

【書評】

玉岡賀津雄(2023)『決定木分析による言語研究』(くろしお出版)¹

張 婧禕²

DOI: 10.18999/stul.37.5

1. はじめに

2023年7月にくろしお出版から『決定木分析による言語研究』(玉岡賀津雄著)^{たまおかか つ お}が出版された。著者の玉岡は1990年にカナダのサスカチュワン大学の大学院で、日本人児童の日本語能力と語彙処理の研究で博士号(Ph.D.)を取得した。その後、心理言語学者として、語彙の音韻・書字・意味・統語情報、句および文構造などの言語処理に関して、第1言語、第2言語、外国語を対象に実証的な研究を幅広く展開している。たとえば、中央アメリカのゲアテマラで話されているマヤ諸語のカクチケル語の文処理を研究して、動詞・目的語・主語のVOSが正順語順であることを実証している。また、スリランカのシンハラ語は、SOVを正順語順としているが、口語表現ではどのような語順でも自由に作ることができる。文処理の研究の結果、SOVが正順語順であるが、SVO語順も準正順語順であることを示した。これについては、SVO語順の英語が日常的に頻繁に使われていることの影響かも知れないと説明している。さらに、南太平洋のトンガ王国で話されている少数言語であるVSO語順の研究も進めている。

2021年3月に名古屋大学を定年退職し名誉教授となった後、2021年9月から2023年8月までの2年間、湖南大学の兼任教授となり、2023年9月からは上海大学の教授となっている。2023年9月までに、SSCI学術誌³に論文を61本執筆しており、その内の30本が筆頭

¹ English title: TAMAOKA, Katsuo (2023) *Linguistic research by decision tree analysis*. (Kurosio Publishers)

² ZHANG, Jingyi, University of Miyazaki, Japan, E-mail: jingyizhang@cc.miyazaki-u.ac.jp

³ Web of Knowledge 掲載の impact factor が算出されている国際学術誌を指す。

著者である。2021 年4月から日本の小出記念日本語教育学会の会長に就任した。心理学研究, 認知科学, 言語学分野などで学術誌の編集委員および査読者として国内・国際の学術団体に貢献している。

玉岡は本書の出版を契機に, 2023 年6月に「言語実証プロジェクト」(<https://sites.google.com/view/srpls>)を立ち上げた。このプロジェクトは, ネット・AI 時代に対応し, 世界で研究に携わっている人々に学術情報を提供することを目的としている。これは従来の研究者の直観による記述的アプローチに対し, 客観的かつ科学的手法に基づいて言語的特徴を証明・反証する必要性を提唱するもので, 実験・調査による研究技法と分析方法をオンライン講演会の開催で主張している。玉岡はこのプロジェクトのリーダーとして, 実証研究の成果を広く世界に向けて発信することに努めており, 研究情報の共有と教育の促進および持続可能な社会の構築に寄与している。

近年, 母語および外国語に関する言語研究において, 反応時間パラダイムによる言語処理, 脳波による言語処理メカニズム, 視線計測による時間継続的な言語処理など高度な測定技術を駆使した実証的研究が盛んになってきた。こうした新しい研究の流れに乗って, 最先端のテキストデータ分析技法およびそれらの現実社会への応用が注目を浴びている。また, 大学をはじめとした教育機関では, 「文系か理系か」という学問領域の壁を越えて, 新しい知の創造に向けた多角的かつ柔軟な教育活動の試みが展開されており, 大量のデータを解析できる人材を養成しようという動きが高まっている。こうした研究および教育の新しい潮流の中で, 本書では多変量解析法の一つである「決定木分析」を駆使した研究を研究理論, データ編集, 分析, 考察まで詳しく説明している。

筆者は玉岡教授のもとで, 言語の理論に基づいたさまざまな解析法を学び, 博士号を取得した。この度, 筆者が本書の書評を得る機会を得たことは望外の喜びである。

2. 決定木分析とは

決定木分析(Decision Tree Analysis)は, 現在, データマイニング, 機械学習, マーケティング情報分析などの分野で幅広く使用され, データ分析の中核的なツールの1つとして重要な役割を果たしている。決定木とは, ある意思決定プロセスを視覚的に分析できるツールの1つで, 当初はビジネス・経営戦略関連分野で利用された。決定木というアイデアがいつごろ始まったのかについてはいろいろな説がある。理論的枠組みとして確立されたの

は、アメリカの経済学者である Wagner が、*Principles of Operations Research: With Applications to Managerial Decisions* という著書が 1969 年に出版されたのに遡る。この著書から、1960 年代にはすでに決定木の概念がオペレーションズ・リサーチ⁴分野で広く普及し始めていたことが分かる。ただし、基本的な理論的枠組みが確立された当時は、予測精度が低いなどの問題点があった。1980 年代に入ってから、決定木分析の計算方法の開発が加速した。特に、Kass (1980) が CHAID (Chi-square automatic interaction detection) の計算方法を *Journal of the Royal Statistical Society, Series C (Applied Statistics)* で提案した (Tamaoka, et al., 2010)。ビッグデータ時代に入り、現在もマーケットリサーチなど実務的な目的で決定木分析の新たな計算方法やアプローチが研究されており、進化を続けている。

決定木分析では、ある従属変数を複数の独立変数で予測する統計手法である。従属変数は、調査・観察・研究の目的になるため目的変数と呼ばれ、それを予測する独立変数は説明変数と呼ばれる。目的変数を予測する分析の結果が、説明変数の予測力を基準に樹形図で階層的に描かれる。より強い説明変数は樹形図の上位に配置され、より弱い説明変数は枝が分岐する形で下位に表示される。樹形図に描かれない要因は、目的変数を有意に予測しない変数と判断される。つまり、樹形図の成長プロセスで、独立変数間に有意な相互作用が見つかった場合、枝を分岐する形で描かれるのである。これにより、複数の説明変数を階層的に可視化することができるため、ビジネスや経営戦略などの分野ばかりでなく、言語研究や第2言語・外国語習得における背景諸要因の分析に有効である (Ihara, et al., 2009; 林ほか, 2011; Tamaoka, et al., 2010; 張, 2017; 張ほか, 2018 など)。

決定木分析は、IBM SPSS から *Decision Trees* の名称で統計解析ソフトとして販売されている。このソフトを利用する際の最大の問題は価格であろう。*Decision Trees* は、基本ソフトウェア IBM SPSS Statistics Base のオプションとして提供されている。両方を購入すると、2023 年 11 月 30 日現在の IBM SPSS Statistics 29 の永久ライセンス・アカデミック価格(税別, 保守なし)で (<https://spss.mediamart.jp/price/>を参照), IBM SPSS Statistics Base が 154,400 円, そのアドオンソフトの IBM SPSS Decision Trees が 85,300 円で、合計で 239,700 円になる。これに IBM SPSS Advanced Statistics を追加すると、さらに 85,300 円が加算され、325,000 円になる。他のオプションを追加していくと 50 万円くらいになってしまう。239,700 円(税別)で決定木分析を使うに必要なソフトは揃うものの、高額な支出となろう。現在では、統計解析

⁴ operations research; 数学的・統計的モデルやアルゴリズムで最も効率的な計画になるような判断をするための科学的技法。

用のコンピュータ言語として、無料の R が普及している。R はオープンソースで無料のプログラミング言語であり、統計解析に特化したプログラミング言語である。近年、R を使って統計解析を行う研究者が増えている。しかし、現段階では、R が提供する決定木分析は、IBM SPSS Decision Trees に機能の面で追いついていない。今後、R でより進化した決定木分析プログラムが提供されることが期待される。

3. 本書の構成と概要

『決定木分析による言語研究』は全 10 章からなり、各章で言語研究に関連するテーマが 1 つずつ設けられている。各テーマについて、研究背景から、研究目的、研究対象と調査・実験の実施手順などの具体例を紹介し、IBM SPSS Statistics のアドオン・モジュールである IBM SPSS Decision Trees (以下、Decision Trees) による解析手順と分析結果の解釈を進めている。

第 1 章は、主語の有生性と動詞の自他性の関係をテーマにしている。本書の序章として、日本語の文における主語の有生性を予測する研究を糸口に、従来の記述研究の考え方、および人間の直感との違いから、著者の視点や実証的なアプローチを導いている。この章では、日本語の文における主語の有生性の有無という仮説の検証を巡り、決定木分析の由来、決定木分析を使うための基本概念、および樹形図の見方などを導入している。決定木分析は従属変数のデータの違いによって、「分類木分析」と「回帰木分析」の 2 種類に分けられる。「分類木分析」とは、質的データを扱う予測手法で、カイ 2 乗検定を繰り返すことにより、独立変数間の有意な交互作用効果を自動的に検出するものである。すなわち、複数の独立変数のなかから従属変数を有意に予測した変数を選び、予測力の強い順に、交互作用を含み込んで階層的に分析する手法である。これに対して、「回帰木分析」は、量的データを扱う予測手法である。F 検定を繰り返すことにより、独立変数間の有意な交互作用効果を自動的に検出してくれる。この章の最後は、2 種類の決定木分析の特徴について詳細にまとめている。どちらの種類の決定木分析を使うべきかを判断し、各決定木分析における計算方法をより理解することが重要である。

第 1 章では、分類木分析の自他対応動詞 3 セット (6 動詞) の例が挙げられているが、これはあくまで例として挙げられているだけで、自他動詞に対する主語の有生・無生の頻度についての実際の研究 (玉岡, 張, 牧岡, 2019) では、32 種類の自他対応動詞 (合計 64 動詞)

の出現頻度を対数変換して対応のあるサンプルの検定で、主語の有生性の頻度を実証している。そのため、この例をそのまま適用して少ない数の動詞で、主語の有生性の頻度が証明できると考えてはならない。また、例として挙げられた6つの動詞は、すべて自動詞で無生主語を多くとる典型的な例だけなので、異なる傾向の動詞も含んだほうがよかったと思う。

第2章のテーマは、マレーシア人日本語学習者のスピーチレベルシフトである。語用論に関する研究を分類木分析の事例として、マレーシア人日本語学習者のスピーチレベルシフトの研究 (Jamila, 2008) を紹介している。スピーチレベルシフトは、談話ストラテジーの1つである。会話において、ある口調から別の口調に一時的に変わったり、戻ったりすることを意味する。これが顕著にみられるのは、デス・マスの丁寧体から普通体に変わるダウンシフトとその逆に普通体から丁寧体に変わるアップシフトである。この章では、マレーシア人日本語学習者の会話における普通体と丁寧体の使用頻度が、来日前後で大きく変わることを証明している。そのために、6名のマレー語を母語とする日本語学習者を対象に、年上の面識のない日本人と話すという場面設定で、来日1カ月前、来日4カ月後、来日 11 カ月後および 20 カ月後の4回にわたり、約 30 分の会話を収録して、普通体と丁寧体の使用頻度を計算している。普通体と丁寧体の使用頻度は、質的変数であるため、日本に来る前と日本に来てからの普通体と丁寧体の使用頻度の変化を分類木分析で予測している。

第1章ですでに決定木分析の基本概念とその後の章の内容の理解に必要な情報を提供しているため、第2章ではデータセットとSPSSへの読み込みなど一連の具体的な操作については説明されていない。第2章では、統計分析ソフト **Decision Trees** による分析手順に重点が置かれており、分類木分析の分析手順と樹形図の枝の成長法について詳しく説明している。特に分析用のデータセットの作成、SPSSへのデータセットの読み込み、ケースの重み付けについての解説を詳細に行っている。

分類木分析の結果、マレーシア人日本語学習者は来日前には丁寧体を頻繁に使用していたが、来日後は普通体の使用が多くなり、日本滞在期間が長くなるとさらに普通体の使用が多くなることから分かる。これは、丁寧体だけの使用では、よそよそしい印象になるため、スピーチレベルシフトにより、適度に普通体を使用することで距離感を調整するというコミュニケーション方略がとられるからであると考えられる。樹形図をみると、このパターンは6名のマレーシア人全てについて言えることであるが、性差はなく個人差が大きいことが視覚的に見て取れる。

ここで注意したいのは、来日1カ月前からの時間の経過と共に決定木分析の樹形図が描かれているものの、これは時間的変化を考慮した分析ではなく、たまたま普通体と丁寧体の使用頻度が時間経過と共に綺麗に変化したということである。その意味で、Jamila(2008)の研究は仮説に従ったきわめて綺麗な結果であるが、類似の調査を実施した場合に、必ずしも時間変化に従わない結果になるとは限らないことも念頭に置いておく必要がある。つまり、今回の結果は時間経過に綺麗に当てはまっているものの、あくまでカテゴリーカルな変数として設定しただけなので、初めに来日4カ月後がきて、次に来日1カ月後がくるという順になることもありうるのである。

第3章のテーマは行為要求表現における丁寧度の変化である。ここでは量的変数を扱う回帰木分析の紹介をしている。日本語の行動要求表現では、相手になんらかの行動を起こさせることを目的とし、「くれる／くださる」や「もらう／いただく」などのような授受の補助動詞を含む表現がよく使われる。これらの表現の丁寧度は、「くれる＜もらう＜くださる＜いただく」の順に高くなることが予想される。これに加えて、ストレートに要求するのではなく、「～てもらえますか」と相手に可能かどうかと尋ねることで、より丁寧さが増すとされる。また、「～ていただけますか」の肯定の疑問形よりも、「～ていただけませんか」の否定の疑問形のほうが丁寧であるといわれている。こうして、肯定形か否定形か、平叙形か疑問形か、という言語特性によって丁寧度がどう変わるかを考察している。すなわち、ここでは4種類の動詞と平叙形の「～くれる」、平叙形・可能の「～もらえる」、丁寧・可能の「～てもらえますか」、尊敬の「～てくださいますか」、謙譲・可能の「～ていただけますか」という5種類の行為要求表現とその否定表現を用いて、回帰木分析で予測している。

第2章と異なるのは、この研究では従属変数としての丁寧度について5段階のリッカート尺度を使って日本人母語話者に主観的判断してもらっている点である。そのため、分類木分析のような頻度データではなく、従属変数のデータは丁寧度という量的変数であり F 検定で平均の差を比較している。その結果は非常に綺麗である。たとえば、否定形にするとより丁寧になるという議論については、肯定形の表現の丁寧度が高い場合に限ることが示されている。樹形図によって、先行研究で議論されてきた内容をより限定して示すことができる。また、この章では、回帰木分析の手順を詳しく説明しており、回帰木分析による研究手法と分散分析 (analysis of variance, ANOVA) による研究手法の差異および実験調査を実施する前に研究手法の選択についても説明している。

ここまで、第1章の決定木分析の導入、第2章の分類木分析、第3章の回帰木分析の分析

手順の説明を通して、決定木分析による研究手法の全体像を示している。第4章からは、より多様な応用事例の説明に入り、2種類の決定木分析の実用性と言語研究における応用可能性が示されている。

第4章のテーマは、中国人日本語学習者による間接発話の理解である。学習者による慣習性と非慣習性の間接発話の理解と日本語習熟度を使って、間接発話の理解を回帰木分析で予測している。Taguchi(2008)は、日本語では「慣習的」な間接発話のほうが「非慣習的」な間接発話よりも正答率が低いと報告している。これに対し、李・玉岡(2019)は、中国人日本語学習者に対して、慣習性と非慣習性の間接発話の理解と日本語習熟度(下位・中位・上位群)を測定した。その結果、英語と同様に、日本語でも「慣習的」な間接発話のほうが「非慣習的」な間接発話より容易に理解されることを指摘した。これにより、Taguchi(2008)の問題点を示している。

さて、これが第3章で紹介した回帰木分析と異なるのは、この章では、慣習性と非慣習性および日本語習熟度の説明変数でダミー変数(0と1からなる変数)に変換された間接発話の理解を予測する回帰木分析を行った点である。ダミー変数を説明変数とする場合には、回帰木分析と分類木分析のいずれも使うことができる。李・玉岡(2019)では回帰木分析を使っているが、分類木分析を使うほうが **Decision Trees** で相対リスクがすぐに計算され、その解釈が容易である。この章では、間接発話の研究背景から説明し、日本語習熟度と間接発話の理解に関して各種の能力測定のためのテストの作成とテスト項目の構成までも詳しく解説している。本章は回帰木分析の手法だけでなく、言語研究に関する調査のためのデザインや言語能力測定のためのテストの開発にも貢献するものとなっている。

第5章のテーマは、山口方言話者のアクセントにおける世代間の変化である。山口県内だけに居住している若年層と老年層を対象に、特殊拍(促音・撥音・長音)と二重母音/ai/を含む語を発音する際のアクセント核を置く頻度を世代間で分析した研究(池田・玉岡, 2013)を紹介している。これにより実際にどのくらい方言が消えているかを確認している。方言研究では、特定の典型的なその地方の方言が話せる高齢者数名の会話やインタビューを録音して、表現を多様な視点から研究することがよく行われる。しかし、高齢の同じ地域の人達が同じような方言を話すとは限らない。そのため、分類木分析で、客観的に複数の被験者の世代間の違いを見る方法を紹介している。

その結果、同じ山口方言話者であっても、老齡層と若年層の世代間でアクセント核を置く位置の頻度が大きな変化しており、方言が消えつつあることを実証的に示している。さらに

面白いのは、老齡層の人達の中でも、語彙の特性によって山口方言と思われる位置にアクセント核を置くとは限らないことも指摘している。また、本調査における樹形図から特殊音および二重母音のアクセント核の置き方は聞こえ度の階層 (sonority hierarchy) に従うのではないかという提案もしている。これにより、特定の典型的な方言話者だけを調査するだけでは、世代間の変化を観察することができないことが示されている。

第6章は、連濁をテーマとして扱っている。連濁は、直接的には分節音に関する現象であるものの、語種、音韻環境、語構造など多くの言語事象と関わっていると指摘されている (テイモシーほか, 2017)。第6章では、連濁に関するライマンの法則の普遍性を検証した例 (Tamaoka, et al., 2016) が紹介されている。ライマンの法則とは、複合語の第2要素の第2拍目以降にすでに濁音 (有声阻害子音ともいう) が含まれている場合には、第2要素のはじめにくる清音が濁音にならないという規則である。北海道開拓使として日本政府に雇われた地質学者であったアメリカ人のベンジャミン・スミス・ライマン (Benjamin Smith Lyman) が、アメリカに帰国してから 1894 年に発表した連濁に関する論文 (Lyman, 1894) が日本語に翻訳され、著者名をとってライマンの法則と呼ばれるようになった。

不連濁規則であるライマンの法則の背景には、OCP (Obligatory Contour Principle ; Leben, 1973 ; McCarthy, 1986 など) があるといわれている。OCP とは、類似した要素の繰り返しを避けるという規則である。ライマンの法則に当てはめて考えると、複合語の第2要素に濁音が含まれていれば、同じ濁音が2つ重なるので、それを避けて、連濁を起さなくなるという説明ができる。OCP がライマンの法則など類似性の連続を阻止するという普遍性のある「原理」であるならば、中国人および韓国人日本語学習者がたとえライマンの法則を知らなくても、ライマンの法則によって有声阻害子音の濁音が複合語の第2要素に含まれていれば不連濁規則を直観的に認知し、連濁を避けると考えられる。

そこで、Tamaoka, et al. (2016) は、中国語と韓国語を母語とする日本語学習者を対象に知覚調査を行い、中国語と韓国語を母語とする日本語学習者を対象に、ライマンの法則の普遍性について調査した。中国語と韓国語の日本語学習者による第2要素に濁音がある場合、無い場合と比べて連濁を避けることを実証した。この章では、Tamaoka, et al. (2016) の研究データを使って、分類木分析の方法を説明している。また、中国語または韓国語を母語とする日本語学習者を直接比較するためには、両グループの特性を統制することの重要性を強調している。つまり、中国人と韓国人の2つの日本語学習者のグループの個別の特性が調査の目的に影響しないように、(1) 全員が日本の大学に留学して日本に住んでいるこ

と、(2)性別、(3)年齢、(4)日本語の学習期間、(5)日本滞在期間、(6)日本語の語彙知識、(7)日本語の文法知識、の7つの特性を、韓国人と中国人の2つのグループで統一している。そのための統計手法として、独立したサンプルのt検定を紹介している。そして、中国人および韓国人日本語学習者が、ライマンの法則を知らなくても、直観的に濁音の連続を避けることが示され、ライマンの法則がOCPに導かれた普遍的な原理であることを証明している。この章では、研究の目的を達成するために前提としてなにが必要かを説明しており、決定木分析を行う前段階の準備の重要性を示している。

第7章は、幼児は「なに＝どこくだれ＝どう＝なぜ＝いつ」の順番で、疑問詞を発話できるようになるという先行研究(Clancy, 1989)の仮説を証明するために、分類木分析を使った例を紹介している。Ito ほか(2020)は、345冊の絵本を対象に、読み聞かせの対象年齢別に疑問詞が出てくる絵本数を数えている。表現の繰り返しが多い絵本もあるので、疑問詞の出現頻度ではなく、本の冊数を頻度としている点でユニークである。それを読み聞かせの対象年齢別に疑問詞が出てくる絵本の冊数を集計して、分類木分析を行っている。分類木分析の樹形図から幼児の疑問詞の習得順序は Clancy(1989)が提案した幼児の発話における疑問詞の出現順序はほぼ一致していたことを報告している。幼児の発話における疑問詞の出現頻度を扱う研究は少なくないが、Ito ほか(2020)のように、疑問詞が出現する絵本の数を頻度として、疑問詞の出現順序を考察する研究事例は稀である。本章は分類木分析が多様な研究のアプローチに応用できることを示している点で新奇性に富む章である。

第8章は、副詞と共起する接続助詞の文中・文末の共起頻度をテーマとしている。玉岡(2006)で、「何しろ」「何せ」「現に」「どうせ」「実際」「本当に」「せっかく」の7種類の副詞と共起する「から」「ので」「のに」の3種類の接続助詞の文における出現位置を、分類木分析を使って予測している。ここでは文中または文末で出現する共起頻度を、滝沢(2002)のコーパスを使って集計している。この分類木分析は、接続助詞が文中または文末に置かれる頻度を、接続助詞と副詞の交互作用で予測するというモデルである。こうしたコーパスにおける頻度解析にも、分類木分析が有効であることを示しており、コーパス研究に応用した事例研究の一つである。分類木分析の結果、「から」が文末にくる場合が多く、その際に「何しろ」が最も多く、「せっかく」は少ないことを示している。すなわち、「何しろ…から」は多様な場面で使われ易いものの、「せっかく…から」という表現は、なんらかの理由がなくては使われず、使用条件が限定されるので、文末共起そのものも頻度が少なくなっているようである。このような考察ができるのも、分類木分析で共起表現を樹形図に描くことができるためであり、コ

一パス研究における分類木分析の有効性を示している。

第9章は、早川ほか(2015)の中国人日本語学習者による精神動詞および物理動詞のニ格とヲ格の付与を紹介している。日本語は膠着語で、主語や目的語の名詞句との格関係は助詞によって示される。一方、中国語は孤立語で、日本語のような格標識(casemarker)を持たない。そのため、中国語のような語順によって文の意味が決定する言語を母語とする日本語学習者には、日本語の格助詞の習得が難しいといわれている。精神動詞は、思考や感情などを対象としており、助詞としてニとヲの両方が用いられるが、日本語学習者にはどちらを選択すればいいのか難しいことがよくある。

第9章では、「作る」や「押す」などの物理的な活動性を示す動詞群と「驚く」や「喜ぶ」などの精神的な活動性を示す動詞群の特性が等質になるように、ターゲットの動詞群の語種、日本語能力試験の配当級、使用頻度、その自然対数変換した数値を計算して、物理および精神動詞の特性に違いがないことを示している。ここでもやはり、研究対象のグループの統制が準備段階として必要であること説明している。その際に、頻度はベキ分布になること、自然対数変換することで正規分布に近づけることができることがグラフで視覚的に示されていて分かり易い。そのような前段階の準備をして、中国人日本語学習者に対し、助詞の「ニ」と「ヲ」の選択テストを行い、決定木分析で学習者の習得状況を検証するという流れを説明している。また、この章では、決定木分析の応用として、ダミー変数を目的変数とする場合は、回帰木分析と分類木分析のどちらも実行できることを紹介している。そして、ダミー変数を使用する場合は、分類木分析のほうが、相対リスクなどの点で回帰木分析よりも結果の解釈が容易であることを説明している。第9章は、この点で決定木分析の方法に深く踏み込んで解説している。

最終章の第10章は、語用論関連の研究で、中国人日本語学習者の助言の難しさを決める諸要因をテーマとしている。この章では、Brown & Levinson(1978, 1987)のポライトネス理論の枠組で、中国人日本語学習者の助言場面における意識と行動に影響する諸要因を考察した研究(黄・玉岡, 2015)を紹介している。ポライトネス理論の紹介から始まり、その枠組みに、「助言により相手を恥ずかしがらせる程度」という心理的変数を追加して、(1)社会的距離、(2)力関係、(3)性差、(4)助言により相手を恥ずかしがらせる程度、(5)文化・言語的背景、の5つを説明変数とし、助言の難しさを目的変数とする回帰木分析を行っている。5つの変数を同時に分析できるのは、まさに回帰木分析の醍醐味である。

黄・玉岡(2015)の回帰木分析のデータは、36の質問に対して、33名が2種類の質問紙

に答えており、延べ 66 名の中国人日本語学習者、67 名の中国人日本語未習者、40 名の日本語母語話者で、36 問×173 名=6,228 データになる。こういうデータ配列が回帰木分析には必要とされる。また、この種の回帰木分析では、データ数が多くなるが、多くの説明変数を同時に分析しても、十分に目的変数を予測することが可能である。分析の結果は非常に興味深く、Brown & Levinson (1978, 1987) のポライトネス理論の枠組みに含まれていない「相手を恥ずかしがらせる程度」という心理的要因が最も強く影響する説明変数となった。もともと、ポライトネス理論は社会学的な観点からの公式であり、心理学的な視点は含まれていないことがより鮮明に回帰木分析の樹形図で示された。

Banerjee & Carrell (1988) が心理的な相手を恥ずかしがらせる程度という指標を提案したのは、Brown & Levinson (1987) の理論が再提案(初回は 1978 年)された 1987 年から 1 年後の 1988 年である。これはポライトネス理論が社会学的観点の主張であることに対する心理学的側面からの批判を込めていたのかもしれない。黄・玉岡 (2015) は、回帰木分析により、Banerjee & Carrell の相手を恥ずかしがらせるという心理的要因が助言の難しさにきわめて強く影響することを示した。そして、Brown & Levinson のポライトネス理論に心理的要因を追加することを提案して、より厳密なフェイス侵害度を見積もる以下のような公式を提示している。

$$W_x = C_x + D(S, H) + P(H, S) + R_x$$

C_x …心理的なコンテキスト(context)「相手を恥ずかしがらせる程度」など

X …言語行為

W_x …フェイス侵害度 (weightiness of the face threatening acts)

S …話し手 (speaker)

H …聞き手 (hearer)

D …社会的距離 (distance)

P …タテの人間関係における聞き手が話し手に及ぼす力 (power)

R …言語行為が特定の社会文化内で持つ負担の度合い (rating of imposition)

黄・玉岡 (2015) の研究は、ポライトネス研究で援用されることの多い Brown & Levinson の理論に新たに重要な要因を追加している点で、画期的なものである。それを可能にしたものこそ、Brown & Levinson や Banerjee & Carrell が活躍した 1970 年代から 1980 年代にはなかった多変量解析の回帰木分析である。

4. おわりに

「決定木分析の欠点は何か?」という質問をする人がいる。決定木分析は、多変量の予測分析なので、それ以外のことはできない。例えば、船は海を移動する手段であり、船に空を飛べというのは無理である。そのため、欠点というより、決定木分析が何をする道具なのかを理解することが、決定木分析をうまく使いこなすコツである。決定木分析は研究の目的としている変数を予測する分析である。その際、予測される変数(予測変数)が質的変数の場合が分類木分析であり、量的変数の場合が回帰木分析である。予測変数を説明する変数(説明変数)は、IBM SPSS Decision Treesを使う限りにおいては、質的変数でも量的変数でもよい。さらに、初期設定のCHAIDを使用すれば、多分岐の樹形図のかたちで結果を示してくれるため、理解し易く、予測変数を構成する説明変数の影響関係と構造を把握することができる。また、仮説に従って、樹形図のはじめの変数を指定することもできる。決定木分析はこうしたことができる道具であることを説明したのが本書である。

[参考文献]

- 池田史子・玉岡賀津雄(2013)「山口方言の特殊モーラを含む語の産出におけるアクセント核の位置に関する世代間比較」『山口県立大学学術情報(共通教育機構紀要)』6, 25-31.
- 黄郁蕾・玉岡賀津雄(2015)「中国人日本語学習者の助言場面における意識と行動に影響する諸要因」『言語文化と日本語教育』48/49合併号, 11-21.
- 早川杏子・玉岡賀津雄・初相娟(2015)「第2言語における動詞特性と格助詞習得の関係: 中国人日本語学習者を対象とした活動動詞におけるニとヲの習得」『Studies in Language Sciences(言語科学)』14, 163-183.
- 林炫情・玉岡賀津雄・宮岡弥生(2011)「否定によって日本語の行為要求疑問文はより丁寧になるのか」『日本学報』86, 143-153.
- Jamila, Mohd(2008)『マレー語母語話者の日本語話体習得に関する縦断的研究—初対面場面における事例研究—』名古屋大学大学院国際言語文化研究科博士論文.
- 滝沢直宏(2002)「文末の接続助詞(+終助詞)と不連続に共起し易い副詞—『から』の場合を中心に—」Makino, Seiichi (ed.), *The Tenth Princeton Japanese Pedagogy Workshop: Proceedings* (pp. 132-142). Princeton, NJ: Princeton University.
- 玉岡賀津雄(2006)「決定木」分析によるコーパス研究の可能性: 副詞と共起する接続助詞

- 「から」「ので」「のに」の文中・文末表現を例に」『自然言語処理』13(2), 169-179.
- 玉岡賀津雄 (2023) 『決定木分析による言語研究』くろしお出版.
- 玉岡賀津雄・張婧禕・牧岡省吾 (2019) 「日本語は無生名詞主語の自動詞文が好まれる言語か?—日本語文の主語となる有生・無生名詞の頻度比較—」『日中言語対照研究論集』21, 124-141.
- 張婧禕 (2017) 「中国人日本語学習者の漢語同形語習得—同形類義語(Overlap語)を中心に—」『愛知工業大学研究報告』52, 6-13.
- 張婧禕・玉岡賀津雄・初相娟 (2018) 「中国人日本語学習者は日本語の漢字の書き取りが正しくできるのか?」『中国語話者のための日本語教育研究』9, 52-68.
- ティモシー・J・バンス・金子恵美子・渡邊靖史【編】(2017) 『連濁の研究: 国立国語研究所プロジェクト論文選集』, 開拓社.
- 李璐・玉岡賀津雄 (2019) 「中国人日本語学習者の間接発話行為の理解—慣習性と習熟度の影響—」『中国語話者のための日本語教育研究』10, 12-28.
- Brown, P., & Levinson, S. (1978). Universals in language usage: Politeness phenomena. In Esther N. Goody (Ed.), *Questions and politeness*, (pp. 56-289). Cambridge: Cambridge University Press.
- Banerjee, J. & Carrell, P. L. (1988). Tuck in your shirt, you squid: Suggestions in ESL. *Language Learning*, 38, 313-347.
- Brown, P., & Levinson, S. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Clancy, P. (1989). Form and function in the acquisition of Korean wh-questions. *Journal of Child Language*, 16(2), 323-347.
- Lyman, B. (1894). *The change from surd to sonant in Japanese compounds*. Oriental Club of Philadelphia.
- Ihara, M., Tamaoka, K., & Murata, T. (2009). Lyman's law effect in Japanese sequential voicing: Questionnaire-based nonword experiments. Collection of the papers selected from the 18th International Congress of Linguists, 1007-1018.
- Ito, N., Tamaoka, K., & Michael P. Mansbridge (2020). A Picture-Book Based Corpus Study on the Acquisition of wh-words in Japanese. *Glottology. International Journal of Theoretical Linguistics*, 10 (1-2), 85-102.

- Kass, G. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29 (2), 119-127.
- Leben, W. (1973). *Suprasegmental phonology*. Doctoral dissertation, MIT.
- McCarthy, J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, 17(2), 207-263.
- Taguchi, N. (2008). Pragmatic comprehension in Japanese as a foreign language. *The Modern Language Journal*, 92 (4), 558-576.
- Tamaoka, K., Lim, H., Miyaoka, Y., & Kiyama, S. (2010). Effects of gender-identity and gender-congruence on levels of politeness among young Japanese and Koreans. *Journal of Asian Pacific Communication*, 20, 23-45.
- Tamaoka, K., Hayakawa, K., & Timothy J. Vance (2016). Triple operations of rendaku processing: Native Chinese and Korean speakers learning Japanese. *Journal of Japanese Linguistics*, 32, 31-55.
- Wagner, M. (1969). *Principles of Operations Research: With Applications to Managerial Decisions*. Prentice-Hall.

[Web 参考サイト]

言語実証プロジェクト: <https://sites.google.com/view/srpls>

統計解析ソフトウェアIBM SPSS Statistics価格表: <https://spss.mediamart.jp/price/>