

混合効果モデルを用いた単語認知研究(Kida, 2016)の再分析による頑健性の検討

鬼田 崇作

1. はじめに

先行研究の再現可能性 (replicability) は、2015年にOpen Science Collaboration (2015) による大規模なプロジェクトによって、心理学では大きな関心を集めた。このプロジェクトでは、認知心理学および社会心理学の国際誌で発表されている論文100編に対して追試 (replication) が試みられた。その結果、オリジナルの研究結果が再現されたのは、全体の4割に満たなかった。この結果は、「その程度の研究結果しか再現されなかった」というように、心理学の世界ではおおむね否定的に捉えられたようである。その後は、再現可能性の低さの原因について様々な議論が行われ、研究内容の事前登録を行うことや、データとその分析のためのコードを公開し、社会と共有するなどの運動などが起こり、研究結果の再現性を高めるための様々な対策が取られている。

現在では、一般に再現可能性と呼ばれる概念について、いくつかの分類がなされているようである。例えば、National Academies of Sciences (2019) は、再生性 (reproducibility)、頑健性 (robustness)、再現性 (replicability) を区別している。Nosek et al. (2021, p. 4) では、それぞれを “testing the reliability of a prior finding using the same data and same analysis strategy” (再生性)、“testing the reliability of a prior finding using the same data and different analysis strategy”(頑健性)、“testing the reliability of a prior finding with different data”(再現性) と説明している。これらのうち、一般的に、追試による先行研究の結果の再現

可能性を問うことは、上記の3つのうちの再現性に相当するものだと考えられるが、その他の再生性または頑健性を問うことも、重要な取り組みであろう。

また、関連するその他の話題として、研究方法、データ分析方法の高度化を挙げることができる。先行研究の再現可能性を上記の3つのタイプで分類する場合、例えば頑健性については、同じデータに対して異なる分析を行うこととなる。近年、計算機の発達によって、数十年前にはできなかったような高度な統計的分析が実行できる環境が整ってきている (e.g., Bates et al., 2015; R Development Core Team, 2013)。その1つの例が、伝統的に t 検定や分散分析を用いてきた心理言語学実験における混合効果モデルの利用である。語彙性判断課題など心理言語学で用いられる課題においては、通常、1つの条件の中で複数の項目へ反応したデータが得られる。例えば、日本でカタカナ語として定着している英単語とカタカナ語としては定着していない英単語の間で、語彙性判断の早さや正確さに違いがあるか否かを検証する実験を考えた場合、カタカナ英語条件と非カタカナ英語条件のそれぞれにおいて、使用される項目は1つではなく、複数の項目が用いられ、同じ被験者がカタカナ英語条件と非カタカナ英語条件の双方について語彙性判断を行う反復測定の実験デザインとなることが多い。このような場合、通常、条件毎に複数の項目から得られた反応時間を平均 (aggregate) した値を、各人のカタカナ英語条件と非カタカナ英語条件の反応時間データとするであろう。データの分析においては、このようにして得られた複数人からの反応時間データについて、 t 検定や分散分析がなされ、条件間で統計的に有意な平均値の差がみられるか否かが調査される。さらに、被験者を誤差項とする被験者分析とともに、項目を誤差項とする項目分析がなされることも多い。

以上のような伝統的なデータ分析は、様々な点で批判がなされている (e.g., 新井・Roland, 2016; 井関, 2020)。例えば、語彙性判断課題など反応時間を得る典型的な課題では、全体として反応時間の短い被験者、長い被験者、実験

条件間の差が小さい被験者、大きい被験者など、反応時間の分散に個人差がある。しかし、1つの条件下で得られた複数の項目に対する反応時間データを平均して1つの値としてデータの分析を行うことは、被験者ごと、項目ごとの反応時間の違いである分散を無視していることになる。また、被験者分析では有意な差が得られたが、項目分析では有意な差が得られなかったという場合、結果の解釈が困難となる。このような結果は、項目数の少ない実験においてしばしばみられる (e.g., Jiang & Nekrasova, 2007; Kida, 2016)。

近年、上記のような問題への解決策を提供する方策として、混合効果モデルによる分析が広く利用されるようになってきた。日本語で書かれた詳細な文献もあることから (e.g., 新井・Roland, 2016; 井関, 2020)、詳細はそちらへ譲るが、混合効果モデルを用いれば、反復測定を行う個々の項目の反応時間を用いた分析ができる上に、被験者分析と項目分析を合わせた統計モデルを構築できる。そのため、被験者分析と項目分析で異なる結果が得られ、解釈が困難となるような状況は生まれない。また、欠損値を柔軟に扱うことができること、分析者が比較的自由に統計モデルを構築できることなど、伝統的な t 検定や分散分析よりも利点が多くある。

第二言語習得や応用言語学研究も含めた広い意味での英語教育学の研究では、以上のような近年の研究動向をどのように捉えているだろうか。心理学における再現可能性の議論以降、英語教育学の研究領域においては、活発な議論がなされてきたとはいえない状況にある。例えば、柳瀬 (2017) が示すように、英語教育学という応用的学問においては、科学性を志向する客観性や再現性という概念自体が相性が悪いという議論も可能であろう。しかし、再生性、頑健性、再現性などの観点から、先行研究から得られた知見を検証することは、科学的研究における本質的な営みの1つであろう。そこで、本稿では、研究結果の再現可能性とデータ分析手法の高度化という近年の話題を鑑み、旧来のデータ分析手法を用いて既に発表した研究のデータについて、混合効果モデルを用いた再分析を行うことにより、元の研究結果と同様の結

果が得られるか否か、つまり頑健性が認められるか否かに取り組む結果を報告する。

2. 先行研究の概観

第一言語、第二言語を問わず、読解においては、視覚的に提示された単語の適切な認知が不可欠であるため、単語認知過程を解明することは、言語処理研究の中心的な研究課題の1つである (Davis, 2010)。単語認知とは、視覚的または聴覚的に提示された刺激が、心的辞書 (mental lexicon) と呼ばれる認知システムの中に貯蔵されている情報と一致する刺激であるとの決定がなされるまでの、一連の情報処理過程を指す。第一言語の単語認知に関するこれまでの研究では、この単語認知過程について様々な認知モデルが提案されており (e.g., Davis, 2010; McClelland & Rumelhart, 1981; Morton, 1970; Whitney, 2001)、現在も活発な議論がなされている。

単語認知モデルの研究は、成人がすでに心的辞書内に貯蔵している語がどのように認知されるのかを示すものである。他方、言語習得の観点からは、その認知メカニズムがどのように獲得されるのかが興味の対象となる。伝統的には、単語認知のメカニズムがどのように発達し、獲得されるのかは主要な研究課題ではなかったが (Castles & Nation, 2006, 2008; Sibley et al., 2008)、近年、学習された語が心的辞書内でどのように表象されるのかについての研究が行われるようになってきている (Qiao & Forster, 2013)。例えば、第一言語における代表的な単語認知の仮説として、Perfetti (1992) による Lexical Quality 仮説を挙げる事ができる (図1)。

LEVEL 1	LEVEL 2	LEVEL 3
ir*n	iron	iron
t*g*	t*ng**	tongue
uk*	ukil*	uk*I*I*

図1. *iron*, *tongue*, *ukulele*における表象の正確さの発達過程 (Perfetti, 1992, p. 158)

図1は、*iron*, *tongue*, *ukulele*の3単語について、それぞれの表象の発達過程を示している。いずれの単語においても、習得の初期段階 (LEVEL 1) では、完全な語彙表象は獲得されておらず、形式が部分的に表象されている。他方、発達が進めば (LEVEL 2), *iron*は完全な表象が獲得される。他方、その他の2語については、LEVEL 1のときよりも正確さは増しているものの、まだ完全な表象の獲得には至っていない。発達がさらに進めば (LEVEL 3), *iron*に加え、*tongue*の表象が完全に獲得されている。しかし、*ukulele*の場合は、発達がLEVEL 2からLEVEL 3への過程で徐々に表象の正確性は高まっているものの、最後まで完全な習得へは至っていないことが示されている。このように、Lexical Quality仮説に従えば、語彙表象の発達においては、(a) 語彙表象はall or nothingではなく、部分的に表象される発達段階を経て完全な表象へ至ること、(b) 発達の程度は単語ごとに異なること、の少なくとも2点が示唆される。

Perfetti (1992) は、図1に示す*iron*や*tongue*のように、完全な語彙表象へ至ることのメリットの1つとして、その段階に至った単語は、形態的に類似する他の語との読み間違い、つまり単語認知のエラーが起こらないことを挙げている。英語は限られた数のアルファベットを組み合わせて単語を作成する言語であるため、形態的隣接語 (orthographic neighbor) と呼ばれる、綴りの似た単語が複数存在する言語である。正確な語彙表象が獲得できていない場合、ある単語の視覚的認知において、その語自身の表象と形態的隣接語の表象の両方が活性化されてしまい、隣接語と読み間違えるエラーが起こる。他方、正

確な語彙表象が獲得されていれば、例え形態的に類似する隣接語が存在してたととしても、隣接語の表象は活性化されず、正確な単語認知が可能となる。第二言語読解の先行研究では、学習者は形態的に類似する単語同士を読み間違えるエラーを起こすことが多く報告されている (Bensoussan & Laufer, 1984; Laufer, 1988)。Lexical Quality仮説に従えば、その理由は第二言語学習者がこれらの単語の表象の正確さを十分に発達させていないことであると考えられる。

学習者が正確な語彙表象を獲得できているかどうかは、どのように調査できるのであろうか。現在までの単語認知研究では、マスク下の形態的プライミング法 (masked form-priming) という実験手法が主に用いられている。プライミングとは、PCモニターに先に提示される刺激 (プライム) の処理が後に提示される刺激 (ターゲット) の処理に与える効果を指す。例えば、*CONTRAST*というターゲットに対して、*bontrast*というターゲットと形態的に類似するプライムを示す条件では、形態的に類似しないプライムを示す統制条件よりも、語彙性判断課題の反応時間が早くなる (Forster & Davis, 1984)。語彙性判断課題とは、PCモニターに提示される文字列が実在の単語が単語に似せた疑似単語かを早く正確に判断する課題であり、視覚的単語認知研究において使用される中心的な認知課題の1つである。

通常、この実験手法では、言語刺激を隠すためのマスクとして、#####が用いられる。次に、プライムが小文字で極めて短い時間 (50ミリ秒ほど) 提示される。最後に、ターゲットが大文字で提示される。被験者には、ターゲットについての語彙判断や命名などの課題が課せられる。プライムの提示時間は非常に短く、また、その前後でマスクとターゲットが連続して提示される。通常、被験者は、プライムが提示されたことに気づかないため、この実験手法では、被験者の意識的な処理 (strategic processing) を排除できるとされる (Forster et al., 2003)。通常、マスク下の形態的プライミング法による実験では、ターゲットとプライムの形態的類似度を操作し、類似度の異なるプライム条

件間でターゲットの語彙性判断や命名などの反応時間とその正確さがどのように異なるのかが調査され、その結果を基に単語認知モデルの検証や単語認知プロセスの提案がなされる。

Perfetti (1992) によるLexical Quality仮説に従えば、ターゲットの語彙表象が十分に発達していなければ、プライムとターゲットの間で形態的類似性のある実験条件では、プライムの提示によって、ターゲットの表象が活性化されてしまい、プライムとターゲットの間で形態的類似性が無い統制条件と比べ、語彙性判断課題の反応時間が異なるであろう。他方、ターゲットの語彙表象が十分に発達していれば、プライムの提示はターゲットの語彙表象を活性化させないため、実験条件と統制条件間で語彙性判断課題の反応時間に違いは見られないだろう。

以上の内容を実証的に示した研究として、Castles et al. (2007) を挙げる事ができる。この研究では、マスク下の形態的プライミング法による語彙判断課題を用いて (図2)、英語を母語とする子供がGrade 3とGrade 5に在籍時に単語認知実験を行うことで彼らの単語認知システムの発達過程を調査し、その結果を大人の単語認知と比較した。彼らの研究では、高頻度語のターゲット (e.g., *play*) に対し、非単語のsubstitution neighbor (SN), transposition neighbor (TN), controlという3種類のプライムが使用された。SNとは、ターゲットと1文字異なる文字列であり (e.g., *rlay*)、TNとは、ターゲットとは一部の隣り合う二文字の位置が異なる文字列である (e.g., *lpay*)。また、control条件のプライムはターゲットと同じ文字を含まない文字列である (e.g., *meit*)。SNとTNはどちらもターゲットの形態的隣接語であるが、これまでの研究からは、おおむねSNよりもTNのほうがよりターゲットへの形態的類似性が高いとされる (Chambers, 1979; Forster et al., 1987; Perea & Lupker, 2003a, 2003b; Perea et al., 2005)。実験の結果、Grade 3では、SN, TN条件ともに、control条件よりも反応時間が早かった。この結果は、Perfetti (1992) のLexical Quality仮説に従えば、Grade 3の児童の語彙表象の不正確さを示す結果であ

ると解釈される。2年後、Grade 5時点では、SN条件の反応時間はcontrol条件の反応時間と同程度であったが、TN条件の反応時間はcontrol条件の反応時間よりも早いままであった。この結果は、Grade 3のときよりも語彙表象の正確さが高まったため、ターゲットと形態的類似度の比較的低いSN条件のプライムではターゲットの表象が活性化されず、語彙性判断課題におけるターゲットの反応時間が早まらなかった一方、ターゲットとの形態的類似性が高いTN条件では、プライムの提示によりターゲットの表象が活性化され、それによりターゲットの語彙性判断が早くなったと解釈される。最後に、18歳以上の大人に対して同様の実験を行った結果、SN条件、TN条件ともに、語彙性判断の反応時間はcontrol条件の反応時間と変わらなかった。つまり、PerfettiによるLexical Quality仮説に従えば、ターゲットの語彙表象を十分に発達させている状態だと解釈される。以上の結果から、単語認知の発達過程として、SN条件、TN条件ともにcontrol条件よりも語彙性判断の反応時間が早くなる発達の初期段階（Grade 3の状態）、TN条件でのみcontrol条件よりも反応時間が早くなる発達の中間段階（Grade 5の状態）、SN条件、TN条件ともに反応時間が早くならない発達の最終段階（大人の状態）の3段階が想定できる。

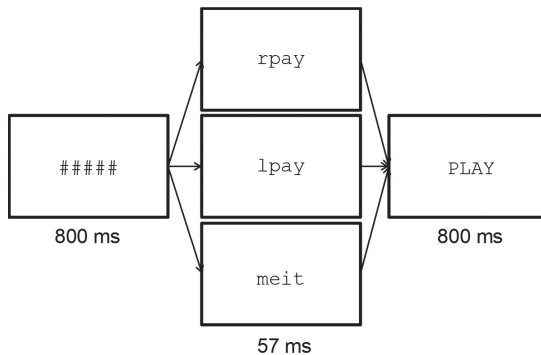


図2. Castles et al. (2007) におけるマスク下の形態的プライミング法による手続き

このCastles et al. (2007) による単語認知の発達過程を基に、日本人英語学習者を対象とした研究としてはKida (2016) がある。この研究では、日本人英語学習者の大学生を対象として、(a) 彼らの単語認知の発達段階は、Castles et al.の3段階からなる発達過程のどの段階に位置づけられるのか、(b) 教育的介入によって、その発達段階を変えられるのか、の2点を調査した。¹⁾ 実験は、必修科目である英語のリーディング授業の一環として行われた、12週間の授業内多読活動に参加した大学一年生21名が対象とされた。多読活動は、週に1回、授業内の30分ほどで行われた。参加者はPenguin ReadersのGraded Readersの中から読みたい本を自由に選ぶことができた。多読活動は、(a) 可能な限り多くの本を読むこと、(b) 内容について興味を持てる本を選ぶこと、(c) 楽しんで本を読むこと、(d) 面白くなければすぐに別の本を選ぶこと、(e) 辞書を頻繁に引かなくても読める程度の難易度の書籍を選ぶこと、という多読の基本的ルールに則って行われた。12週間の多読活動を行う前後に、Castles et al. (2007) と同様のマスク下の形態的プライミング法による語彙性判断課題を行った (pretestおよびposttest)。実験の結果、多読活動の前後ともに、SN条件、TN条件の反応時間はcontrol条件の反応時間よりも早く、この結果は、Castles et al.の実験におけるGrade 3の児童の結果と同様であった。このことから、(a) 日本人英語学習者の大学生の単語認知の発達段階は、初期段階に留まること、(b) 12週間の授業内多読活動では、単語認知の発達段階に大きな影響は与えられないこと、の2点が示された。

Kida (2016) の分析は、伝統的な分散分析によるものであった。以下、図3と図4にKida (2016) における事前テストと事後テストの反応時間分布を示す。これらの図が示すとおり、各条件の反応時間には、全体として反応時間が短い被験者や長い被験者、プライム条件間の差が小さい被験者や大きい被験者など、分散に個人差があることがわかるが、この個人差は分散分析では無視されている。また、Kidaの分析では、被験者分析と項目分析を別々に行った結果、プライムの主効果について、被験者分析では有意な結果が示された

一方、項目分析では非有意な結果が示され、解釈に困難が生じている。これらの問題点を克服するため、本稿では、混合効果モデルを用いてKidaのデータを再分析することにより、分析結果の頑健性について検討を行う。

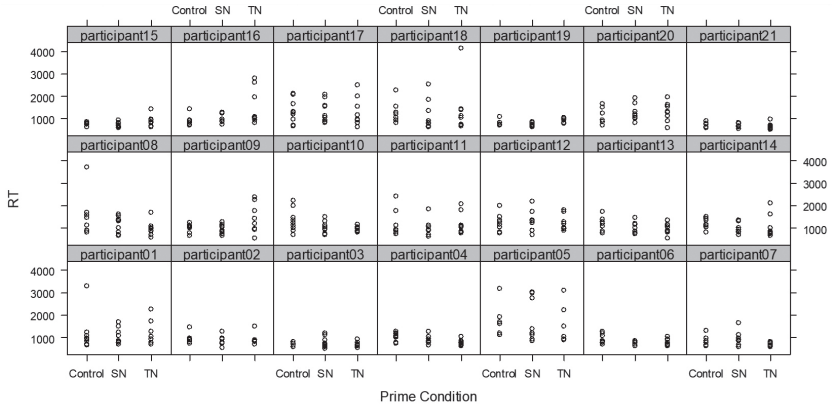


図3. pretestにおける各プライム条件の反応時間分布

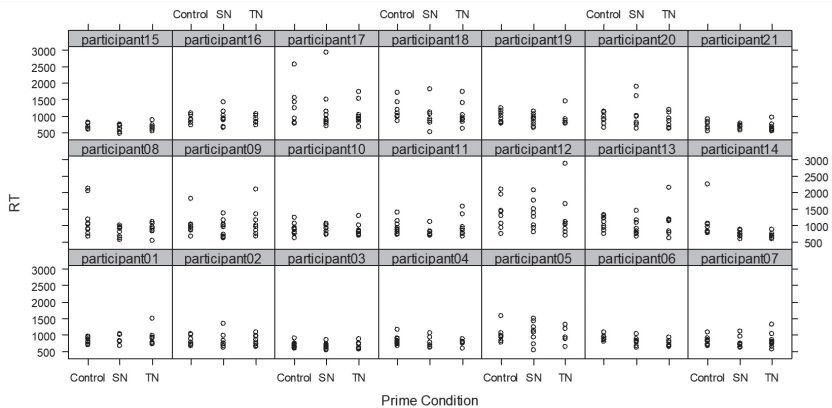


図4. posttestにおける各プライム条件の反応時間分布

3. 本研究

本研究では、Kida (2016) における正反応の反応時間データおよび誤答率に対し、混合効果モデルをフィットさせる。²⁾ 分析は統計解析環境のR (R Development Core Team, 2013) を用いて、混合効果モデルによる推定はlme4パッケージ (Bates et al., 2015), 固定効果のp値の推定にはlmerTestパッケージ (Kuznetsova et al., 2014) を用いた。分析のためのコーディングは、プライム要因については、controlを参照レベルとするダミーコーディング、テスト要因については、pretestを-0.5, posttestを0.5とする対照コーディングとした。また、反応時間データについては、一般的に生データでは正規分布のような左右対称の分布にはならず、正に歪む分布となるため (Balota & Yap, 2011; Ratcliff, 1993; Wagenmakers & Brown, 2007), 本研究では-1,000を各反応で割る逆変換を施したデータを分析対象とする (Kezilas et al., 2017; Nakayama et al., 2016)。

データにフィットさせるモデルは、固定効果として、プライム (SN, TN, control) およびテスト (pretest, posttest), 両者の交互作用を指定し、被験者の変量切片および変量傾きとして、プライムとテストの交互作用、項目の変量切片および変量傾きとして、テストを指定した。しかし、このモデルは反応時間および誤答率のどちらのデータにおいても収束しなかったため、変量効果の構造を順次単純化していき、初めに収束したモデルを最終的なモデルとした。反応時間データについての最終モデルの変量効果は、被験者の変量切片および変量傾きとして、プライムとテストを指定し、項目の変量切片および変量傾きとしてテストを指定した。また、誤答率データについての最終モデルは、被験者の変量切片、項目の変量切片および変量傾きとしてテストを指定した。別途、VIF (Variance Inflation Factor) を計算したところ、最大の値を示したものは両モデルともテスト要因であり、その値は反応時間のモデルでは2.26, 誤答率のモデルでは2.57であり、どちらのモデルにおいても

多重共線性は認められないものと判断した。

4. 結果と考察

反応時間データについての結果を表1に示す。分析の結果、プライムとテストの交互作用は有意ではなかった。また、SN条件とcontrol条件の比較およびTN条件とcontrol条件の比較においては、どちらにおいても有意な差が見られた。以上の結果は、Kida (2016) と同様であり、混合効果モデルを用いた本研究の再分析においても、結果の頑健性が示された。

表1. 線形混合効果モデルを用いたKida (2016) の反応時間データの再分析結果

	固定効果				変量効果	
	Estimate	SE	t	p	被験者	項目
					SD	SD
Intercept	-1.08	0.04	-25.34	< .05	0.17	0.04
Substitution vs Control	-0.08	0.03	-2.72	.01	0.05	—
Transposition vs Control	-0.08	0.03	-2.34	.03	0.08	—
Test	-0.09	0.03	-2.70	.01	0.08	0.01
Substitution x Test	-0.04	0.04	-1.04	.30	—	—
Transposition x Test	-0.03	0.04	-0.79	.43	—	—

また、誤答率データについての結果を表2に示す。分析の結果、プライムとテストの交互作用は有意ではなかった。また、SN条件とcontrol条件の比較およびTN条件とcontrol条件の比較においても、有意な差は見られなかった。反応時間の結果と同様に、以上の誤答率の結果も、Kida (2016) と同様であり、結果の頑健性が示された。

表2. 一般化線形混合効果モデルを用いたKida (2016) の誤答率データの再分析結果

	固定効果				変量効果	
	Estimate	SE	<i>z</i>	<i>p</i>	被験者 <i>SD</i>	項目 <i>SD</i>
Intercept	-2.66	0.26	-10.26	< .05	0.55	0.27
Substitution vs Control	-0.46	0.33	-1.39	.17	–	–
Transposition vs Control	0.01	0.30	0.03	.98	–	–
Test	-0.36	0.40	-0.91	.36	–	0.11
Substitution x Test	0.73	0.61	1.19	.24	–	–
Transposition x Test	0.62	0.55	1.12	.26	–	–

5. おわりに

本稿では、日本人英語学習者を対象に授業内で多読活動を行うことにより、語彙表象の発達が見られるか否かを調査したKida (2016) のデータの再分析を行うことにより、結果の頑健性の検証を行った。伝統的な分散分析ではなく、近年その利用がより広がっている混合効果モデルによる分析を行った結果、元の研究結果と同様の結果が得られ、頑健性が示された。

再生性、頑健性、再現性という3つの観点からの検証を行うことは、先行研究で得られた知見を再検討する有効な方法として利用できるであろう。今後、過去の実証的な英語教育学研究の知見が再検討され、我々の知見がさらに蓄積されることを期待したい。

註

- 1) 実際には、Kida (2016) の研究では、Castles et al. (2007) の追実験による語彙表象の発達に加え、変動係数 (coefficient of variation) を用いた単語認知の自動化 (automatization) の調査が行われている、本稿では前者のデータ分析のみを対象にデータの再分析を行う。
- 2) 本稿では、Kida (2016) の分析の再検討という観点から、便宜上、反応時間のデータと誤答率のデータに対して、2つの異なる混合効果モデルをフィットさせて分析を行った。しかし、語彙性判断課題などの強制的二択課題における反応時間と誤答率の分析においては、拡散モデル (diffusion model) という両者を同時にモデリングする数理モデル (Ratcliff & Rouder, 1998) が提案されており、今後はこのようなモデリングによる研究を行うことが望ましいだろう。

参考文献

- Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, 20, 160–166. doi:10.1177/0963721411408885
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01
- Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, 7, 15–32. doi:10.1111/j.1467-9817.1984.tb00252.x
- Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology*, 97, 165–182. doi:10.1016/j.jecp.2007.01.006
- Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 151–179). Psychology Press.
- Castles, A., & Nation, K. (2008). Learning to be a good orthographic reader. *Journal of Research in Reading*, 31, 1–7. doi:10.1111/j.1467-9817.2007.00367.x
- Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior*, 18, 225–241. doi:10.1016/S0022-5371(79)90136-1
- Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological*

- Review*, 117, 713–758. doi:10.1037/a0019738
- Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 680–698. doi:10.1037/0278-7393.10.4.680
- Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *Quarterly Journal of Experimental Psychology*, 39A, 211–251. doi:10.1080/14640748708401785
- Forster, K. I., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: State of the art* (pp. 3–37). New York: Psychology Press.
- Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91, 433–445. doi:10.1111/j.1540-4781.2007.00589.x
- Kezilas, Y., McKague, M., Kohlen, S., Badcock, N. A., & Castles, A. (2017). Disentangling the developmental trajectories of letter position and letter identity coding using masked priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 250–258. doi:10.1037/xlm0000293
- Kida, S. (2016). Automatization and orthographic development in second language visual word recognition. *Reading in a Foreign Language*, 28, 43–62.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer Objects of lme4 Package). R Package Version 2.0-10. Available at: <http://CRAN.R-project.org/package=lmerTest>
- Laufer, B. (1988). The concept of ‘synforms’ (similar lexical forms) in vocabulary acquisition. *Language and Education*, 2, 113–132. doi:10.1080/09500788809541228
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88, 375–407. doi:10.1037/0033-295X.88.5.375
- Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.
- Nakayama, M., Kinoshita, S., & Verdonschot, R. G. (2016). The emergence of a phonemized unit in L2 speech production: Evidence from Japanese–English bilinguals. *Frontiers in Psychology*, 7, 175. doi:10.3389/fpsyg.2016.00175
- National Academies of Sciences, E. (2019). Reproducibility and replicability in science. <https://doi.org/10.17226/25303>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler,

- F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2021, February 9). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*. <https://doi.org/10.31234/osf.io/ksfvq>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716, doi:10.1126/science.aac4716
- Perea, M., & Lupker, S. J. (2003a). Does judge activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory and Cognition*, *31*, 829–841. doi:10.3758/BF03196438
- Perea, M., & Lupker, S. J. (2003b). Transposed-letter confusability effects in masked form priming. In S. Kinoshita & J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 97–120). Psychology Press.
- Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception and Psychophysics*, *67*, 301–314. doi: 10.3758/BF03206493
- Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Erlbaum.
- Qiao, X., & Forster, K. I. (2013). Novel word lexicalization and the prime lexicality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1064–1074. doi:10.1037/a0030528
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–532. doi:10.1037/0033-2909.114.3.510
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356. doi:10.1111/1467-9280.00067
- R Development Core Team (2013). 'R: A language and environment for statistical computing [Computer software],' Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Sibley, D. E., Kello, C. T., Plaut, D. C., & Elman, J. L. (2008). Large-scale modeling of wordform learning and representation. *Cognitive Science*, *32*, 741–754. doi:10.1080/03640210802066964
- Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*, 830–841. doi:10.1037/0033-295X.114.3.830
- Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SEIOL model and selective literature review. *Psychonomic Bulletin and Review*, *8*, 221–243. doi:10.3758/BF03196158

新井学・Douglas Roland (2016)「言語理解研究における眼球運動データ及び読み時間データの統計分析」『統計数理』第64巻第2号201-231.

井関龍太 (2020)「心理学者は反応時間をどう分析するか」『基礎心理学研究』第38巻第2号243-249.

柳瀬陽介 (2017)「英語教育実践支援研究に客観性と再現性を求めることについて」『中国地区英語教育学会研究紀要』47巻83-93.

Synopsis

Testing the Robustness of a Word Recognition Study (Kida, 2016) with Mixed Effect Models

Shusaku Kida

Recent discussion on the replication crisis in psychological research has classified three aspects of approach for this issue: reproducibility, robustness, and replicability. Among them, robustness refers to the test of previous findings using the same data with new analyses. The purpose of this paper is to show the robustness of Kida's (2016) masked form priming lexical decision data with mixed effect models. Kida examined the effects of in-class extensive reading of graded readers on the acquisition of precise word recognition in English by university-level Japanese learners of English. The rationale of using the form priming lexical decision was that when the participants did not acquire precise representations of target words in the lexical decision task, the presentation of orthographically related nonword primes activated target words' representation, which lead to faster reaction time in the task. On the other hand, when they acquired precise representations, these orthographically related nonword primes did not affect the processing of the target; therefore no facilitatory reaction time should be observed. The results demonstrated that orthographically related primes fastened the processing of targets, suggesting that the participants' representations of targets were not precise. The results were telling, but Kida used analyses of variance (ANOVA) for the complex data, which is recently regarded as questionable practice. Instead, researchers are now encouraged

to use mixed effect models to analyze such complex data. This paper, therefore, reanalyzed the Kida's (2016) data with the linear mixed effect model for the RT data and the logistic mixed effect model for the error rates and examined the robustness of his findings. The mixed effect models showed almost the same results as the original study by Kida, which shows the robustness of the original finding reported in Kida.