はじめに　私はなぜ IBM SPSS Decision Trees なのか？

IBM SPSS Statistics のアドオン・モジュールである決定木(decision trees)分析を使った言語研究の方法を紹介します。決定木分析は，音韻，語彙，格助詞，文理解，敬語，ポライトネス，方言，言語習得，日本語教育など多様なテーマに活用することができます。IBM SPSS Decision Trees が，言語研究に有効な理由は以下の6つです。

第1に，決定木分析は，予測する目的変数(従属変数)が質的変数(名義尺度)である場合には分類木分析(classification trees)，量的変数(スケール)である場合は回帰木分析(regression trees)を自動的に選んで実行してくれます。

第2に，決定木分析は，量的変数と質的変数の両方を，目的変数を予測する説明変数(独立変数)として使用することができます。

第3に，決定木分析は，説明変数の予測力を統計検定によって評価してくれます。その際，目的変数に対して，説明変数の値が統計的に等質であると判断された場合は結合し，異質であると判断された場合は保持して，説明変数を最適に分類してくれます。

第4に，決定木分析は，CHAID またはその修正版である Exhaustive CHAID を使用することで，結果の樹形図を多分岐で描くことができます。

第5に，決定木分析は，説明変数で目的変数を予測した結果を，視覚的にわかり易い樹形図で描いてくれます。この樹形図のおかげで，難しい統計解析の手法がわからなくても，直感的に結果を理解することができます。

第6に，決定木分析は，1層目の説明変数を指定することができます。もちろん，有意でない説明変数を指定することは意味がないのですが，理論的に検証したい説明変数を1層目に置くことができるのは，研究目的の全貌を把握するのに非常に有効です。

この本の執筆にあたり，一橋大学の早川杏子先生がテキストの内容を確認してくださり，また宮崎大学の張婧禕先生が章のイラストなどのデザインをしてくださいました。心より感謝申し上げます。

2023年4月26日　名古屋大学にて

玉岡　賀津雄

目次

第1章 決定木分析

主語の有生性と動詞の自他性の関係

決定木分析(けっていぎぶんせき，けっていぼくぶんせき)は，ある目的の事柄(目的変数)があり，それを複数の背景要因(説明変数)で説明したり，あるいは構成する背景要因の構造を把握したりするための解析法です。分析の結果が樹形図で描かれるので，統計の知識がじゅうぶんではなくても，視覚的に結果を把握することができます。とても実用的な手法です。もともと，マーケットリサーチのために開発された解析法ですが，近年，医学，生物学，心理学などさまざまな分野で広く使われています。言語研究にも大いに活用できるはずです。この章では，主語の有生性を予測する研究を例に，決定木分析で描かれる樹形図の見方と名称を紹介します。そして，決定木分析が言語研究に効果的に応用できることを実感していただきます。決定木分析には，予測したい事柄のデータの違いによって，分類木分析と回帰木分析の 2 種類があります。最後に，これら 2 種類の決定木分析の特徴を説明します。

分析練習用のデータのダウンロード
https://tamaoka.org/download/index.html の jyogen.xls です。
注：このデータは分析の練習用に作成したものです。オリジナルのデータとは異なりますが，この章の分析結果と同じになります。

1 決定木分析の目的

決定木分析は，英語では decision tree analysis といわれます。英語の decision が「決定」「決断」「解決」「判断」などと訳されることからもわかるように，なにかを決めることを目的とした分析法です。決定木分析は，もともとマーケティングリサーチの分野で開発されました。企業は，顧客の要求に応える商品やサービスを提供します。そして，それらの情報を顧客に効率的に流し，顧客を満足させることで利益を得ます。その際，どういう商品やサービスが市場で求められているか，どのような人達が商品やサービスを購入する見込みがあるか，さらに商品やサービスに対して顧客がどの程度満足しているのか，これらのことを効率よく把握しなくてはなりません。こうした一連のマーケティングリサーチにおける問いに答えるために決定木分析が活用されてきました。決定木分析は，結果を**樹形図**(dendrogram)で視覚的にわかり易く描いてくれます。そのため，統計の知識がなくても分析の結果が直感的に理解できるきわめて実用的な解析法です。

決定木分析は，予測あるいは説明したい事柄について，それを決める背景となる諸要因を解明することを目的としています。そのため，言語研究でも大いに活用することができるはずです。まず，簡単な例で考えてみましょう。日本語では，人間・動物が主語である場合には「いる」，それ以外には「ある」が用いられ，有生か無生かが区別されています。**有生**(animate)か**無生**(inanimate)かの区別を**有生性**(animacy)と呼びます。「先生が(生徒に)たずねる」の主語の「先生」，「犬が吠える」の主語の「犬」は有生です。一方，「山が聳(そび)える」の主語の「山」は無生の主語です。有生と無生の代わりに，有生物と無生物という言い方をすることもあります。また，これらはそれぞれ有情物と非情物，活動体と不活動体などという言い方をすることもあります。ヨーロッパ言語では，無生物が主語になることが多いといわれます。一方，日本語では，主語になるのは有生の名詞のほうが無生の名詞より多いというのが一般的な理解ではないでしょうか。

私達が典型的だと考えるのは，有生の誰かがなにかを行うという他動詞文でしょう。たとえば，図1-1のように，「ケンジとマサルがカレーを作った。」という文です。主語は「ケンジとマサル」で，主語を示す格助詞であ

第2章

分類木分析

マレーシア人日本語学習者のスピーチレベルシフト

分類木分析（ぶんるいぎぶんせき，ぶんるいぼくぶんせき）は，決定木分析の一種です。ある複数の**説明変数**（独立変数）で名義尺度の目的変数（従属変数）を予測する分析法です。カイ２乗検定の有意水準で樹形図の分岐が決められます。樹形図の分岐をたどることで目的変数を予測する諸要因（説明変数）を視覚的に把握することができます。この章では，マレーシア人日本語学習者が初対面の目上の人と話す際の，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を調べた研究（Jamila, 2008）を紹介します。統計分析ソフト IBM SPSS Decision Trees による普通体と丁寧体を予測する分類木分析の手順を説明します。

分析練習用のデータのダウンロード

https://tamaoka.org/download/index.html の speech.xls

注：このデータセットは，Jamila（2006, 2007, 2008）に印刷されている頻度表から作成しました。

https://tamaoka.org/download/index.html の baishin.xls

注：このデータセットは，日高・伊藤（2007）に印刷されている頻度表から作成しました。

1 言語研究における決定木分析の目的

目的変数が名義尺度である場合に分類木分析を使います。説明変数としては，名義尺度，順序尺度，量的尺度のいずれも使えます。第 1 章では，主語が有生または無生のいずれをとるかという仮説を分類木分析で検証しました。「上がる」「変わる」「始まる」の 3 つの自動詞では，毎日新聞 18 年の記事を調べたところ，出現頻度(以下，頻度)で有生の名詞が主語になることはあまりなく，無生の名詞が主語になる場合がほとんどでした。一方，対応する「上げる」「変える」「始める」の 3 つの他動詞では有生と無生の主語の頻度に違いはみられませんでした。ただし，第 1 章で紹介した 6 つの動詞はあくまで分類木分析を説明するために使用した頻度データです。有生性と自他性との関係については，64 種類の動詞を対象に研究した玉岡・張・牧岡(2019)を参照してください。なお，新聞のサンプル数は約 3 億語で，非常に大きいので，玉岡・張・牧岡(2019)の結果の予測力はじゅうぶんであると思われます。そのため，今後書かれる新聞記事においても，これらの動詞について類似した主語の有生性の頻度傾向を予測することができると思われます。

すでに，決定木分析の全体的なイメージは，第 1 章でつかめたと思います。そこで，第 2 章では，IBM SPSS Statistics のアドオン・ソフトである Decision Trees を使って分類木分析を実際に行ってみましょう。以下では，(1)Excel を使ってどのようにデータを配置するか，(2)SPSS を使ってどういう手順で分析するか，(3)結果をどう解釈するか，を説明します。

2 会話での丁寧体と普通体の使用

分類木分析を行う前に，どのような仮説を証明しようとしているかを把握しておくことが大切です。そこで，まずデータ収集の理論的な背景について説明します。

日本語には体系的な敬語表現が存在します(菊池, 1997)。日本語の会話では，同じ内容であっても，対人関係，場面，状況，目的などに応じて表現を変えることがあります。「見る」と表現するにしても，尊敬語の「ご覧になる」，謙譲語の「拝見する」，丁寧語の「見ます」と，いろいろな表現があり

第3章

回帰木分析

行為要求表現における丁寧度の変化

回帰木分析（かいきぎぶんせき，かいきぼくぶんせき）は決定木分析の一種で，複数の説明変数（独立変数）で量的尺度の目的変数（従属変数）を予測する解析法です。回帰木分析では，*F* 検定で平均の差を比較して，樹形図を分岐します。日本語の行為要求表現は，平叙形の肯定の「～てくれる」よりも否定の「～てくれない」のほうが，さらに疑問形でも肯定の「～てもらえますか」より否定の「～てもらえませんか」のほうがより丁寧だといわれています。この章では，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主観的判断で検討した研究（林・玉岡・宮岡, 2011）を例に，SPSS の Decision Trees を使って回帰木分析の手順を解説します。

分析練習用のデータのダウンロード

https://tamaoka.org/download/index.html の polite.xls

注：このデータは分析練習用に作成したものです。オリジナルのデータとは異なりますが，この章の分析結果と同じになります。

1 行為要求表現と丁寧度の関係

はじめに，この章の回帰木分析に使うデータの背景について説明します。相手になんらかの行動を起こさせることを目的とする**行為要求表現**(国立国語研究所, 1992)は，「話して」「話しなさい」と直接に要求する形式から，「話してくれる」「話していただけませんか」といった疑問形で間接的に要求する形式まで多様です。表現の形式としては，授受の補助動詞「くれる／ください」や「もらう／いただく」を含んだ表現が頻繁に使用されます(岡本, 2007)。これらは，「くれる」＜「もらう」＜「くださる」＜「いただく」と相手の許可を求める際の丁寧度も変化すると予想されます('＜'は丁寧度の順位を示す)。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たずねることで，より丁寧さが増すといわれています(岡本, 1986, 2000; Okamoto, 1992)。さらに，日本語の要求表現は，「～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ます。その理由としては，依頼の相手に対して選択の余地を与えることで，心理的負担を軽減し，より丁寧な印象を与えるからだと説明されています(菊池, 1997; 岡本, 2007)。これらの先行研究(菊池, 1997; 岡本, 1986, 2000; Okamoto, 1992)から，日本語の行為要求表現において，否定や疑問などの言語特性が丁寧度を上げる要因となっているかどうかを実証的に検討したのが，林・玉岡・宮岡(2011)の研究です。この章では，林ほか(2011)の研究で使われたデータを基に，回帰木分析の手順を説明します。

2 行為要求表現の丁寧度の測定

依頼される側である聞き手が感じる丁寧度は，どのように測定するのでしょうか。表3-1に示したように，林ほか(2011)は，平叙形の「～てくれる」「～てもらえる」から，疑問形の「～てもらえますか」「～てくださいますか」「～ていただけますか」という丁寧度が異なると予想される行為要求表現を設定しています。さらに，これらの5つの表現を「～てくれない」「～てもらえない」「～てもらえませんか」「～てくださいませんか」「～ていた