

# 日本語教員支援用「MyClass 分析支援ツール」の設計と開発

中尾 桂子  
大妻女子大学 短期大学部

森下 淳也  
神戸大学 国際文化学研究所

本研究は、毎学期の日本語教師の授業の振り返りを簡易にできるような小規模事例の量的分析用支援ツールを設計、開発するものである。その目的は、当該担当の学生の記述分析に基づいて得られた知見を、その該当学生に還元するという分析の本来の目的を支援することにある。本稿では、WEB上に公開するサイトに、指定された方法でアップした記述を、統計処理ソフトにかけるために表の形式に調整して出力するという設計と、その開発の方向性について述べるものである。

## A design and development of the "MyClass analysis supportive tool" for a Japanese language teacher

Keiko Nakao  
Junior College Division  
Otsuma women's University

Jyun-ya Morishita  
of Intercultural Studies  
Kobe University Graduate School

This research designs the supportive tool for quantitative analysis of a small-scale example, which a Japanese teacher can look back on his lesson easily. The purpose is to support originally intended analysis as returning the knowledge acquired from the student's description to the student. This paper describes a design and the directivity of the development, which the student's description is adjusted and outputted by raising in the specified method to the exhibited site on WEB.

### 1. まえがき

言語教育では、教師が、学習者への指導のあり方、達成度、学習者が抱える問題の傾向を確認するために、学習者の産出した記述や発話、また、その発話が生じる環境としての授業を分析する[1]。本来、その目的は、その学期中に該当学生に指導内容等を還元することにあるが、実際には、日々の授業の準備、実施、提出結果の採点・評価で終始してしまい、教師が指導自体を評価することや指導学生の傾向を学期中に分析するのは十分に行えないことも多い。

もし、学習者の分析が簡易にできるのであれば、教師自身の振り返りを授業に還元しやすくなると期待できる。

そこで、本研究では、毎学期の日本語教師の授業の振り返りを簡易にできるような小規模事例の量的分析用支援ツールを設計、開発する。その目的は、当該担当の学生の記述分析に基づいて得られた知見を、その該当学生に還元するという分析の本来の目的を支援することにある。

### 2. 学習結果の振り返り

日本語教育では、他の言語教育同様、学習後に学習者の達成状況を評価する。日本語教育における評価では、どのレベルでも、基本的に、作文や口頭発表等での産出を調べ、文法的に正確な形で

使用されているか、すなわち、学習した語彙が的確な意味と形で使用されているかを見ることから始まる。もちろん、初級から上級に学習者の能力が上がるにつれて、文法形式が1語どうしの関係から、連語、語の共起から1文全体における適切生、ついで、2文以上の関係、さらに、段落どうし、章節を伴った文章構成へと、確認する範囲が広がるが、基本的には、語の単位での形と意味の適切性を確認することになる[2]。

一方、評価ではなく、学習者の習得状況や授業内容、活動方法などを分析するような場合、学習者の適切性を評価するよりは、使用語彙の傾向に差があるか、傾向として顕著なスタイルがあるかという点を確認することが多くなる。誤用が多く見られる形式の有無や、誤用ではなくとも、過剰な使用が見られるか、反対に、過少使用の傾向が認められるかといった母語話者との差の有無や差の程度、さらには、学習の前後での学習者の産出能力の変化や、条件の違いで生じる学習者の産出物への影響などを調べるが多くなる。

また、習得研究における産出結果からの情報と学習環境の情報は、1つずつ質的に分析していく分析か、量的に分析するものかで、おおむね、2つのタイプに分けられる[3-4]。質的な分析も、量的な分析も、分析対象である言語形式、構成など、着眼点は、ほぼ同様で、ただ、分析の手法が異なるという場合も多い。そのため、昨今では、量的に傾向をふまえて、質的に使用状況の背景や原因

を考察するという二段階を経た分析も増えている[5].

習得研究における量的な分析の場合は、学習者の産出した結果そのものの分析だけでなく、学習者の学習環境との関係で、学習者の産出結果を分析することも多い[6].

学習者の産出結果自体から導きだされるものとは、具体的には、文字(拍)数、語彙数(異なり語数、延べ語数)、文数、段落数、文字種別の数、出現位置、使用文型、構文パターン、品詞別比率、共起語のパターンといったデータ内部から導きだされる形あるものの数である。これらの数は、種類を示す「異なり語数」や全数を示す「述べ語数」の数とともに、粗頻度そのものと、粗頻度を100万語辺りの割合に調整した頻度など、基本的には数字に置き換えて利用される。

また、学習者の学習環境に関する情報とは、産出物の外部にある学習者を取り巻く関連情報のことである。具体的には、学習経歴、母語、学習年数、成績、クラス人数、目標、動機などである。

習得研究の一環で行われる授業分析は、学習者の達成状況を評価するや授業目標自体を再検討することよりは、どの部分を重点的に指導すべきか、弱点や注意点はどのようなものかを具体的に探すことが目的となる場合が多い。そして、習得研究は、次の年度の学生にも当てはまるかどうか判断できない事例研究を、通事的や横断的に繰返していくことで一般化される事象を見出すような流れで行なわれることが多い。

単体の事例研究は、汎用性が低い場合も多いが、分析対象の学生に還元される場合には、その限りではなく、十分に意味があるものとなる。そのため、学期中に概略的にでもよいので、習得状況の傾向を還元できるのであれば、学生にとっても、研究のために記述を提供する甲斐があると考えられる。

### 3. 量的研究の視点

学期の途中などで、折に触れて、学習者が持つクラスとしての傾向を調べ、かつ、具体的にどのような部分に問題があるかをクラス単位の概略として探すことを考えた場合、分析は、質的なものではなく、量的な手法で行なうのが適当である。

量的分析の手法や分析観点は、もちろん目的により異なるが、今回は、分析対象を、学習結果の記述とし、テキスト分析で、ごく一般的な分析ができる程度であることを優先する。

習得研究における分析では、まず、差の有無や関係の有無を判断するようなものがある。たとえば日本語母語話者と日本語学習者の記述における特定の種類の語の使用率の差[7]や、ある教授法を実施したクラスと実施していないクラスの達成度の差[8]など、複数のものの間にどの程度の差があるかを分析するもの、また、その差が性

別や学習年数などとの関係があるかを分析するようなものである。統計的には、検定、相関を調べるものとなる。

また、これらの差や関係性の有無について、具体的に、その程度を明らかにしたい場合は、モデルを設定し、そのモデルとの差の程度を見るという回帰分析を行なうことになる。例えば、教師か教師でないかにより、学習者の産出に対する評価ポイントが異なるかという問題[9]や、母語話者と学習者の言語使用上の意識の違いを見るような場合があげられる[10].

その他、特定のグループが同様かどうかについてはクラスター分析を行なって確認することもある[11]. また、ある特定の語の使用状況がグループ毎に異なるかどうかを視覚的に明示するような場合にはコレスポンデンス分析が考えられる[12].

言語習得における授業実践等の分析で扱われる量的研究の観点として、以上のような先行研究に代表されるようなものが散見されることから、今回のツールにて実行する量的分析の手法は、品詞や特定の語どうしがテキスト毎にどのような傾向を示すかを調べるものとし、統計的手法として想定されるものは「相関分析」「回帰分析」「判別分析」「クラスター分析」「コレスポンデンス分析」とする。

### 4. ツールの概要と入出力ファイル

本ツールが行う処理は次のようなものである。Web上のサイトに記述データファイルをアップロードすると、ファイルがサーバに転送され、テキスト部分の形態素解析(Chasen[12])処理後にデータベース(以下、DB)に格納する。そして、DBに対して指定した条件に基づく形式を結果としてブラウザ上に返す(図1)。

出力の際の条件を何も指定しなければ、基本情報として①全語彙リストが返される。これはChasenの出力結果のうち第2層までの品詞情報に従い、自立語、付属語別に、各語の粗頻度を降順でリストにしたCSV形式での出力である(表1)。

出力形式の条件を指定した場合は、条件毎の特定の形式に語彙数が整理された結果が返される。

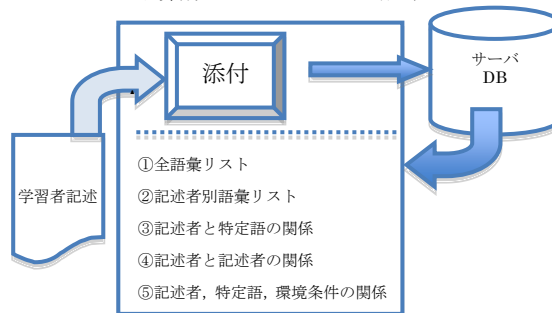


図1 基本の処理の流れ

この条件の指定は、アップロードする添付ファイル内の記述で行なう。すなわち、「###」で始まる行に「メタデータ」という文字列がある場合、その行の「##」で区切られた文字列を、学習者の環境を表すメタデータとして、語彙リストとともにメモリ上に保有し、出力時の指定形式に応じて利用するようにしている。

したがって、使用の手順は次のようになる。

- ①分析対象の記述を全てテキスト形式の1ファイルにまとめる
- ②分析対象の各記述の開始位置に「###メタデータ」、次行に「###本文」と書く
- ③「###メタデータ」の行に「##」で区切って記述者情報を入れる
- ④開始位置の次行の「###本文」で始まる行に、出力結果として指定したい条件を書く
- ⑤ファイルをアップロードする

ファイルがアップロードされると、サーバに取り込まれる。テキスト部分を形態素解析し、その結果とあわせて、ファイルの先頭から順に自動的にふった記述テキストIDと、記述された分のメタデータがDBに保有する。IDは、開始位置の「###メタデータ」から、次に現れる「###メタデータ」の直前までを、一記述者の記述テキストとして認識している。

なお、「###本文」という行に書く条件に応じてメタデータが調整されるのであるが、出力結果として指定できる条件は、今回は、Chasen出力の結果で特定できる単一の語に限っている。図2は、条件を指定してアップロードする際の学習者の記述例である。

###メタデータ##女性##2###79##留学生  
 ###本文##である。##です。##だ。##…  
 学生にとってアルバイトはそれほど重要なものではない。なぜなら、学生にとって最も優先すべきことは勉強することであるからだ。ただし、社会に出る前に社会の仕組みを学ぶという目的で考えれば、アルバイトも

図2 学習者記述ファイルの例

アップロードされるファイルは、複数の記述者別のテキストが「###メタデータ」で区切られて1ファイルに記載されていなければならない。

また、処理の速度と1クラスあたりの学習者の数を考え、今回は記述者テキストの上限を100人分、ファイル容量を1Mとしておく。さらに、学習者の記述内の誤用は、形態素解析処理結果に影響することから、今回、テキストは全て修正済みであることを前提としている。

出力する結果は、次の5種である。①全語彙リスト、②記述者別の語彙リスト、③記述者と特定語を関係づけたもの、または、④記述者どうしを関連づけたものか、それに⑤記述者に関する環境情報(メタ情報)を加味して関係づけたものかである。

①と②はテキスト中の全語彙リストを、アップロード全体か、記述者別かに分けたものであるが、③はクラスター分析、相関分析、検定、コレスポネンス分析に対応できる形式で、④は判別分析、⑤は、環境情報を加味して、語彙頻度を調整し、重回帰分析を行なうことが可能な形式に対応させて調整し、語彙とその頻度(100万語あたりの調整頻度)で整理した形式で出力される。それぞれの出力結果の形式は、表1~5のようになる。

表1 全語彙リスト例(一部)

名詞	9003	サ変名詞	3700	形容動詞	917
女性	119	意味	186	必要	31
人気	117	喫煙	96	様々	30
言葉	116	存在	72	非常	27
作品	103	生活	61	温暖	26
若者	102	影響	59	重要	25
自分	89	使用	48	危険	24

表2 記述者別語彙リスト例(一部)


表3 記述者と特定語の関係例(一部)

連体詞	留学生	日本人学生	論文
あの	78	13	87

ある	310	139	261
こうい	52	13	5
この	362	328	966
そのような	26	13	19

表 4 記述者と記述者の関係例 (一部)

両群記述者	しかし	…	確かに	群
0001	78	…	13	1
0002	310	…	139	1
0003	52	…	13	1
0004	362	…	328	…
0005	26	…	13	2

表 5 記述者・特定語・環境条件の関係例 (一部)

記述者	句点	読点	性別
0001	78	13	女
0002	310	139	男
0003	52	13	男
0004	362	328	女
0005	26	13	男

## 5. 出力ファイルの応用

4章で述べた出力までが、本ツールの支援の範囲で、この後の統計処理は、それぞれの利用者が行なうことを想定している。

ここでは、フリーの統計マクロ Seagull Stat2003[13]と、福山平成大学の福井正康氏が作成した社会システム分析+αプログラム College Analysis Ver. 5.0[14]を利用して、本支援ツールで出力された結果を統計的に分析した結果例を報告する。

まず一つ目は、平成 21 年度に 0 短期大学部でアカデミック・ライティングの指導を受けた学生 60 人の記述の均一性について調べたものである。

「学生のアルバイトの是非」に関する意見文毎に、その中の語彙情報から、異なり語数 (Type)、異なり語数の延べ語数中の割合 (Type Token Ratio)、名詞、動詞、副詞、接続詞、段落、文、句読点の 9 項目の数を記述能力の判断指標と位置づけ、表の形式に整理する。その表に、学年を群として追加し、判別分析用に整理した表を作成する。これを Seagull Stat2003 で分析したところ、次の表 6, 7, 図 3 のような結果が得られた。

表 6 判別結果

前 \ 後	1 群	2 群	正判別率
1 群	23	7	76.7%
2 群	9	21	70.0%
総合			73.3%

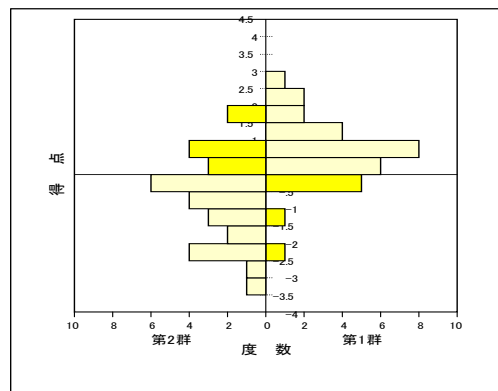


図 3 線形判別関数による得点度数

表 7 判別関数係数

1 群～2 群線形判別関数の係数

変数	1 群-2 群
異なり語	0.167882
TTR	-0.03978
段落数	0.047956
名(名, サ, B, C, )	-0.17076
動詞(動, B)	-0.46296
副詞(副, 可, B)	-0.34602
接続詞	0.449785
句点	-0.23952
読点	-0.05448
定数項	3.847675
マハラビス D <sup>2</sup>	1.264926
誤判別率	0.286941

二つ目は、文脈の結束性を表す語である指示語を指標に比較して、文脈の結束性からみた学術性の有無や差を確認した例である。

比較対象は、文系学部 1, 2 年次の留学生の日本事情を調査する課題として提出された学期末レポート 30 本、文系学部 1 年次の日本人学生が文学の授業で提出した学期末レポート 75 本、さらに、都内の大学図書館に寄贈された研究科論文集内の文系院生の論文 15 本、CiNii[15]でフリーアクセスになっている文系学術論文 100 本のテキストとしている。これら 4 種のグループの各記述毎に、それぞれ「こそあど」の含まれる指示語を取り出し、100 万語辺りの調整頻度に整理した表 (表 9) を作成する。これを利用して、College Analysis Ver. 5.0 で分析したところ、図 4 のようなクラスタ分析の図が確認された。

表9 指示語 (PMW)

指示語	留学生	日本人学生	院生	論文
あの	78	13	68	87
ある	310	139	217	261
こういう	52	13	8	5
ここ	362	328	652	966
こちら	26	13	17	19
この	3643	3046	3981	3719
これ	1060	1271		1426
これら	155	378		517
こんな	130	38		43
そういう	52	265		36

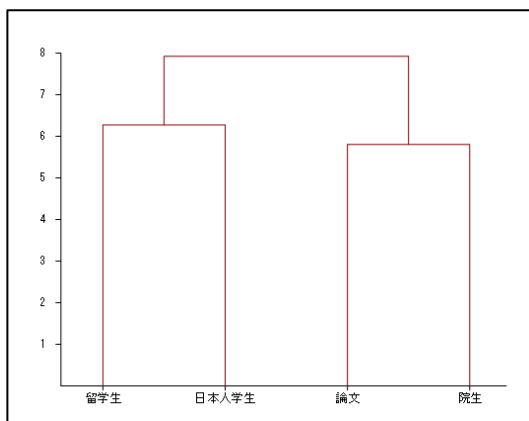


図4 4つのテキストの学術性の差

## 6. 支援ツールとしての可能性と限界

5章では、条件を指定して出力した表形式に整理された結果を、フリーの統計ソフトを用いて分析した例を2点紹介した。

1点目の場合、Chasen 出力の結果で特定できる単一の語を品詞と、全語数から計算する異なり語数と延べ語数を利用している。異なり語は全体語彙リストで出力される各語の頻度数のことで、それと延べ語数で計算した結果である TTR は百分率で表される数字である。

出力結果の条件指定は、基本的には語として形式が明確なものや、品詞情報で指定することを想定していたが、実際には、このほかに、記述データの量的分析には、その記述データの基本情報として考えられるような、総語彙数等の情報が必要だということに気がついた。例えば、5章の例で利用した「異なり語数」、「延べ語数」「TTR」、「文数」、「段落数」などである。そして、これらを利用して頻度データを計算できると、分析にはより便利だと考えられた。

DB 格納時に自動的に計算して保有しておくことは難しくないのだから、必要かどうかとは別に、使いそうな情報を、記述データをアップロードす

る際に、裏であらかじめ計算しておくのとよいと考える。

また、DB を利用することからすると、記述データをアップロードするときに指定する条件に、検索のための SQL を記述して条件を指定することも考えられる。この方法は応用例であり、本ツール開発の目的である「簡易」分析とは相容れない方法であるが、応用の際の一つの可能性を示していると考え、基本情報の計算とあわせて、サイトの使用案内に、別枠の応用使用例として注記しておきたい。

さらに、今回、統計的な手法を想定した量的分析のための出力結果を念頭に置き、出力結果を、①全語彙リスト、②記述者別の語彙リスト、③記述者と特定語を関係づけたもの、または、④記述者どうしを関連づけたものか、それに⑤記述者に関する環境情報(メタ情報)を加味して関係づけた5種として、①～⑤のように名付けたが、やはり、統計の知識に関係なく、用語の正確性には配慮すべきではないかと考える。説明を付けるのは冗長であるが、何も書かないのも誤解を招く恐れがある。また、ツールの仕組みを理解していなくても、データアレンジの限界についての予測が立つような説明も必要だと考えることによる。ただし、どのような説明や例示方法がよいのかについては今後の課題として検討していきたい。

5章で例示した分析のように、今回の出力結果の利用方法確認のため、フリーソフトを利用したが、表形式に調整した出力結果であるため、ケースと変数の位置づけ、目的として分析目的も合わせて考えれば、出力の結果を利用者がアレンジすることにより、さらに、様々な利用法へと応用が可能だと考える。

ただし、何を持って「一般的」と言えるかは問題があり、どのような「一般的」な統計手法を支援するのが妥当かについては検討する必要がある。今回のツールに於ける条件指定は、ごく基本的なもので、さほどの充実した出力結果が備わっているわけではない。したがって、基本的な使用範囲、応用の際の考え方、応用の範囲を明記すべきだが、まだそれ程多くの例を検証していない。今回の例示の他の応用例についても説明できるように、様々な分析を試しておきたいが、これも今後の課題である。

## 7. あとがき

本研究は、学期中の日本語教師が、各自の授業の振り返りのために小規模事例の量的分析を簡易に行なえるように支援するツールを設計するというものであった。目的は、当該担当の学生の記述分析に基づいて得られた知見を、その該当学生に還元するという学期中の分析を簡便にすることで、学習者分析の本来の目的を支援することである。

今回は、WEB上に公開するサイトに、指定された方法でアップした記述を、統計処理ソフトにかけるために表の形式に調整して出力するという設計で開発したが、ごく基本的な下処理までは可能であることが確認できた。

しかしながら、授業実践の振り返りの観点として、平均的なものが想定しきれていない問題、ならびに、形態的に明確で、検索が極力簡単に行なえる語以外への対応には不適合であることから、それほど強力な支援ツールではないため、依然、考慮すべき問題も多く残っている。今後の課題である。

## 参考文献

- 1) 新版日本語教育辞典, 文野峯子: 授業研究, 8 言語習得・教授法, 812l-813r (2005) .
- 2) 新版日本語教育辞典, 縫部義憲: 授業時の評価, 8 言語習得・教授法:795r-797l (2005) .
- 3) 新版日本語教育辞典, 迫田久美子: 量的研究, 7 言語・言語教育研究の方法:613l-614l (2005) .
- 4) 新版日本語教育辞典, 箕浦康子: 質的研究, 7 言語・言語教育研究の方法:614l-615l (2005) .
- 5) 原田三千代: 作文の変化にピア・レスポンスがどのように関わったか—中級日本語学習者の場合—, 小出記念日本語教育研究会論文集, 15, pp55-69, (2007).
- 6) 尹喜貞: 授受補助動詞の習得に日本語能力, 及び学習環境が与える影響—韓国語学習者を対象に, 日本語教育(130), 120-129, (2006).
- 7) 伊集院郁子, 高橋圭子: 日本語の意見文に用いられる文末もダリティ—日本・中国・韓国語母語話者の比較—, 東京外国語大学留学生日本語教区センター論集, 36:13-27, (2010) .
- 8) 田中信之: ピア・レスポンスが推敲作文に及ぼす影響—分析方法とフィードバックの教示に注目して—, アカデミック・ジャパニーズ・ジャーナル, 3, pp.9-20, (2011).
- 9) 一二三朋子: 非母語話者との会話における母語話者の言語面と意識面との特徴及び両者の関連—日本語ボランティア教師の場合—, 教育心理学研究, 47:490-500, (1999) .
- 10) 飯田香織・玉岡賀津雄・初相娟: 中国人日本語学習者の音象徴語の理解, 日中言語研究と日本語教育, 第5号:46-54, (2011).
- 11) 平成 19-21 年度科学研究費補助金研究「PAC 分析法を活用した学習者が日本語教材から受ける影響と学習者要因の解明」(課題番号:19520449).
- 12) 阿部新: 言語学習ピリーフの多変量解析—調査項目のクラスター分析と数量化Ⅲ類—, 名古屋外国語大学外国語学部紀要(39), 71-90, (2010).
- 13) Chasen<<http://chasen.naist.jp/>>
- 14) 石川慎一郎・前田忠彦・山崎誠: 言語研究のための統計入門, くろしお出版(2011).
- 15) 社会システム分析+αプログラム  
CollegeAnalysisVer.5.0<<http://www.heisei-u.ac.jp/ba/fukui/analysis.html>>
- 16) CiNii<<http://ci.nii.ac.jp/>>