニュアルの中で、「単語にはスターを付けることができ、その単語だけをクイズで出題するよう設定することができる」と説明されている。この機能から考えれば、「利用者が単語にスターを付けるのは、単語が覚えにくく、特に練習が必要だと思った場合である」と解釈するのが最も自然だろう。実際、これまでのアンケート結果<sup>3</sup>や筆者が直接学生から耳にしたところによると、その解釈に合致する場合がほとんどである。本稿では、この解釈に基づいてスター関連のデータを分析する。

図1 スター



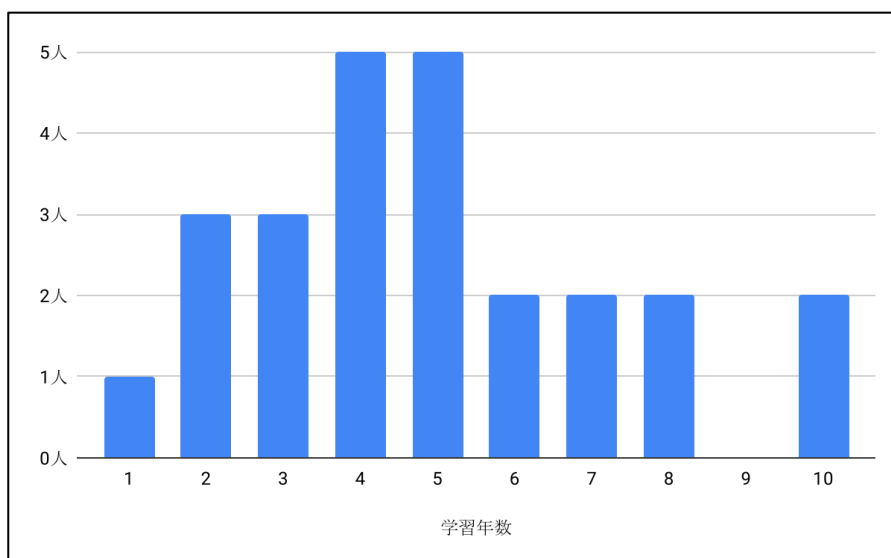
## 4 スター関連データの検討対象

本稿が検討の対象とするのは、2017-18 年度ならびに 2018-19 年度に WebKIC アカウントを作成したレギュラーコースの学生から、筆者が使用許可を得たデータである<sup>4</sup>。各年度

それぞれ15名、計30名の学生から利用許可を得た。ただし、そのうち4名はスター機能を用いておらず、データが記録されていなかった。また、他の1名はスター機能を特殊な目的で用いていたので、今回の検討対象からは除外した<sup>5)</sup>。したがって、利用することができたデータは実質25名分である。

IUC レギュラーコースは入学時に2～3年以上の日本語学習歴がある学生を主な教育対象として想定しており、KICの想定対象もこれと同じである。今回のデータ検討対象である25名の日本語学習歴は下記図2のように分布しており、ほとんどがKICの想定している学習者であるといえる。ただ、学習歴が1年と短い1名も中国語（マンダリン）母語話者であり、漢字学習の上では他の利用者と比べて特段に不利な条件ではないと言えるだろう。母語について言えば、25名のうち22名が英語であり、3名が中国語（マンダリン）であった。

図2 分析対象のIUC入学時における日本語学習歴



## 5 スター関連のデータの検討

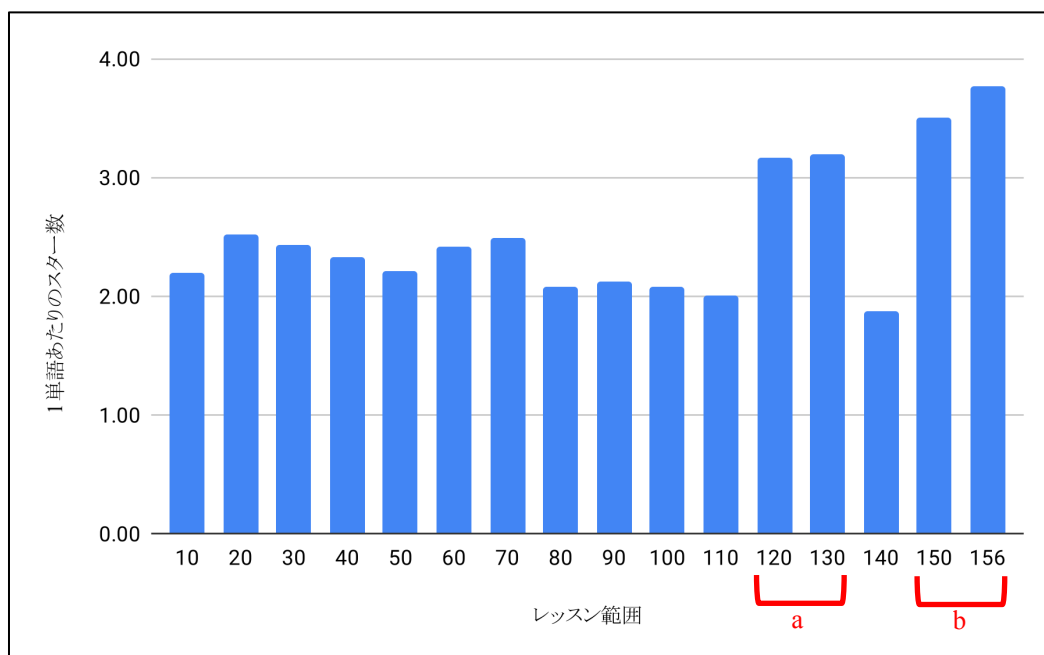
### 5-1 10レッスン（回）ごとのスター数の傾向

KICの冒頭解説によれば、KICは日本語の学習段階に対応して常用漢字を7つの水準に分けている。第1水準から第5水準までは初級から上級に対応しており、第6水準は「特別な領域や分野の特殊な語彙の中でしか使われない漢字<sup>6)</sup>」、そして第7水準は2010年の改定常用漢字表<sup>7)</sup>で新たに加えられた漢字である。それぞれの水準については、「漢字の重要度や難易度を基本に、字形や意味のつながり、語彙のつながりなど、様々な面に配慮して、水準ごとに方針を立てて配列しました。<sup>8)</sup>」とされている。このように配列された漢字とその語彙は全156回のレッスンに分けられており、学習者はそれを第1回から順番に学

んでいく。ここから、学習する単語の難度も回を追うごとに上がっていくであろうと想定される。

では、WebKICの利用者自身はKICの単語の学習難易度をどのように感じているのだろうか。全25名の利用者が合わせてどれくらいの数のスターを付けたか、これをレッスン単位で見れば、各レッスンの難しさの傾向が分かるであろう。ただし、レッスンごとに含まれる単語の数が異なるので、単純に合計スター数を比較するだけでは難しさを比較することにならない。そこで、レッスンごとの合計スター数を当該レッスン所収の総単語数で割ることで、「あるレッスンにおける1単語あたりのスター数」を算出することとした。こうすれば、レッスンの難しさを相互に比較することができる。ただ、1レッスン単位で傾向を見るのは煩瑣に思われるので、10レッスン単位で範囲を区切り、範囲ごとに1単語あたりのスター数を計算した。それが図3である。

図3 10レッスンごとの、1単語あたりスター数



興味深い傾向が見て取れる。それは、階級120と130が示すレッスン第111回～130回の範囲（以後、範囲aと呼称）と、階級150と156が示す第141回以降の範囲（以後、範囲bと呼称）で突出してスター数が多いということである。つまり、学習者は範囲aとbに含まれるユニットだけを特に難しく感じているということだ。KIC学習者への指導上、教員はこのことに留意する必要がある。

範囲aとbの突出を除けば、スター数に増加の傾向は見られない。むしろ、わずかながらも減少しているように見える。KIC所収の漢字と単語が基本的には「易」から「難」へと並べられているにも関わらず、1単語あたりスター数にはレッスンを追うに従って全体

的な増加傾向が見られないことは、何を意味するのだろうか。WebKIC 利用者に、実は学習を進めて漢字に習熟するに従い新しい漢字や単語の習得が容易になっていくという傾向があり、それが漢字ごとあるいは単語ごとの難度の増加を上回っているのかもしれない。また、データの検討対象は全て IUC レギュラーコースの学生であり、日々の授業と課題を通じ、KIC の自律学習と並行して日本語を学んでいく。これも、KIC 単語の習得難度が上がっていかない要因の一つに挙げられよう。あるいは、次節に掲載の図 4 に見られる通り、KIC の各レッスン所収単語の数は回を追うごとに減っていくので、そのことも影響しているかもしれない。範囲 a と b 以外の部分における KIC 単語の学習難易度の傾向に関しては、今後の調査でさらに検証する必要があるだろう。

## 5-2 スター数の特に多いレッスン範囲

1 単語あたりスター数が突出して多いレッスン範囲に再び目を向ければ、範囲 b は KIC の第 7 水準（第 144 回～156 回）、つまり 2010 年の改定常用漢字で追加された範囲にほぼ重なるので、学習者にとって単語が難しいと感じられるのも致し方ない。問題は、範囲 a である。問題の所在を明らかにするために、図 3 と同じく 10 レッスンごとに区切った KIC 所収単語の数を次に示す。

図 4 10 レッスンごとの KIC 所収単語数

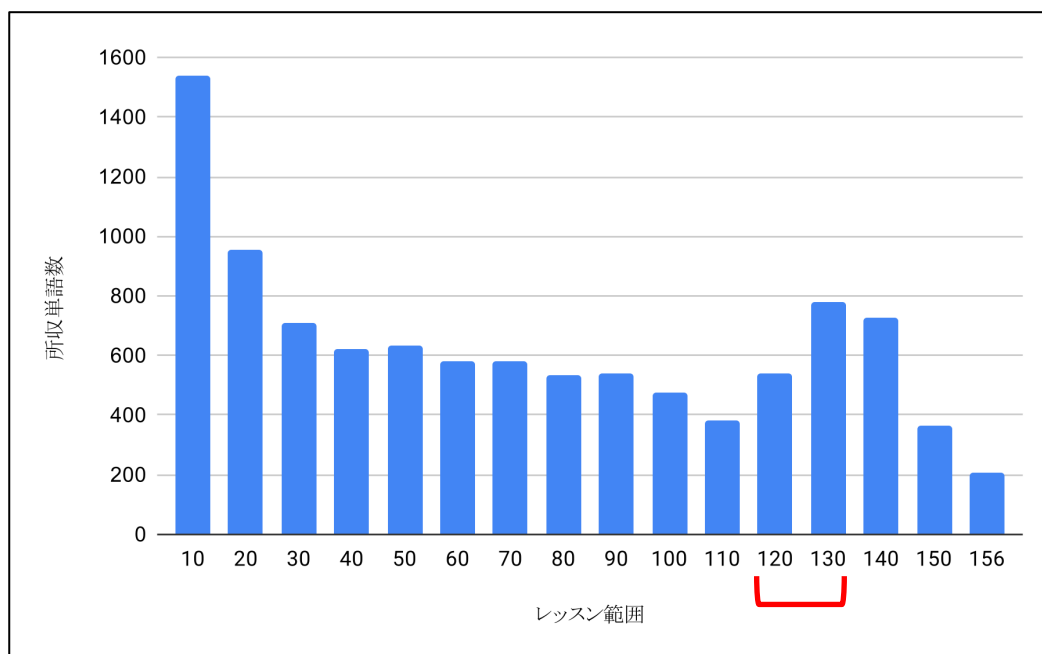


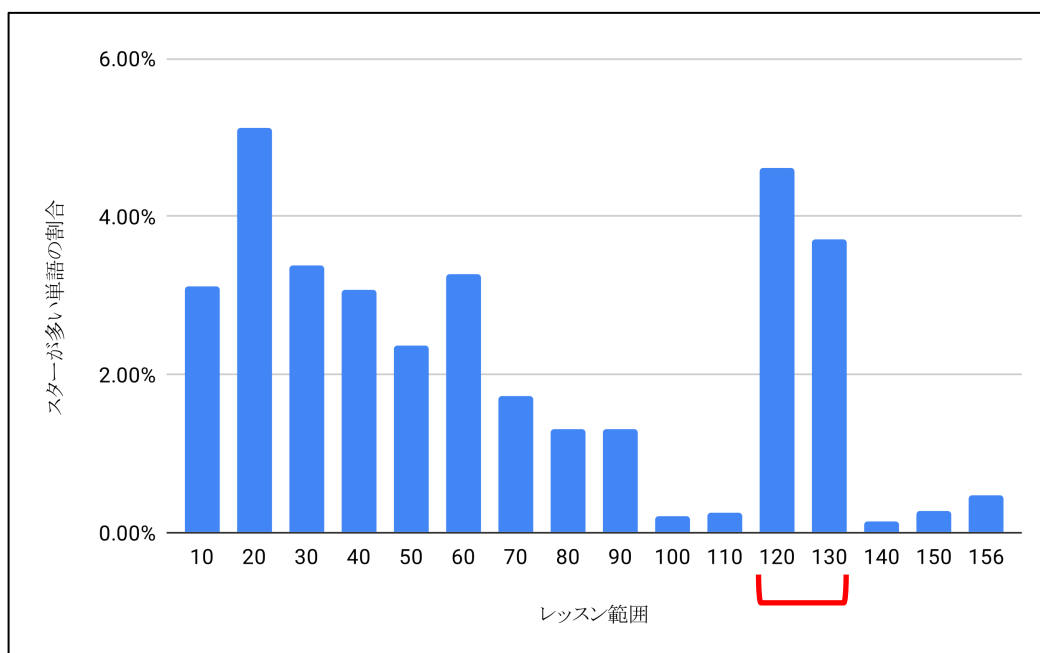
図 4 において終盤のレッスンで単語数ににわかに増加する範囲と、図 3 の範囲 a が重なっているのである（特に階級 130、すなわちレッスン第 121 回～130 回の範囲）。つまり、1 単語あたりの学習難度の高い範囲が、単語数が多い範囲と一致しているということであ

る。学習者にとり、学びの困難がにわかに強く感じられてくる範囲だと言えるだろう。

### 5-3 スターが特定単語に集中する傾向

利用者が付けるスターが特定の単語に集中しているか、あるいは多くの単語に分散しているかを見ると、また違った傾向が浮かび上がる。検討対象の半数以上 (=13 名以上) がスターを付けた単語が、10 レッスンごとに区切った範囲内にどの程度の割合で存在しているかを、下記図 5 に示す。

図 5 10 レッスンごとの「スター数が 13 以上の単語」の割合



スター数が集中している単語は KIC 冒頭では 3%から 5%の割合で存在するが、その割合は大きく落ち込んでいき、階級 110 すなわちレッスン第 101 回から 110 回の範囲では 0.26% にまで低下する。一方、5-1 の図 3 によって、冒頭から階級 110 にかけての「1 単語あたりスター数」はほぼ横ばいであることが示されている。この 2 つの結果を考え合わせると、利用者が付けるスターは、KIC 冒頭では特定の単語に集中する度合いが高いが、レッスンが進むにつれて多くの単語に分散していく傾向が見て取れる。つまり、利用者のスターの付け方が、レッスンが進むにつれて多様化していくのである。

これは何を意味するのだろうか。一つの可能性として、早いレッスンでは基本的で誰にとっても必要な単語が多く、どの単語が覚えにくいかという傾向に利用者による差が出にくい反面、レッスンが進んで高度な単語が多くなってくると、利用者の専門分野や関心領域に応じて覚えにくさを感じる単語が分散していく、という状況が考えられるのではないだろうか。

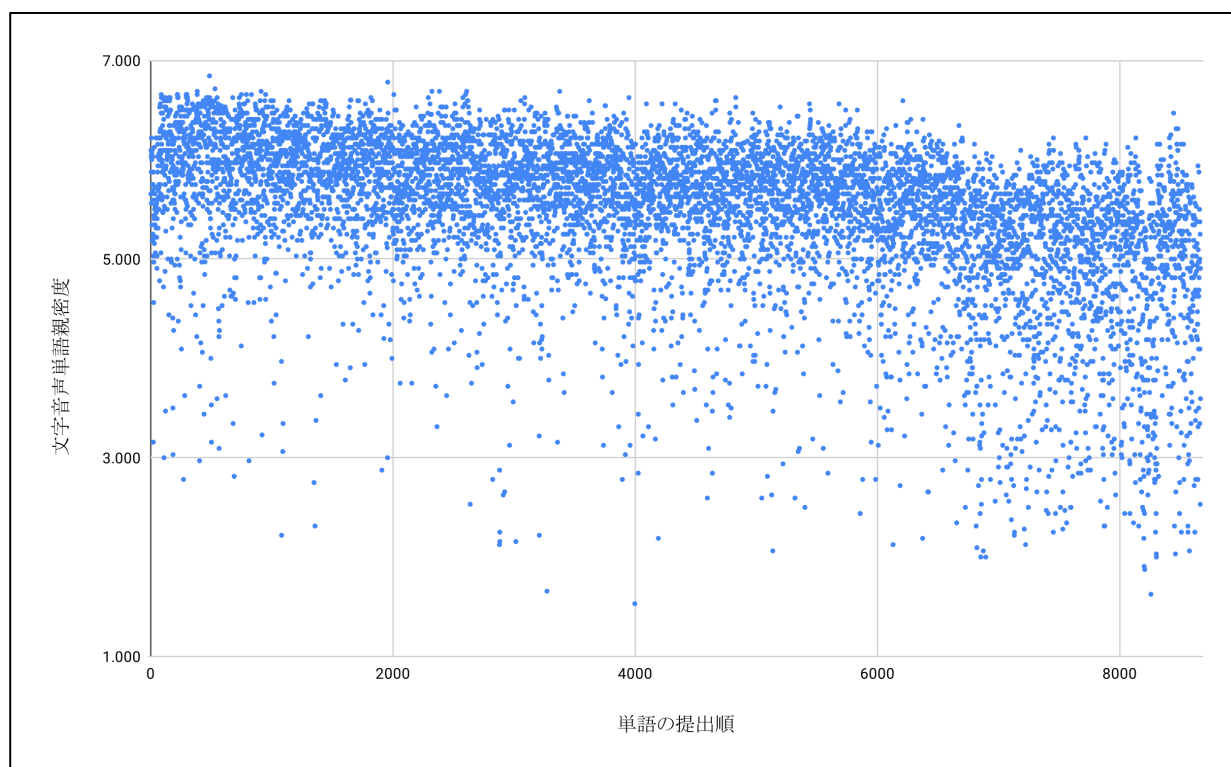
一方、階級 120 と 130 すなわちレッスン第 111 回～130 回の範囲については、スター数の特に多い単語の割合が顕著に高くなっている。これは、図 3 の範囲 a と一致する。つまりこの範囲では、レッスンあたりの学習単語数が急増するのに加え、含まれる単語が全体的に難しいという印象を学習者が持っていると同時に、学習者を問わず覚えにくいと感じられる単語も増えているということである。

## 6 KIC 単語の親密度と習得難易度

### 6-1 KIC 単語全般の文字音声親密度

単語の難しさを考える時、その一つの指標として、天野・近藤 (1999) による単語親密度が挙げられるだろう。そこで本節では、KIC 収録単語全般の親密度をそのデータベースから参照し、KIC が単語親密度の観点からどのような傾向を持っているか概観する。データベースには文字音声単語親密度と文字単語親密度、そして音声単語親密度の 3 種が掲載されているが、本稿では文字音声単語親密度の値を採用することとした。その結果を下の図 6 に示す<sup>10</sup>。

図 6 KIC 収録単語の文字音声親密度の分布



横軸方向には KIC 提出順に単語が並んでおり、縦軸には各単語の親密度をプロットしてある。両者には中程度か弱い逆相関関係が見られる。参考までに、単語提出順と親密度の

相関係数は-0.452であった。

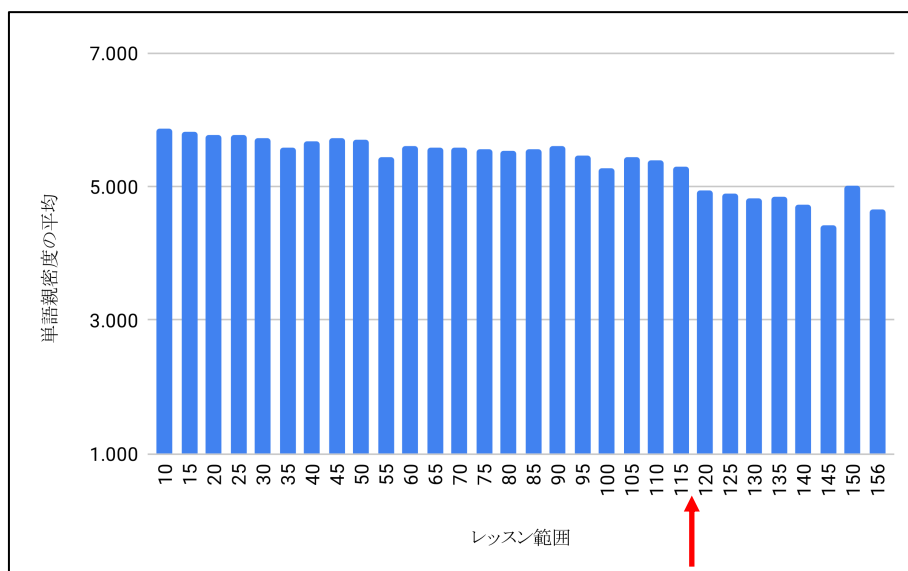
図6を見ると、3つの傾向が見て取れる。

1. KIC 所収単語の親密度は、全体的に高い（平均値は 5.468）。
2. 親密度の高い単語によって形成されている塊が、わずかに右肩下がりの傾向を示している。
3. はじめのうちはその塊の幅が狭く境界線が明瞭だが、後に下るにつれて境界が曖昧になり、幅も広くなる。

この結果は、KIC が基本的に日本語母語話者にとって親密度の高い単語を収録しつつ、レッスンを追うに従って親密度の低い単語も配置していくという傾向を示すものである。KIC は、その編集方針が適切に具現化されているということが分かる。

しかし、図6の右1/3の部分を見ると、親密度の高い単語の塊が全体の傾向以上に低く落ち込んでいることに気づく。それは、この部分に至ると所収単語の親密度がそれまで以上のペースで下がるということを意味する。より詳しく見るため、5レッスンごとの所収単語の平均親密度を下の図7によって示す。

図7 5レッスンごとの、単語親密度の平均



レッスン115まで（階級115まで）とそれ以降（階級120以降）との境目で、単語親密度がやや急な低下を示している。このポイントは、またもや図3でいう範囲aに含まれている。つまり、ユニットあたりの学習単語数が急増する範囲と、KICの学習者にとってにわかに覚えにくさが高まる範囲、利用者を問わず特に覚えにくいと感じられる単語が増える範囲、そして、単語親密度の低下がやや著しくなる時点が重なっているのである。



## 6-2 範囲 a でスター数が特に多かった単語の親密度

5-3 の図 5 で、特にスター数の多い単語が範囲 a において急増することを指摘したが、本節ではそれらの単語ならびに親密度を見てみたい（下記表 1）。表中の各単語の横の数値は、左から、スター数、レッスン、親密度となっている。単語はスター数の多い順で上から下へ、そして左から右へ配列されている。天野・近藤 (1999) のデータベースにエントリーが存在しなかった単語については、親密度は記載していない。

表 1 範囲 a でスター数が 13 以上だった単語の親密度

憤り	16	127	4.875	漆	14	121	5.031	鑑みる	13	126	2.656
行脚	16	123	4.656	滴る	14	121	4.5	晩酌	13	125	5.5
委嘱	16	117	4	携える	14	120	5.031	謹む	13	125	4.594
米俵	16	117	5.156	水浸しになる <sup>11</sup>	14	120	5.156	紛糾	13	124	4.438
詰問	15	124	4.562	拘束	14	119	5.469	粉碎	13	123	4.969
紡ぐ	15	124	4.25	如実に	14	118	4.062	狩猟	13	121	5.125
拙い	15	119	3.344	惨めな	14	118	5.25	雨漏り	13	121	5.312
唆す	15	117		恐喝	14	117	5.719	搾る	13	120	4.875
滞る	15	114	4.75	年俸	14	117	5.5	控除	13	119	5.125
潜む	15	114	4.938	苦渋	14	114	4.906	挿絵	13	119	4.812
滑らかな	15	114	5.562	潜る	14	114	5.531	俊才	13	117	3.625
醸し出す	14	128	4.344	罷免	13	129	4	潜伏期間	13	117	
朗らかな	14	127	5.25	頒布	13	128	3.844	麻	13	116	5.875
管轄	14	125	5.062	醸造酒	13	128		患う	13	111	4.75
諭す	14	124	4.188	欺く	13	127	4.594	潔癖な	13	111	5.594
諮る	14	124		鎮まる	13	126	5.219	装う	13	111	5.656
陥れる	14	122	4.188	鋳物	13	126	4.094				
滴	14	121	4.594	鋳造	13	126	3.75				

表 1 は、KIC 学習の上で大きな山場であることが判明した範囲 a の中でも、特に学習者が覚えにくいと感じている単語の一覧であり、教員として指導する際には注意を要する語が集まっていると考えられる。

一方、親密度を見ると、全 52 語のうち 40 語つまり 76.9%が、今回親密度データを参照できた KIC 単語の平均値である 5.468 を下回っている。平均を上回っている単語も、ほぼ平均付近の値にとどまっている。これらの単語の親密度は、KIC 所収の単語の中では比較的到低いといえる。

単語親密度は、日本語学習者にとっての学習難易度を調査したものではない。また、KIC

所収の単語はそもそも全般的に親密度が高い。しかし、学習者が特に覚えにくいと感じている単語は、他の KIC 単語と比べて親密度も低めになっているという傾向が見出せたことは興味深い。

## 7 結論と、それが示唆するもの

本稿は、WebKIC の利用者がどの単語にスターを付けたかというデータを複数の角度から検討してきた。

KIC は、範囲 a つまりレッスン第 111 回から第 130 回までにおいて、各レッスン所収の単語数がにわかに増加するとともに、学習者が単語の学習難度を特に高く感じていることが分かった。

レッスンが進むにつれて利用者のスターの付け方が多様化していくという傾向が KIC 冒頭から見られたものの、範囲 a では、その傾向に反して特定の単語にスターが集中する割合が顕著に高まっていた。これは、利用者を問わず習得が困難に感じられている単語が範囲 a には多いということである。また、それらの単語は文字音声親密度も低めにとどまっていた。

こうして見出された範囲 a の難しさに鑑みれば、KIC を使った漢字と語彙の学習を教員として指導するにあたり、学習者がそこまで勉強を進めた時には難しさと単語の多さに挫折しないよう気を配ることが望ましい。また、多くの学習者にとって習得が困難だとみられる単語（表 1）については特に指導上の工夫が求められるだろう。

範囲 a の難しさは、KIC の改訂によって対処できる可能性もある。単語親密度の観点からすれば KIC の語彙とその配列は全般的に適切であると言えるが、範囲 a に関しては改善の余地がありそうだ。例えば、表 1 に含まれる単語を親密度の高いものに置き換えれば、多くの学習者にとって学びやすさが高まるかもしれない。ただ、KIC はあくまでも常用漢字とその読み方ならびに意味を網羅的に学ぶための教材であるから、漢字一字を訓読みすることで成立している単語のように、親密度が低いからといって他で置き換えるわけにはいかない語も多く、その方法はあまり現実的ではない。より実現可能性の高い方法としては、範囲 a に含まれるレッスンの数を増やして各レッスンあたりの所収単語数を減らす、あるいは、範囲 a に含まれる漢字や単語の一部を他の範囲に分散させるといった手段が考えられるだろう。

今後も多くの WebKIC 利用者からデータを収集し、単語の習得難易度についての考察を深めると同時に、学習者への指導や教材の改良にその考察を役立てていきたい。

## 注

- 1 WebKIC はブラウザから利用するアプリケーションであるが、現時点で、利用できるのは本センター教員と学生、そして卒業生のみである。開発にいたった経緯については秋澤 (2015) を、機能の詳細については秋澤 (2018) をそれぞれ参照のこと。
- 2 スターならびに学習履歴のデータが記録されるのは、WebKIC にサインインするためのアカウントを持っている利用者だけである。現在のところは、センターの学生と卒業生のみがアカウントを作成できる。
- 3 秋澤 (2016) 、秋澤 (2018) を参照のこと。
- 4 データの活用のため、各年度末に全学生を対象としてメールで協力を依頼した。返信があった学生についてのみ、データの使用許可を得たものとした。データの使用を許可してくれた学生諸氏には、この場を借りて心から感謝を申し上げる。依頼メールの文面は、以下の通りである。「WebKIC 上のみなさんの学習データを、研究のために活用させてもらえませんか。もしデータを使ってもいい場合、お手数ですが、今月中にこのメールに返信を下さい。メール本文には、何も書かなくていいです。もちろん、データは全て匿名 (anonymous) で用います。利用者個人を特定することは決してありません。例えば、ある単語が覚えやすいか覚えにくい、といったことが知りたいと思っています。研究結果は、センターの発行する「教育研究年報」などで発表する予定です。使いたいデータは、以下の通りです。・どの単語にスターやチェックを付けたか。・どの練習を何回おこなったか (例：フラッシュカードでこの単語を～回見た、など)。では、どうぞご協力をお願いします！」
- 5 その学生は、「この漢字は知っている」と判断した単語にスターを付けて、その単語は飛ばして学習を進めたと申告している。
- 6 KIC 冒頭の解説 p.[9]。
- 7 [http://www.bunka.go.jp/seisaku/bunkashingikai/sokai/sokai\\_10/pdf/kaitei\\_kanji\\_toushin.pdf](http://www.bunka.go.jp/seisaku/bunkashingikai/sokai/sokai_10/pdf/kaitei_kanji_toushin.pdf)
- 8 KIC 冒頭の解説 p.[9]。
- 9 グラフ横軸「レッスン範囲」における各階級は、それぞれ「表記の数マイナス 9 から表記の数まで」の範囲を示している。たとえば、階級 10 の棒グラフはレッスン第 1 回から第 10 回までの範囲である。ただし、階級 156 は第 151 回から第 156 回 (最終回) までの範囲を示す。
- 10 天野・近藤 (1999) の調査結果に含まれるエントリーとは表記が完全に一致しない KIC 単語は、検討対象から除外した。ただし、名詞的形容詞 (例：「十分な」)、連体詞 (例：「唯一の」)、そして一部の副詞相当句 (例：「気軽に」) は、親密度エントリーには句末の「な」「に」「の」が含まれない形で収録されており (例：「滑稽」「唯一」「気軽」)、KIC に所収の表記とは完全には同じではないが、本稿では同じものと

して判断し、親密度データを採用した。結果、KICに含まれる全10,183語（重複を含む）のうち、8667語の親密度を参照することができた。KICで複数の読み方が示されている単語は、読み方ごとに別々の語としてカウントした。天野・近藤(1999)において複数のアクセントパターンが存在する語については、WebKICに収録されている音声と同じアクセントの親密度を採用した。

- 11 天野・近藤(1999)の単語親密度調査には「水浸しになる」というエントリーはなく、本稿では「水浸」という表記の親密度を代用した。

### 資料

本稿のために集計したデータ（Microsoft Excel形式）は、[このリンクからダウンロードすることができる](#)。ただし、図2に関する詳細なデータ（国籍等、利用者個々人のデータ）は開示しない。

### 参考文献

- 天野成昭・近藤公久編著、NTTコミュニケーション科学基礎研究所監修(1999)『日本語の語彙特性 第1巻 単語親密度』三省堂
- アメリカ・カナダ大学連合日本研究センター(2013)『KANJI IN CONTEXT 中・上級学習者のための漢字と語彙 [改定新版]』the Japan Times
- 秋澤委太郎(2015)「『Kanji in Context』をWebから使う—長期的な使用に耐える多目的教育リソースの構築とその活用例—」『日本研究センター教育研究年報』(4) アメリカ・カナダ大学連合日本研究センター pp.38-58
- 秋澤委太郎(2016)「WebKICの改良とそれに対する学生の反応」『日本研究センター教育研究年報』(5) アメリカ・カナダ大学連合日本研究センター pp.35-58
- 秋澤委太郎(2018)「WebKICへのアカウント機能と音声再生機能の追加—機能の詳細と実装上の困難、そして学生による活用状況—」『日本研究センター教育研究年報』(7) アメリカ・カナダ大学連合日本研究センター pp.81-102