

外国語教育研究と直交表を用いた実験計画 —実験計画の効率化を求めて—

草薙 邦広

名古屋大学大学院生

概要

本稿の目的は、実験計画法のひとつである直交表実験計画の概要を述べながら、外国語教育研究分野におけるその応用可能性を探ることである。外国語教育研究における実験的研究においては元来、(a) 標本サイズが調整しにくい場合が多い、(b) 実験参加者にとって実験自体が不公正な処置になってしまう、(c) 厳密な実験計画では実験参加者の負担が大きすぎる場合が多い、といった実験実施上の実用的問題がボトルネックとなる場合が少なくない。そのため、「実験計画の効率化」といった実験実施上の側面が重要となってくる。本稿では、実験計画法、特に直交表を用いた実験計画の基礎に触れ、外国語教育研究における研究事例(川口・草薙, in press)の紹介を行う。

Keywords: 実験計画法, 直交表, 研究方法論, 統計処理, 品質工学

1. 背景

近年の外国語教育研究は、しばしば様々な分野やアプローチの複合体のような様態を成しており、その研究テーマや研究手法は非常に多岐に渡る。このような様態を、簡便な二元論的区分けで見ると、一方は、教師のナラティブ、ケーススタディ、教育実践データの蓄積、アクションリサーチといった手法に関連付けられる、現実的でしばしば研究者や学習者自身の観察を伴うような研究であろう。もう一方は、脳機能イメージングに代表的である生理学的実験、それに準じる心理言語学的実験、またはシミュレーションなどを伴う数理的アプローチといった、ややもすれば理想的で、現実に対して積極的に干渉を行う研究と捉えられよう。もちろん、このような区分けは一元的で、非常に視野の狭いものであり、その中間も段階を成すと考えられる。学習者を対象とした言語テストや、質問紙調査、そして指導・干渉型実験などがその中間部に位置するといえよう。

本稿は、以上のような区分における中間的な研究に関するものである。現実の、統制が極めて困難と考えられる教育実践自体の観察や、精度の高さや知見の再現可能性を極めて重視する、理想化された状況の研究ではなく、そのちょうど中間ないし、やや理想化の方向に落ちるような研究を本稿の対象とする。具体的には言語テストや、質問紙調査、

指導・干渉型実験などが主な視野である。

このような類の研究は、しばしば推測統計といった量的手法に基づくものであり、測定の妥当性や信頼性はもちろんのこと、母数の推定の精度や、統計的仮説検定などを用いる場合はその質、そしてデータの要約についてもその再現可能性などが厳しく問われる。しかし、現実的制約がかかるこの類の研究では、実験実施上の様々な問題点はそのボトルネックとなる。以下に典型的な例を三つ示そう。

まず、標本サイズの大きさである。統計的仮説検定においては、各種検定統計量やそれに対応する有意確率は、標本サイズに大きく依存することが知られている。例えば、平均差の検討に用いられる t 検定の場合、標本平均差や標本効果量が同一であるとき、標本サイズの増大に伴い、 p 値は単調低下する。また、回帰分析や因子分析も同様に、標本サイズが十分でないと分析が正確でない場合が多い。しかしながら、我が分野の研究は、実験を任意の回数繰り返したり、予算の限り任意の数の実験用検体を用意したりできるような分野とは異なり、学校教育なりの現実的制約から独立したものではいられない。例えば、実在するクラスを対象とする場合は、クラスサイズの制約が伴う。

望ましい統計分析を行うためには、適切な標本サイズを予め決定する必要がある。研究者は、任意の有意水準 (α)、先行研究に基づく効果量の目安、そして目標とする検定力 ($1-\beta$ 、しばしば .80) を満たす標本サイズを計算する検定力分析 (事前の分析) を用いて標本サイズを決定することができる。また、任意の信頼区間の幅を満たす標本サイズを決定する正確度分析など、研究計画において望ましい標本サイズを統制するための分析はある。しかしながら研究者は、必ずしも統計的に望ましい標本サイズを統制できるわけではないのである。仮に事前の検定力分析により、望ましい標本サイズが 81 人だったとして、実験を受け入れられるクラスサイズが 20 であり、如何とも標本サイズを満たすことができない、という状況があり得よう。このように、標本サイズの統制の困難さは、統計上、実験計画上の望ましさの実現をしばしば阻む。もちろん、逆に大きすぎる標本サイズは、実質科学的に意味の重要性が相対的に低い微小な効果量に対して有意差を検出する可能性を高める。また、その 1/10 の標本サイズを以ってして望ましい検定力を再現できる効果量が極めて大きな研究であれば、9/10 の標本の必要性が研究倫理上問われても仕方がない。このような事にも留意すべきであろう。

次に、指導・干渉型実験によく見られるものであるが、実験自体の処置による不公正さが不可避な出来事として生じる場合がある。仮にクラス A とクラス B にそれぞれ指導法 A' と B' による処置を施し、その後同一のテストで成績を比較した場合、平均差が有意であったり、標準化平均差が大であったりする場合は外国語教育研究で少なくない。このような場合、クラス B は不公正な処置を受けたことになる。このため、対照群を用意するような望ましい完全配置の実験を行うことが困難な場合も多々ある。

更に、同様に倫理的配慮によるものであるが、教育現場の実践性から、実験や調査

研究などの負担を避けなければいけない可能性もある。また純粋に、実験参加者の疲労という要因も潜在的脅威であるが、実験実施時間が限られるならば、そもそも反復測定において複雑な実験計画によるデータを取ることが不可能であったり、困難であったりし得る。残念ながら昨今の研究がこのような問題を明示的に論じる機会を多く持つとは言い難いが、実験計画上の必要性は指摘できるだろう。

これらの点は、全て研究、特に実験計画（ここでは調査研究も含む）の効率化の必要性を訴えるものである。外国語教育研究では、上記のような点から、(a) 検定力が低く統計的に望ましくない解釈を行う、(b) 微小の効果量を検出できなく、研究の視座が精細でない、(c) 望ましい完全配置の実験計画が無理な場合がある、(d) データ自体が限られる、といった実情が鑑みられる。このような点の克服について取り組むことが重要であることに異論はなからう。

一方で、国内では理想的実験環境を目指すあまり、多元配置分散分析を多用する研究や、標本サイズの適切さに言及がなく、更に実験参加者への倫理的配慮、疲労要因についての配慮がみられない研究が一定数存在する。また、特に多元配置分散分析に関しては、一般に三要因交互作用が見られることは少ないことが知られており、実際に存在しそうな交互作用を取り上げ、また、そのために実験の負担を増やすことで結果的に検定力が下がることもあり、望ましい分析とは言い難い場合もある。また、多水準の実験計画は解釈が困難であったり、統計的に多重比較の問題を招いたりする可能性もある。

しかし、このような研究は一概に批判されるべきではない。外国語教育研究において、研究の興味たる要因は無数にあり、その交互作用も検討すべきである場合が大多数である。交互作用の検討は二元配置以上の計画でなければ不可能であり、また理論的に四次の交互作用の解明が必須という現象もあり得ないとはいえない。このように、私たちの研究における要因やモデルは極めて複雑であり得るが、それを表現し得る豊かなデータは極めて得にくい、といった事態が恒常的である。もちろん、教育研究における一般化の難しさを考えれば、単発実験の精度の高さや実験計画の複雑さを求める必要性には疑義が投げかけられるであろう。また、メタ分析や追行研究（replication）の必要性と重要性も知られるところである。だからこそ、記述的で広範囲に渡る「当たり前の教育データ」の記述と共有こそが、精緻な知見を得ることを目標とする研究に併存する形で望ましいとはいえないだろうか。

直交表を用いた実験計画が、このような状況に対して一助となるかもしれない。

2. 直交表を用いた実験計画

2.1 直交性という性質

直交表（orthogonal array）とは、直交性（orthogonality、後述）を持つ実験配置を予め整理した表と考えられる。直交表を用いた実験計画は、実験計画法のひとつを成し、実

験の「一部実施法」（部分配置実験）とみなせる。一方，一般的な分散分析を用いた実験計画は，全ての要因の組み合わせについて実験を実施するため，「要因配置実験」（完全配置）などと呼ばれる。本稿の及ぶ範囲ではないが，実験の一部実施法には，ラテン方格法，グレコ・ラテン方格法などが知られる。直交表を用いた実験計画は，主に品質工学の分野における研究や調査に見られ，提唱者にちなみ，タグチメソッド（Taguchi）などと呼ばれる場合もある。

直交性とは，数学的には説明変数同士のベクトルの内積が 0 である性質を示す（相関係数が 0 であるともいえる）。通常の分散分析における二元配置実験計画を例にとる。要因 1 と要因 2 がそれぞれ 2 水準（A, B）を持つとする。直交性がある配置は，各要因においてそれぞれの水準が同一回出現する。一方，直交性が崩れた配置はこの条件を満たさない（表 1）。なお，直交性は通常の（要因配置の）分散分析においても満たされており，むしろ直交性は分散分析の要件でもある。

表 1.
直交性を持つ配置と直交性が崩れた配置の例

直交性あり			直交性なし		
組み合わせ	要因 1	要因 2	組み合わせ	要因 1	要因 2
A ₁ A ₂	A ₁	A ₂	A ₁ A ₂	A ₁	A ₂
A ₁ B ₂	A ₁	B ₂	A ₁ B ₂	A ₁	B ₂
B ₁ B ₂	B ₁	A ₂	A ₁ B ₂	A ₁	A ₂
B ₁ B ₂	B ₁	B ₂	B ₁ B ₂	B ₁	B ₂

注：上記の表の灰色部分は直交性を乱している部分を示す。

要因配置実験において，単純に要因を増やしていくと，実験がやがて莫大な数になってしまう。例えば，二要因二水準では，その組み合わせは 4 であるが，三要因二水準では 8，四要因二水準では 16 であり，五要因二水準では 32 となり，二水準系 n 要因の実験では 2ⁿ 回となる。三水準系はもちろん，3ⁿ 回の実験となる。

直交表を用いた実験計画における最も重要な点は，直交性自体，このような要因配置によらずとも保てる組み合わせがあるということである。例えば，交互作用を計算せず，四要因の主効果のみを検討するのであれば，以下のような配置でよい（表 2）。このように直交性が満たされてさえいれば，各要因（因子）が交絡することはない。例えば，以下の例における要因 1 の二水準間の比較には，単純に 1-4 行目，5-8 行目の平均値（M）および標準偏差（SD）を用いればよい。更に，要因 2 は，1, 2, 5, 6 行目と，3, 4, 7, 8 行目を比較するとよい。以下も同様である。

表 2.

直交性を持つ配置の例（四元配置二水準系）

	要因 1	要因 2	要因 3	要因 4
1	A ₁	A ₂	A ₃	A ₄
2	A ₁	A ₂	B ₃	B ₄
3	A ₁	B ₂	A ₃	B ₄
4	A ₁	B ₂	B ₃	A ₄
5	B ₁	A ₂	A ₃	B ₄
6	B ₁	A ₂	B ₃	A ₄
7	B ₁	B ₂	A ₃	A ₄
8	B ₁	B ₂	B ₃	B ₄

このような直交性を満たす配置は多くないため、直交性を持つ配置（直交表）は既に分かっており、用意されている。直交表自体には様々なものがあり、二水準系、三水準系などが様々であるが、以下に最も基本的な、二水準系、7列8行からなる $L_8(2^7)$ 直交表を紹介する（表 3 に示す）。この直交表の組み合わせに対して相関行列を求めると相関係数が全て 0 であることがわかる。

表 3.

 $L_8(2^7)$ 直交表

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

2.2 直交表を用いた分析

上記のような直交表を用いて実験を計画するには、最初に、列に対して要因を割り付ける。 $L_8(2^7)$ 直交表は、7列から成るため、最大で7要因の割り付けを実行することができる。例えば、前節の例のように、四要因の主効果のみの検討を行うのであれば、1, 2, 4, 7列にそれぞれ因子を割り当てるとよい（これは結果的に表 2 と同様になる）。一方、

任意の要因の組み合わせについて交互作用を検討したい場合は異なる割り付けを行う必要がある。例えば、要因 A・C の交互作用、要因 B・C の交互作用、および要因 A~D までの主効果を検討したい場合、2 列目に要因 C、3 列目に要因 A、5 列目に要因 B、7 列目に要因 D を配置すると、1 列目には要因 A・C、6 列目には要因 A・B の交互作用があらわれる。残りは誤差の計算に用いることができる。交互作用を取り上げられることについては、例えば 1 列目において、要因 A・C の組み合わせが交差してあらわれる点に着目されたい。

上記のような要因の割り付けは容易ではないため、線点図というものを利用するとよい。線点図は、各列に割り当てられる要因を、線（交互作用）と点（主効果）で表す。線点図は、各直交表に対して複数あることが多い。ここでは $L_8(2^7)$ 直交表における線点図のひとつを示す（図 1）。数字が割り当てられる列番号を表す。

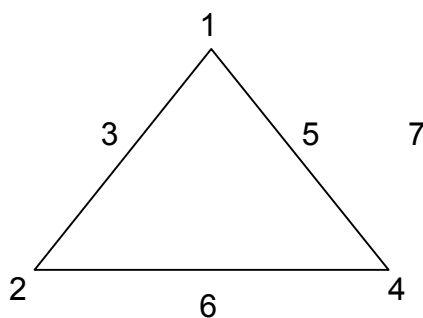


図 1. $L_8(2^7)$ 直交表における線点図のひとつを表す。

仔細は本稿の及ぶところでないが、統計的な分析は、「列毎に集計する」ことで分散分析を行うことができる。つまり、全変動 = 平均変動 + 因子の効果 + 残差と捉え、それぞれに分散を分割し、平均平方和の比率 (F) を求め、更に自由度から p 値を求めるとよい。¹ 重要なポイントは、割り当てを行わなかった列を誤差項とするところである。また、直交表を用いた実験計画ではしばしば「プーリング」と呼ばれる処置を行うことがある。プーリングは交互作用が比較的小さい列の効果を誤差項に回し再分析する手順である。プーリングに関する仔細についても附録で紹介している文献などを参照されたい。

直交表を用いた平均値の比較については独特の可視化の仕方をする場合がある。例えば、四つの語彙特性ごとに語彙性判断課題の反応時間を可視化する場合、図 2 のような方法が用いられる。各要因の線が独立しており、それぞれの水準の差を示している。また平均値のプロットには誤差棒を用いている。あくまでも図 2 は仮想データであるが、誤差棒は母平均の 95% 信頼区間を表している。品質工学など、直交表を多用する分野ではしばしば最適水準を見つけることを目標とするため、このような要因が従属変数に及ぼす影響が一度に検討できる可視化が好まれる。

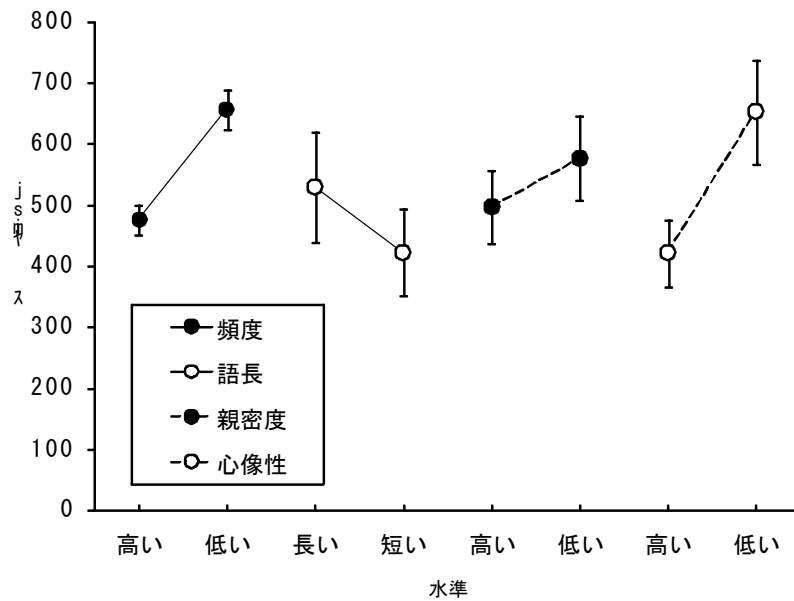


図2. 上記の折れ線グラフは95%信頼区間つきの各要因・水準の平均値を表す。

2.3 直交表を用いた実験計画の展望

このように、直交表を用いると実験計画の効率化を図ることができる。例えば、四要因二水準の要因配置実験では、全水準について実験するためには、16回の実験回数が必要であるが、 $L_8(2^7)$ 直交表を用いるとその半分となる8回の回数で済む。本稿では紹介していないが、三水準系の直交表を用いて主効果のみを検討する場合、四要因三水準配置実験の実験回数の1/9でよいので、大幅な効率化が期待できる。

最適水準の探索と区間推定も重要な手法のひとつである。交互作用を仮説として持っていない場合でも、各要因における最適水準の組み合わせは重要な情報となろう。また、今後の実験の数量的手がかりにもなりうるため、母平均値、不偏標準偏差、母効果量の点・区間推定が重要であり、またこれらは検定力分析や正確度分析などとの併用も望まれるであろう。しかし、このような方法はあくまでも探索的であって、確証的でない。背景の部分と関連するが、このような探索的であってでも効率的な方法と精緻な知見を求める研究が併存することが望ましい。もちろん、その橋渡しの研究も重要性を担うだろう。

また、「魚骨図」ないし「特性要因図」と呼ばれる、現象に影響を及ぼす要因を列挙した図のように、知見の共有や実験計画段階の構想に役立つ可視化も必要である。このような試みはあまり外国語教育研究では見られない。今後の研究が望まれるところである。

直交表は、特にソフトウェアのバグ検出に絶大の効果を示している。このような手法を援用して、教材、ソフトウェア、テストなどの関する多特性の探索的調査を行うこともできよう。例えば、ある文法テストにおける特性；時間制限（強・中・小）、項目の種類（語彙、統語、コロケーション）、形式（2択、3択、記述）、干渉負荷（なし、ノイズ、

書字混合)の効果を効率的に検証することができる。また、予備実験としてこのようなデザインを用いることで、本実験の結果が極めて効果的に予測できるだろう。²

実用的な観点について述べるならば、市販の様々なテキストが表計算ソフトを用いた分析の手法を丁寧に解説している(附録の文献リストを参考のこと)。また、統計解析環境である R は、計算の自由度が高いことから、心得のあるものであれば簡便に分析ができよう。R には専用のパッケージ DoE.base³などが用意されており、これらを用いることで分析はもちろん、要因の割り付けも行うことができる。

3. 研究事例

最後となるが、ここでは直交表を用いた実験計画による探索的研究、川口・草薙(in press)の概要を紹介する。

川口・草薙(in press)は第二言語の語彙特性が語彙性判断課題における判断時間に及ぼす影響は、語彙の即時的運用能力の発達に伴い緩和されるだろうという予測に基づき、直交表を用いた多特性の語彙性判断課題を実施した。実験参加者は36名の日本語を母語とする英語学習者であり、全員が大学生・大学院生であった。対象となった語彙特性は、(a)語長、(b)頻度、(c)親密度、(d)心像性の四つであった。これらの語彙特性を四要因、主効果のみを対象とするように $L_8(2^7)$ 直交表に割り当てを用いて実験材料である語彙の統制を行った。8項目で配置が完結するため、3セットとなる24項目、そしてフィルター項目の24項目を加え計48項目のテストセットを成す。対象となる24項目で、交互作用を持たないまま、各要因12項目 vs. 12項目の対比較を行うことができる。

結果として、語長効果のみが顕著な反応時間の遅れを及ぼすことが分かった。語長が反応時間に及ぼす遅れの度合いを個人個人の標準平均差で変数化し、全体の反応時間、自動化係数(CV_{RT})、TOEICスコア、信号検出理論における弁別力 d' などとの多変量相関分析の結果、このような語彙特性効果の大きさは語彙の即時的運用能力との連関を示す可能性が明らかになった。このように、直交表を用いることで多様な指標を効率的に測定することが可能になる。川口・草薙(in press)は、今後このような技術を用いて、高度なユーザーインターフェースの実装、拡張性の実現、項目の精査などを経た、語彙アクセスにおけるベンチマークソフトウェアの開発を展望として述べている。

4. まとめ

本稿では、直交表を用いた実験計画について紹介しながら、その外国語教育研究における応用可能性を述べた。もちろん、背景で述べたような事柄が全て直交表を用いることで解決できるわけではない。しかしながら、実験の一部のみ実施する直交表を用いた計画が、研究におけるデータ収集や分析の効率化という目的に対する一助となる得る可能性は否定すべきものではないだろう。特に、母数の区間推定や、検定力、正確度分析との併

用や、ソフトウェア開発などと併用することによって、更に応用可能性は広くなろう。今後このような手法が、恒常たる研究上のボトルネックを緩和し、外国語教育研究において新たな視座をもたらさんことを願う。

謝辞

本稿における内容について、外国語教育メディア学会（LET）関西支部メソドロジー研究部会 2013 年度秋田研究会で発表した際、フロアの質疑応答によって発表に加筆した部分がある。中でも貴重なご意見を下さった住政二郎先生（流通科学大学）、水本篤先生（関西大学）に謝意を申し上げたい。また、紹介した研究事例の第一著者であり、本稿の内容にくまなく目を通してくれた川口勇作氏（名古屋大学大学院）に感謝を述べたい。しかしながら、本稿の誤りは全て私のものである。

注

1. 分散分析ではなく、重回帰分析を用いる方法がある。
2. このような分析は、外国語教育研究分野ではあまり見られないものの、コンジョイント分析やベネフィット・セグメンテーションなどに通じるものがある。しかしながらこれらの紹介は本稿の及ばないところである。
3. R のパッケージである DoE.base については、CRAN におけるページ (<http://cran.r-project.org/web/packages/DoE.base/index.html>) または、パッケージのマニュアル (<http://cran.r-project.org/web/packages/DoE.base/DoE.base.pdf>) を参照のこと（2014/3/12 アクセス）

参考文献

川口勇作・草薙邦広 (in press). 「反応時間データにおける語彙特性効果に着目した語彙の即時的運用能力：語長・頻度・親密度・心像性に焦点を当てて」『LET 中部支部研究紀要』25.

附録

参考となる文献リスト

- 大平平 (2013). 『実験計画と分散分析のはなし—効率よい計画とデータ解析のコツ』日科技連出版社.
- 奥原正夫 (2013). 『実践に役立つ実験計画法入門』日科技連出版社.
- 田口玄一 (2010). 『<復刻版> 第3版 実験計画法 上・下』丸善.
- 中村義作 (2013). 『よくわかる実験計画法』近代科学社.
- 永田靖 (2000). 『入門 実験計画法』日科技連出版社.
- 森田浩 (2013). 『図解入門 よくわかる最新実験計画法の基本と仕組み—実験の効率化とデータ解析の全手法を解説』秀和システム.

- 森田浩・奥村清志・今里健一郎. (2011). 『Excel でここまでできる実験計画法——一元配置
実験から直交配列表実験まで』 日本規格協会
- 山田秀 (2004). 『実験計画法 方法編—基盤的方法から応答曲面法、タグチメソッド、最
適計画まで』 近代科学社.