

反応時間計測実験における外れ値の取扱い —L2 心理言語実験の場合—*

橋本 健一
近畿大学

キーワード: 心理言語実験、反応時間、外れ値、L2 処理、自己ペース読み課題

1. はじめに

心理言語学は、人間の言語理解・産出・習得に関わる心的処理はどのようなものであるか、という課題に答えることを目的とする研究領域である。この研究課題に答えるために、様々なレベル（音韻・形態・統語・意味など）で操作された言語刺激を素材として実験が行われているが、その中で広く用いられている手法の1つが、反応時間を計測する行動実験である¹。多くの場合、その前提は、ある条件における反応時間が、別の条件における反応時間よりも長かった場合、前者の方でより複雑な言語処理が行われている、ということである。従来、このような心理言語実験は、第一言語（L1）話者を対象として幅広く行われていて、人間の母語使用・習得に関して、多くの知見が得られている。しかし近年、言語に関わる心理プロセスを明らかにしようとする動きは、第二言語（L2）話者を対象とするものにまで広がってきていて、L2 話者を対象にした心理言語実験も多く行われている。これは、L1 のように流暢に使いこなせるわけではないが、何らかの言語処理システムを使って言語を使用・習得するという点については L2 も同様であり、不完全である分、熟達度に応じた言語発達（あるいはそれに関与する人の認知システム）について、より詳細なデータを与えてくれるのでは、という考え方に基づいている（門田, 2010）。

このように、反応時間を計測する心理言語実験は、L1・L2 話者を対象として幅広く行われている。実験の実施方法、データ分析法については、共通の理解が得られている部分が多いが、研究者によって異なる、という部分も少なからず見受けられる。この内の1つが、反応時間を分析する際の外れ値（outlier）の取り扱いについてである。外れ値とは、実験機器の操作ミス、集中力の低下、外部からのノイズなどによって生じる、極端に短い、あるいは長い反応時間のことである。外れ値をそのまま放置したデータは、分析・解釈に大きな影響を与える恐れがあり、タイプ1・タイプ2のエラーを起こす原因となる可能性がある。そのため、研究者は何らかの方法で外れ値の影響を低減する処置を行うが、その方法は一貫しておらず、どのような時にどの方法を使うのが良いか、ということを提案するガイドラインも少ない。

本稿では、筆者が実施した、日本人英語学習者が英文を読む際の統語解析を調べる実験を例にして、外れ値の取り扱いについて考察する。データ中に含まれる外れ値（と思われ

る値) はどのようなものかを提示して、実際に外れ値の影響を低減する方法をいくつか紹介する。また、外れ値を含んだままのデータと、外れ値の影響を小さくするよう処理されたデータで、どのように記述・推測統計の結果が異なるかを検討する。

2. 心理言語実験の方法

2.1 自己ペース読み課題

L1・L2を問わず、どのような言語処理を観察・研究したいかによって、様々な心理言語実験の方法が存在する。例えば、メンタルレキシコン内の語彙知識へのアクセススピードなどを測る際には、語彙性判断課題 (lexical decision task) やプライミング課題 (priming task) などが行われる。近年の脳研究の隆盛を背景として、脳波の測定や fMRI (functional magnetic resonance imaging : 機能的磁気共鳴画像法) による脳血流量の変化の記録など、言語使用・習得の脳内基盤を明らかにしようとする動きもみられる (心理言語実験の詳細については門田(2010)の Step 4 を参照)。

筆者は L2、特に日本人英語学習者の英文理解の際の統語処理についての研究を主に行っている。この分野においても、色々な実験方法が考えられるが、簡便で様々な示唆を与えてくれる方法として、移動窓方式の自己ペース読み課題 (self-paced reading task with moving window paradigm) がある。これは、下記図 1 のように、コンピューター画面上に刺激文が単語ごと (あるいは句・チャンクごと) に表示され、実験参加者が指定されたボタンを押しながら、英文を読み進めていくという課題である。

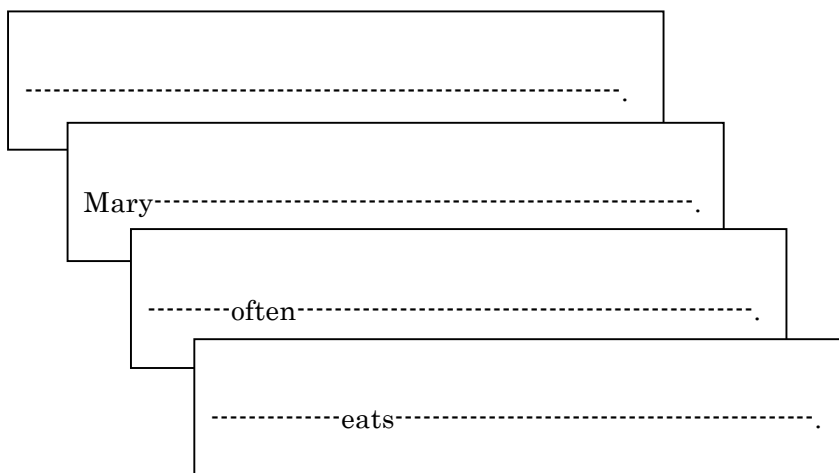


図 1. 自己ペース読み課題の刺激文提示例

最初、コンピューター画面上には、文の長さを表すダッシュが表示されている。実験参加者が指定されたキーを押すごとに、刺激文の始めから 1 語ずつ表示され、直前に表示されていた単語はダッシュに戻る。単語が表示されてから、次にボタンを押すまでの時間が、

その単語の読解時間としてミリ秒単位で記録される。この手続きが、刺激文の最後まで続き、多くの場合、読んだ文を正しく理解できているかを試す、理解度チェック問題が続いて表示される。

2.2 L2 英語関係節文処理研究の概要

以下に、自己ペース読み課題でどのようなデータが得られて、どのように分析するのかを示すが、それに先立って、筆者が実施した、英語関係節文の読解実験について、少し説明をする (Hashimoto, in press)。この実験は、熟達度の高い日本人英語学習者が、(1, 2)のような英語関係節文を読む際にどのような統語処理をしているか、ということ明らかにすることを目的としている。実験参加者は、英語圏の大学に留学している日本人英語学習者 23 名であった。

(1a) 主格関係節・高頻度語文²

The boy who helped the girl passed the exam last week.

(1b) 目的格関係節・高頻度語文

The boy who the girl helped passed the exam last week.

(2a) 主格関係節・低頻度語文

The defendant who insulted the journalist confessed the sin at last.

(2b) 目的格関係節・低頻度語文

The defendant who the journalist insulted confessed the sin at last.

(1a, 2a)の関係節は、関係代名詞 who が関係節内の動詞の主語の役割を果たしており、主格関係節と呼ばれる。一方、(1b, 2b)の関係節は、関係代名詞 who が関係節内の動詞の目的語の役割を果たしており、目的格関係節と呼ばれる。L1 の心理言語学研究を中心に、多くの実験が行われて、主格関係節文(1a, 2a)の方が、目的格関係節文(1b, 2b)よりも処理が容易であるということが示されている。具体的には、自己ペース読み課題や眼球運動測定実験の中で、目的格関係節文の方が、主格関係節文よりも読解時間が長いという結果が報告されている (例えば Gibson, Desmet, Grodner, Watson, & Ko, 2005)。

一方、上記(1)と(2)の文は、文中の内容語の語彙頻度が異なっている。(1)は、JACET8000 の高頻度語上位 2,000 語を用いて刺激文が作られている。(2)では、同じく JACET8000 における頻度順位 2,001~4,000 位の語を用いて刺激文を作成している。単語の出現頻度は、その単語の認知処理に影響を与えるとされていて、高頻度語の方が処理が容易であるという実験結果は、L1・L2 の語彙処理研究で多く報告されている。ところが、このような語彙処理の難易度 (あるいは自動性) が、文の統語処理にどのような影響を及ぼすかという点については、まだ研究がなされていない。L2 学習者の語彙処理自動性が低いために、高頻度

語文と低頻度語文で、関係節を理解する際の統語処理が異なる可能性は十分ある。すなわち、高頻度語文では、語彙処理負荷が低く、統語処理の複雑さに応じた、構造間の読解時間の差が観察されるが、低頻度語文では、処理容量が語彙処理のみで使われてしまい、構造間に統語処理効果（すなわち読解時間の差）は生じないと予測される。

実験参加者は、各構造4文ずつ計16文と、他に関係節を含む文を20文、さらにフィラー文（実験意図を悟られないようにするための文）として関係節を含まない英文を54文加えた合計90文を、単語ごとの自己ペース読み課題で読んだ。

3. 読解時間データの特徴と外れ値

図2は、実験文・フィラー文に含まれるすべての単語について、ある1人の実験参加者の読解時間がどの程度であることを示したヒストグラムである。

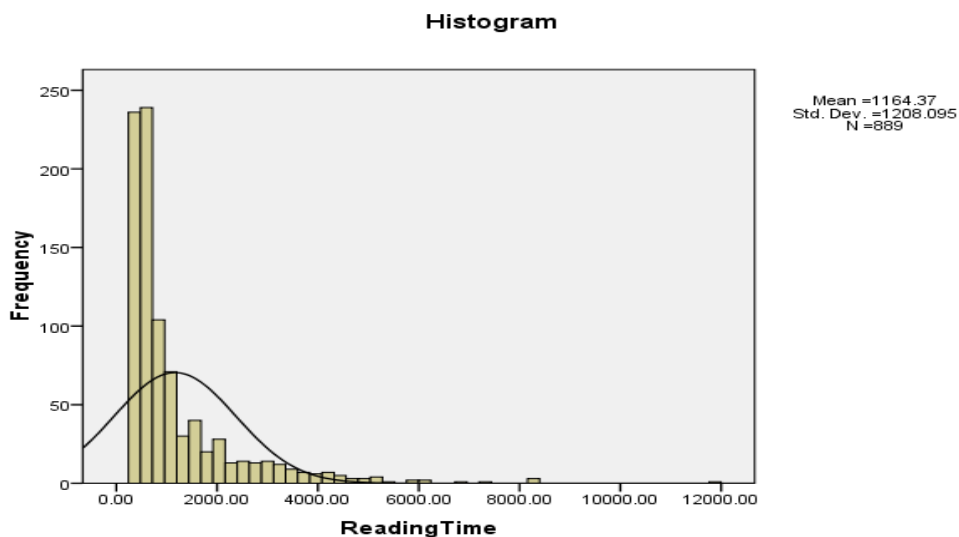


図2. ある実験参加者の単語毎の読解時間のヒストグラム

このグラフから、特に2つのことが注目される。1つは、多くの単語の読解時間は1,000ミリ秒以下で読まれており、グラフのピークが、左側に歪んでいることである。これは、正に歪んでいる分布（positively-skewed distribution）と呼ばれ、心理言語実験では、語彙性判断課題などでもよく見られる分布である。もう1つは、ヒストグラム上では見にくいだが、この実験参加者はほとんどの単語を1,000ミリ秒前後で読んでいるが、12,000ミリ秒付近で1つ、8,000ミリ秒付近でもいくつかの値が検出されている。これは、文を読み進める過程で、ある特定の単語を読むのに8,000~12,000ミリ秒かかっていることを示している。この実験参加者の通常の読みのスピードが1,000ミリ秒前後であると考えると、これらの値

は明らかに長すぎる。複雑な統語構造を含む文を読んでいるので、そのために表れた数値であるとも考えられるが、もう 1 つ考えられるのが、これらの値が外れ値 (outlier) である可能性である。

心理言語実験で言う外れ値とは、実験操作 (例えば主格・目的格の差) に関係のない、極端に大きい (あるいは小さい) 読解時間の値のことである。外れ値が生じる理由は多々ある。例えば、自己ペース読み課題では実験参加者に対して、「各文を理解できる程度で、なるべく速く読んでください」などと指示をする。この指示の伝え方が悪い (あるいは実験参加者が指示に従わない) と、とにかく速く読むことが大事だと思い、非常に短い読解時間のみが観測されることがある。しかしこの場合、各文の後に続く理解度チェック問題に対する正解率が極端に下がるので、正答率・読解時間のデータを照らし合わせることで、分析上不適格なデータとして特定することは、比較的容易である。

これに対して、図 2 で示されているように、ある実験参加者の標準的な読み時間からすると、著しく長い読解時間が計測されることがある。このような長い読解時間につながる要因は多岐にわたる。1 つは外部からのノイズによるもので、例えば実験環境の周辺で突然大声がしたり、非常用ベルなどが鳴ると、実験参加者の注意がそちらに向いてしまい、読み進める処理が一時的に止まるために、このような長い読解時間が生じる。最近では、携帯電話などの音で生じることもある。他の要因としては、実験参加者の集中力不足がある。実験中には、自分のペースで休憩が取れるように設定されていることが多いが、それでも文の途中で疲れて止まったり、空腹感、あるいは読んでいる文やタスクそのものに興味がない、などの影響でも、集中した読みが妨げられることがある。これ以外にも、多くの理由により、長い読解時間が観測されることがある。これらの値は、実験の操作には全く関係のないものである。そのため、このような外れ値の影響は極力取り除かれなければならない。

ところが、外れ値の特定・除去は極めて難しい。それは、多くの場合、外れ値と考えられる読解時間の値は、実際の処理を反映した読解時間の値と重なるためである。とりわけ、実験の操作 (例えば、目的格関係節文を読むとき) により、読解時間が長くなることは、実験仮説上も十分に考えられることである。事実上、外れ値だけを取り除くことは不可能に近いといえる。このため、外れ値の取り扱いとしては、正しい処理を反映したデータをできるだけ残して、外れ値の可能性のあるデータの影響を限りなく小さくする、ということに尽きるようである (Ratcliff, 1993)。

4. 外れ値の影響を低減する方法

それでは、外れ値の影響を低減する方法としては、どのようなものがあるのだろうか。Ratcliff (1993) では、以下のような方法を紹介している。

- ① 一定の基準値を設定して、それ以上の値をすべて取り除き、分析対象外とする方法。
この基準値は、2,000 ミリ秒、3,000 ミリ秒のように、実験参加者間の個人差を問わず絶対値を採用する方法と、個人・グループの平均値+2 あるいは3 標準偏差の値を基準値とする方法がある。
- ② 基準値を設定して、その基準値内に収まらない値は、すべてその基準値の値に変換する方法。基準値の求め方は、上記①と同様、絶対値を採用する、標準偏差値を基に設定する、等の方法がある。
- ③ 各読解時間に対して任意の数を底とする対数を求める対数変換（log transformation：例えば 1,000 ミリ秒の場合、10 を底とする対数を求めると $\text{Log}_{10}(1000) = 3$ など）、読解時間値の逆数を求める逆変換（inverse transformation： $1 / \text{読解時間}$ ）などのデータ変換を行う（Osborne, 2002）。これらの変換をかけると、正に歪んだ分布はより正規分布に近づき、外れ値である可能性がある長い読解時間の影響が小さくなる（下記図 3 参照）。

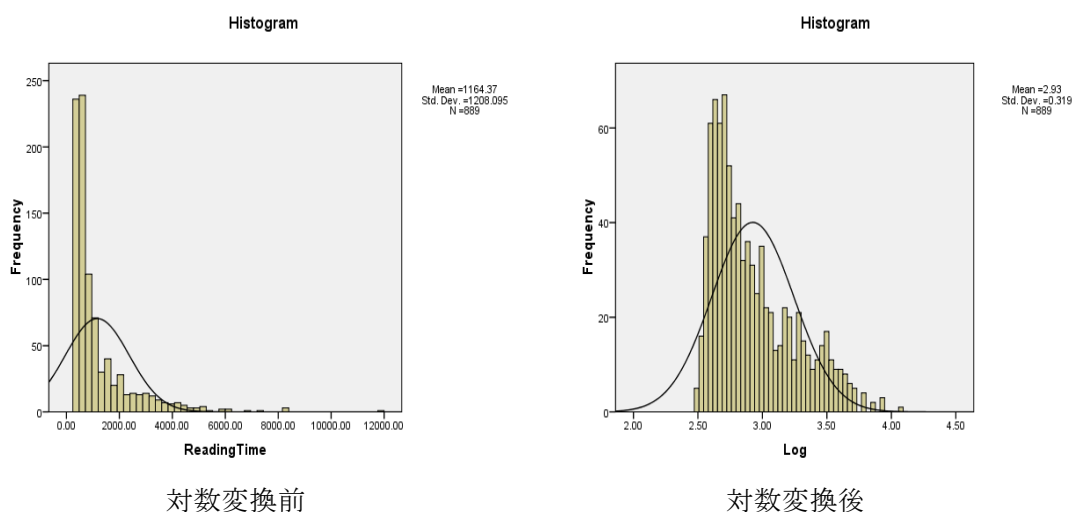


図 3. ある実験参加者の単語毎の読解時間のヒストグラム（対数変換前後）

5. 外れ値がデータに与える影響

それでは、外れ値が含まれている可能性のあるデータと、外れ値の影響を低減する処理を行ったデータでは、どの程度の違いがあるのか、上記の自己ペース読み課題によって得られた読解時間データを用いて比較してみよう。表 1 は、関係節の構造（主格×目的格）と語彙頻度（高頻度×低頻度）ごとの平均読解時間を表している。外れ値処理の方法は、実験参加者ごとに理解度チェック問題に正解した文（関係節文・フィラー文すべて含む）の全単語の読解時間の平均+3 標準偏差の基準値を求めて、その値を超えた読解時間の値を、基準値に戻す方法をとった。この処置により値が変わったデータは、全データの約 4% であった。なお、カッコ内の値は、各セルの標準偏差値を表している。

表 1. 外れ値処理の前後のデータの変化

外れ値処理前のデータ	主格関係節文	目的格関係節文
高頻度語文	771 (319)	1049 (542)
低頻度語文	988 (346)	1039 (493)
外れ値処理後のデータ	主格関係節文	目的格関係節文
高頻度語文	752 (311)	935 (417)
低頻度語文	937 (309)	983 (454)

外れ値の影響を低減する処理の前後を比べると、処理後に平均値・標準偏差値が下がっていることがわかる。特に目を引くのは、高頻度語・目的格関係節文で、平均・標準偏差とも 100 ミリ秒以上下がっている。この条件の読解データに、影響の大きい外れ値が相当数含まれていた可能性が考えられる。これに伴い、高頻度語文における主格・目的格関係節文の平均読解時間の差は、処理の前後で 278 ミリ秒から 151 ミリ秒とかなり小さくなっている。このことから、外れ値の取り扱いによって、データの分析・解釈に大きな影響があるということが言えるだろう。

6. どの方法を使えば良いのか？

6.1 先行研究を参考にする

ここからは、実際に L2 心理言語実験を行った際に、外れ値の影響を低減させるために、どの方法を使えば良いのかを考える。上では、実験参加者ごとに平均+3 標準偏差を基準値として、それを超えた値を基準値に変換した結果を示した。しかし、これが最上の方法であるという保証はどこにもない。この手続きの後で処理効果（主格・目的格の差）がかなり小さくなったと記したが、これは、本来の文理解処理を反映しているが、かなり大きな値であったために、「誤って」外れ値としてその影響を低減させられた読解時間値が多く存在したためであるからかもしれない。このように、どの手法を採用すればよいか、という質問に答えることは、極めて困難なことであると言える。

この問いに対する答えの求め方の 1 つは、先行研究でどのような手法がとられているか、ということである。表 2 は、過去に主な L2 文理解研究で行われた自己ペース読み課題のデータ分析の際に、どのような外れ値処理の方法が用いられたかをまとめたものである。

表 2. 先行研究で用いられた外れ値処理の方法

先行研究	提示モード	外れ値処理の方法
Dussias & Cramer (2006)	word-by-word	200ms以下と2000ms以上を除去
Juffs (1998)	word-by-word	時間切れを除去+分析領域における、各グループの平均+3標準偏差以上を除去+対数変換
Juffs (2005)	word-by-word	各条件における各単語のグループ平均+3標準偏差以上の値を除去+対数変換
Hopp (2006)	phrase-by-phrase	5000ms以上を除去+各分析領域のグループ平均+2標準偏差以上を、基準値に変換
Felser et al. (2003)	segment-by-segment	各条件ごとにグループ平均±2標準偏差の範囲外の値を除去
Papadopoulou & Clahsen (2003)	segment-by-segment	母語話者は、条件ごとに平均±2標準偏差の範囲外の値を除去、L2話者の場合は、±2.5標準偏差の範囲外の値を除去

この表からも、外れ値の取り扱いについて、一貫した方法がないことは明らかである。これは L2 の心理言語学研究だけに特異な状況ではなく、L1 の研究においても大差のない状況である。無論、研究対象となっている言語構造の違いや、実験課題の提示モード（語レベル・句レベルなど）の違いはあるが、これだけ取り扱いにばらつきがあると、先行研究にあたってだけでは、適切な処理方法を 1 つ選択することは不可能である。

6.2 Ratcliff (1993)のシミュレーション研究

外れ値の取扱いは、上記のとおり手法の選択が困難であるが、そのためのガイドラインとなりうる報告が、人工的なデータを用いたシミュレーション研究からなされている。Ratcliff (1993)は、語彙性判断課題に代表される、2 択の素早い反応が要求される課題の反応時間の人工的な分布を条件ごとに作成した。さらにその分布は、主効果の位置、実験協力者間の分散の程度、外れ値の有無なども統制されている。こういった分布を数多く作り出し、上で示した外れ値の影響を低減する処理を行い、それぞれに統計（分散分析）をかけた結果、どの手法が統計的に有意になりやすいかを調べている。その結果は、以下のようまとめられている。

- ・任意の基準値以上の値を除去した場合、最も統計的に有意になりやすい
 - 実験参加者間の分散が小さい場合、また主効果が分布の全体的移動によって現れる場合は、1,000 ミリ秒、2,000 ミリ秒などの絶対値による基準値が有効
 - 実験参加者間の分散が大きい場合、または主効果が分布の tail に現れる場合は、標準偏差をベースにした基準値が有効
- ・データ変換（Log 変換・逆変換）は、共に統計的に有意な結果を生みやすい。
- ・基準値以上の値を、基準値に置き換える手法は、あまり推奨されない。

これらの結果は、L1 の語彙性判断課題などの実験データをシミュレートした分布に基づ

くもので、自己ペース読み課題など、分布の在り方が異なる別の反応時間計測実験に応用するには注意が必要である。さらに、実験参加者間の分散が大きい L2 での実験では、さらに慎重になる必要がある。例えば、基準値以上の値を、基準値に置き換える手法は、Ratcliff (1993)では推奨されていないが、シミュレーションデータに外れ値が含まれない場合に、無駄にデータを削除せず有意な結果を多く示していた、という点では他の手法よりも優れているとも言える。実験参加者間の分散が大きい L2 学習者の反応時間分布では、実際の処理を反映した反応時間と、外れ値に該当する反応時間が多く重なることが予想される。このため、基準値外のデータをすべて除去する手法よりも、一定の基準値に置き換えて外れ値の影響を低減するという手法の方が、好まれる可能性がある。

7. L2 文理解実験のデータを用いた検証

では実際の L2 文理解実験のデータを使って、これまでに述べてきた外れ値の取扱い手法が、分析にどのような影響を与えるかを検証してみよう。表 3 は、表 1 でも示した、外れ値の影響を低減する処理を行っていない、生の平均読解時間と標準偏差の値である。

表 3. 生の自己ペース読み課題結果

	主格関係節文	目的格関係節文
高頻度語文	771 (319)	1049 (542)
低頻度語文	988 (346)	1039 (493)

表 4 は、表 3 のデータに対して、外れ値の影響を低減する様々な手法をかけた後、分散分析を行った際の構造と単語頻度の交互作用の結果である。

表 4. 外れ値処理の手法ごとの構造×単語頻度の交互作用

	F 値	p 値	partial η^2
生データ	7.77	.011	.261
平均+3SD以上を除去	0.60	.447	.027
平均+3SD以上を基準値に変換	4.55	.044	.171
Log変換	6.40	.019	.225
逆変換	5.73	.026	.207
2,000ミリ秒以上除去	1.75	.200	.074
3,000ミリ秒以上除去	3.70	.067	.144

この結果の中で目を引くことが 2 点ある。1 つは、全くいじっていない、生のデータにおいて効果量が最も大きい値を示している、ということである。もう 1 つは、一定の基準値以上の値を除去した場合、統計的に有意な結果が得られにくくなる、ということである。

この傾向は、高頻度語文における構造の単純主効果を調べた際も同様である。表 5 は、生データと平均+3 標準偏差以上の値を除去したデータで、高頻度語文の主格・目的格関係節文の読解時間がどれくらい違うかを表したものである。また表 6 は、高頻度語文における構造の単純主効果を調べた結果を、外れ値処理の手法ごとにまとめたものである。

表 5. 外れ値処理の手法ごとの高頻度語文読解時間データ

高頻度語文	主格関係節文	目的格関係節文
生データ	771 (319)	1049 (542)
平均+3SD以上を除去	705 (259)	764 (277)

表 6. 外れ値処理の手法ごとの高頻度語文の構造単純主効果

	F 値	p 値	partial η^2
生データ	16.65	< .001	.431
平均+3SD以上を除去	2.77	.110	.112
平均+3SD以上を基準値に変換	16.58	.001	.430
Log変換	13.73	.001	.384
2,000ミリ秒以上除去	0.20	.660	.009
3,000ミリ秒以上除去	6.81	.016	.236

これらの結果を基に考えると、どの手法を採用するかというよりも、どの手法を採用すべきではないかということについて、いくつかの提案が可能である。まず、生データによる分析は、有意差指標、効果量とも最大であるが、これは 1 語を読むのに 10 秒 (=10,000 ミリ秒) 以上かかったようなデータも、そのまま分析に反映されているために生じた結果である。この長い時間の間には、言語理解以外の様々な処理が行われた可能性が高いし³、このような数少ない極端な値によって得られたデータを基に、何らかの結論を導き出すとすることは不適當である。反対に、有意な結果が得られなかった、基準値以上の値を除去する手法だが、先行研究の結果と照らし合わせてみると、適切な手法であるとは言い難い。関係節処理における主格・目的格の難易度の差は、ある程度熟達度の高い話者であれば、L1 でも L2 でも頑健に観察される事象である。数値上は、基準値外の値を除去した場合でも目的格関係節文の読解時間の方が長い(表 5 参照)、有意差が検出されないというのは、データ除去の方法が強力すぎたのではないかと考えられる。特に 2,000 ミリ秒のような低い絶対値を基準値とした場合、多くの値が除去の対象になってしまうため、少なくとも L2 の文理解実験においては適していないと言えそうである。

8. まとめ

以上、Ratcliff (1993)のシミュレーション研究の結果と、筆者の L2 文理解実験のデータ

を基にした分析結果から、L2 の心理言語実験における外れ値の取扱いの流れとしては、以下のような手法が勧められる。

- ① ある程度のデータを集めて、その分布を見る。語彙性判断課題などの課題では、条件ごとに分布全体がずれるので、2,000 ミリ秒などの絶対値を基準として、その範囲外の値を除去するという方法も可能である。自己ペース読み課題のように、効果が分布の外側に現れ、かつ実験参加者間の分散が大きい場合は、標準偏差に基づいた基準値設定を行い、除去、基準値変換のどちらも可能性に入れておいた方がよい。
- ② 先行研究のデータの分布に関する情報に当たる。その研究で、外れ値の取扱いについてどのような手法が用いられたかを知ること重要だが、なぜその手法を用いたのかという情報があれば、自分の研究に応用できるかどうかを考える指標になり得る。
- ③ データ変換を試してみる。Ratcliff (1993)は、確認の意味で Log 変換や逆変換をかけたデータでの分析を勧めている。最終的な分析・発表（口頭発表・論文）を、Log・逆変換したデータに基づいて行うことも可能である。ただしその際には、変換で得られたデータは平均値の解釈が難しいという弱点があるので、生データの平均などを同時提示した方がよい。
- ④ 新種の、あるいは予期していない効果なら、追従実験を繰り返して行う。

外れ値の可能性のある値を特定して、その影響をできるだけ小さくしようとする取り組みは、本質的に研究に関係のない数値を除去するために行うものであり、決して統計的有意差を求めやすくするために行うものではない。自分にとって都合のいい手法を、異なる研究で使い分けるということは、少なくとも同一の実験手法を用いている限りは、厳に慎むべきである。また、今回はすでに得られたデータに含まれているかもしれない外れ値をどのように取り扱うか、という点に焦点を絞ったが、副次的なデータ（例：実験中の参加者の様子をビデオ撮影、実験終了後に全刺激を紙で提示して関係のないプロセスがあった部分を指摘させるなど⁴）を得ることによって、さらに詳細に外れ値の特定ができる可能性がある。微細なことではあるが、これらを積み重ねていくことで、より精緻な L2 言語処理モデルの構築が可能になるのではないかと考える。

最後に、外れ値の取扱いについて、最も基本的な心構えを表した 2 つの文を引用して、本稿を閉じたいと思う。

Do not use several methods and choose only the one that is significant. (Ratcliff, 1993, p. 519)

Any transformations, changes of scores, and deletions are reported in the Results section together with the rationale. (Tabachnick & Fidell, 2006, p. 78)

注

- * 本稿は、第4回 LET 関西支部メソドロジー研究部会、第10回言語テスト・第二言語習得合同勉強会（2011年2月27日、於関西大学）において口頭発表したものに加筆・修正を加えたものである。このように文章化する形で勉強する機会を与えてくださったメソドロジー研究部会の先生方、当日の発表で有益なコメントをくださった方々、また本稿執筆中に様々なご助言をくださった、酒井登、平井愛の両氏に、ここに記して感謝申し上げます。不完全な部分も多々含んでいると思われるが、ご意見、ご提案等いただければ幸いです。
- 1 本稿における「心理言語実験」は、反応時間計測パラダイムに従った言語実験に限定している。このようなリアルタイムの言語処理を観察するオンライン法に対して、質問紙形式の文法性判断課題など、処理の結果としての判断・分類データから、言語処理に関する知見を得ようとするオフライン法も、心理言語学では多く用いられている。（郡司・坂本, 1999）
 - 2 例文中の下線は、分析対象領域を示している。各文の読解時間は、この分析領域の平均読解時間によって求められた。
 - 3 例えば、オンラインでの理解処理が破たんして、通常の言語理解処理とは違った形で、何とか文を理解しようとしている可能性なども考えられる。オンラインの言語処理を考える心理言語実験では、このような処理も原則分析対象外である。
 - 4 口頭発表中に、フロアからいただいたご指摘による。

参考文献

- Dussias, P., & Cramer, T. (2006). The role of L1 verb bias in L2 sentence parsing. In D. Bamman, T. Magnitskaia & C. Zaller (Eds.), *Proceedings of the 30th Boston University Conference on Language Development* (pp. 166-177). Somerville, MA: Cascadilla Press.
- Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453-489. doi:10.1017/S0142716403000237
- 郡司隆男・坂本勉 (1999). 『言語学の方法』 東京: 岩波書店.
- Hashimoto, K. (in press). Syntactic processing of L2 English relative clause sentences: The effect of proficiency. *Annual Review of English Language Education in Japan*.
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369-397. doi:10.1191/0267658306sr272oa
- 門田修平. (2010). 『SLA 研究入門 第二言語の処理・習得研究のすすめ方.』 東京: くろしお出版.
- Juffs, A. (1998). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning*, 48, 107-147. doi:10.1111/1467-9922.00034
- Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, 21, 121-151. doi:10.1191/0267658305sr255oa
- Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in

- English. *Cognitive Linguistics*, 16, 313-353. doi:10.1515/cogl.2005.16.2.313
- Osborne, J. (2002). Notes on the use of data transformations. *Practical Assessment, Research and Evaluation*, 8. Retrieved from www.microbiologybytes.com/math/graphics/transformation.pdf
- Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing. *Studies in Second Language Acquisition*, 25, 501-528.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532. doi:10.1037/0033-2909.114.3.510
- Tabachnick, B. & Fidell, L. (2006). *Using Multivariate Statistics*. Boston, MA: Pearson Education.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285-318. doi:10.1006/jmla.1994.1014