

テキストマイニングを用いて「振り返り」を可視化する試み

田中 雅人¹⁾

Visualization of reflection using the method of text mining

Masato Tanaka¹

Key words: text mining, reflection, physical play, active child program

**(Bulletin of Department of Physical Education, Faculty of Education,
Ehime University, 11, 17 - 22, March, 2019)**

キーワード：テキストマイニング、振り返り、運動遊び
アクティブ・チャイルド・プログラム

1. はじめに

自由記述形式での回答や面接記録のようなテキスト型データの分析には、数量的なデータを解析する際に使用する統計的手法を用いることは難しい。テキスト型データを分析する方法のひとつにテキストマイニングがある。テキストマイニングとは、自由記述のような通常の文章からなるデータを定量的に扱うための分析方法で、結果を視覚的に表現することでデータの解釈を容易できるという特徴がある。なお、コンピュータを用いたテキスト型データの計量的分析には、2つの異なるアプローチがある。ひとつは、分析者が作成した基準にしたがって言葉や文書を分類する方法であり、もうひとつは、頻繁に同じ文書の中にあられる言葉のグループを多変量解析によって自動的に発見・分類する方法である。前者は、Dictionary-based アプローチ、後者は、Correlational アプローチと呼ばれている。Dictionary-based アプローチは、テキスト型データの様々な側面に自由に焦点を絞れるという利点があるが、意図のないしは無意識のうちに、理論や仮説に都合の良いコーディング規則ばかりが作成・利用さ

れてしまう危険性を含んでいる。一方、Correlational アプローチでは、客観性に関わる問題は解決されるが、理論や問題意識を自由に操作化し追究することは困難である(樋口, 2004)。そこで、樋口(2004)は、この2つのアプローチを統合する分析用システムとして「KH Coder」を開発している。また、樋口(2017)は、語の分析、コードの分析、テキストの検索と閲覧が「KH Coder」の主要な機能であり、語の分析を行ってから、必要に応じてコードの分析を行いつつ、常にもとのテキストを検索・閲覧するというのが計量テキスト分析の基本的な進め方であると述べている。

野口ほか(2018)は、大学の授業科目である「アウトドアレクリエーション」の受講生を対象に、テキストマイニングによる振り返りの分析を行った。5日間の振り返りシートの記述内容をテキストデータとし、「KH Coder」を用いて共起ネットワークを作成したところ、ネットワーク構造の中で中心的な役割を果たしている語が実習中に変化していることが示された。また、田中(2018)は、「KH Coder」を用いて野外活動(スキーとロングライド)後の振り返りに関する記述を分析した。その結果、階層的クラスター分析を行うことにより、野外活動の意義や目的を知る手掛かりを得ることができた。また、共起ネットワークを作成することによって、野外活動に対する意識が経験や

1) 愛媛大学教育学部
〒790-8577 愛媛県松山市文京町3番

1. Faculty of Education, Ehime University,
Bunkyo-cho 3, Matsuyama-shi, Ehime,
〒790-8577, Japan

性別によって異なることが明らかとなり、自由記述形式で得られたテキストデータを計量分析することは有効な方法であると報告している。

本研究では、テキストマイニングを用いて、アクティブ・チャイルド・プログラムに指導スタッフとして参加した学生の「振り返り」の可視化を試み、その有効性について検討する。

2. 方法

2-1. 「ACP北条」の概要

近年、子どもの基礎的運動能力の低下が問題となっている。アクティブ・チャイルド・プログラムは、子どもの運動能力、体力の向上とスポーツの普及を目的に、公益財団法人日本スポーツ協会が開発した運動遊びのプログラムである。

2016年～2018年に、松山市文化・スポーツ振興事業団主催、愛媛大学総合型地域スポーツクラブ後援で、「アクティブ・チャイルド・プログラム in 北条(以下、「ACP北条」とする)」を年1回、計3回開催した。対象は小学1年生～4年生で、1回の開催につき、約100名が参加した。アイスブレイクを行ったのち、参加者は、5つのグループに分かれて、5つの運動遊びを順に実施した。各プログラムの実施時間は、約20分であった。3回の「ACP北条」で実施した運動遊び、および各運動遊びで身に付くと考えられる体力要素を表1に示した。

指導スタッフとして参加した学生は、3～4人でチームを作り、各運動遊びの指導を担当した。なお、「ACP北条」への参加を通して、スポーツや健康に関する課題に関心を持ち、主体的に課題解決に向かって、理論と指導を連関させた取り組みができること、多世代にわたる対人関係能力およびコミュニケーション能力を獲得し、豊かな地域社会づくりに貢献するための適切な行動ができることを目指した。

2-1. 分析の対象

「ACP北条」に指導スタッフとして参加した教育学部スポーツ健康科学課程スポーツ指導者養成コース4年生、男28名、女17名、計45名の「振り返り」の記述を分析の対象とした。

2-3. 分析の手順

1) 分析ツール

テキストマイニングには、「KH Coder 3」を使用した。「KH Coder 3」の特徴は、「茶筌(形態素解析)」

「MySQL(データベース)」、「R(統計解析, グラフ理論による分析と視覚化)」といったソフトウェアで構成されたフリーソフトウェアであること、また、探索的な分析だけでなく、仮説検証的な分析も可能なことである(牛澤, 2018)。

表1. 運動遊びで身に付く体力要素

No.	運動遊び	体力要素				
		筋力	瞬発力	筋持久力	全身持久力	調整力
1	ねずみ逃がし		○			○
2	ところんオニ		○			○
3	魚とり		○			○
4	ボール集め競争		○			○
5	どかーんジャンケン		○			○
6	氷オニ		○		○	○
7	しっぽとり		○		○	○
8	ひよこの闘い	○				○
9	スリーチーム綱引き	○		○		○
10	大根抜き	○		○		○

2) テキストデータの作成

「ACP北条」の実施後に提出された自由記述による「振り返り」から、ACPに関する部分をテキストデータとして抽出した。なお、「ACP北条」に参加した学生は、愛媛大学総合型地域スポーツクラブのスポーツ教室の指導スタッフとしても活動している。そこで、子ども(小学生以下)を対象としたスポーツ教室と大人(中学生以上)を対象としたスポーツ教室との分類を外部変数とし「G1:子ども, G2:大人」の2群に区分した。G1は26名、G2は19名であった。抽出したテキストデータと外部変数は、Excel(Microsoft社)を用いて入力し、入力終了後にCSV形式で保存した。

3) データの編集

a) 複合語の検出

ひとつの名詞が連続する複数の名詞に分割される場合(例えば、「運動」と「遊び」)には、ひとつの名詞(「運動遊び」として扱う必要がある。そこで、「茶筌」を用いて複合語を検出したのち、ひとつの名詞として扱う語のリストを作成、CSV形式で保存し、分析の際にはこのファイルを読み込んで使用した。

b) 未使用語の検出

分析に使用しない語のリストを作成、CSV形式で保存し、分析の際にはこのファイルを読み込んで使用した。

4) 分析の手続き

分析は、田中(2018)が野外活動に関する記述に対して実施したテキストマイニングの手法を用い、階層的クラスター分析、共起ネットワーク、対応分析を行った。

a) 語の抽出

名詞、動詞、形容詞の3つの品詞を分析の対象とした。なお、「KH Coder 3」では固有名詞、組織名、人名、地名として区別して抽出される語も名詞として扱った。各品詞別に出現回数が高かった順に語を抽出して分析の対象とした。

b) 階層的クラスタ分析

階層的クラスタ分析を実施し、デンドログラムを作成した。この分析では、出現パターンの似通った語の組み合わせが明らかになる。

c) 共起ネットワーク

語と語の関連性を視覚化するために、共起ネットワークを作成した。なお、共起ネットワークでは、対象語の出現回数が円の大きさと、共起性・関連性の強さが線の太さで表現される。また、外部変数を用いて、性別、スポーツ教室での指導対象による違いを分析した。さらに、キーとなる語がどのような文脈で使用されているのかを分析するため、KWIC (Key Words in Context) コンコーダンス検索を実施し、コロケーション統計を行った。

d) 対応分析

クロス集計の結果を視覚化するために対応分析を行った。なお、クロス集計には、外部変数を用い、性別とスポーツ教室での指導対象の違いによって抽出語がどのように異なるのかを示した。

3. 結果と考察

3-1. 語の抽出

対象となる記述(文)の数は849で、ひとりの平均記述数は18.9であった。出現した異なり語数は1,398、出現回数の平均値は5.6 (SD=24.37)であった。

3-2. 階層的クラスタ分析

出現回数20以上の名詞、動詞、形容詞を選択したところ、62語が対象となった。階層的クラスタ分析を行ったところ、図1に示したデンドログラムが作成され、8つのクラスターに分割された。出現パターンの似通った語のまとまりを概観すると、「運動遊びの活動内容」、「運動遊びで身に付く動き」、「運動遊びの楽しさ」、「子どもの反応や感情」、「指導の工夫や伝え方」、「子どもへの対応の難しさ」、「スタッフとしてのイベントへの関わり」であった。これらは、運動遊びの実施や運動遊びが子どもに与える影響に関する語と指導方法や子どもへの関わり方に関する語に大別された。なお、クラスターの数はい任意に設定でき

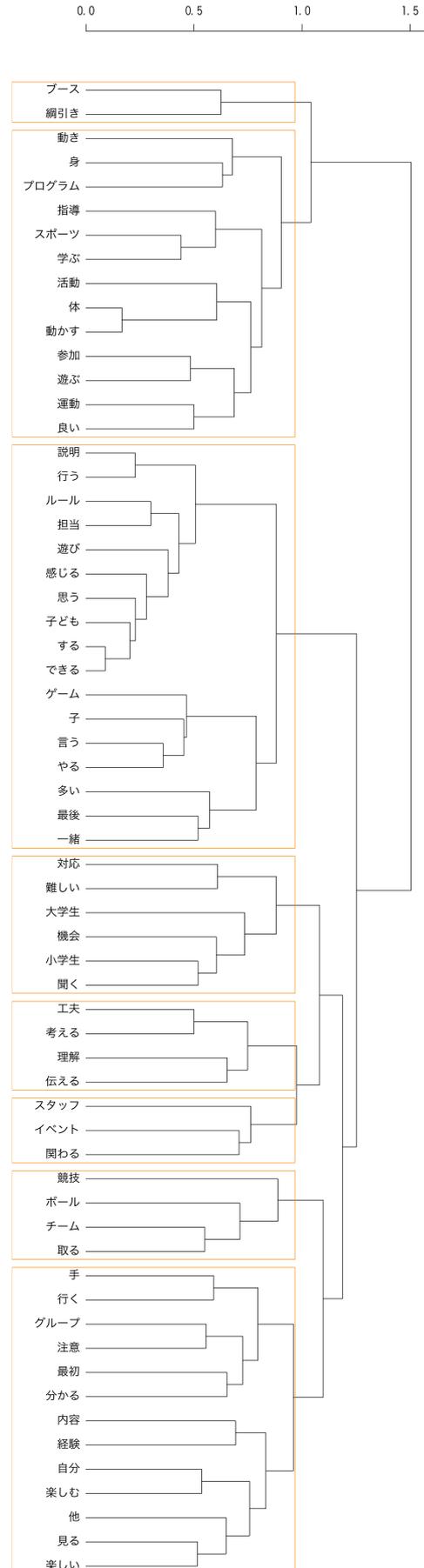


図1. 階層的クラスタ分析

「できる」と関連する名詞は、多くはなかったが、「理解」, 「交流」, 「参加」等があり, このイベントへの参加が, 子どもたちとの交流や子どもの理解につながったことがうかがわれる。

表1-1. 「する」と関連する名詞

説明 (41)	活動 (11)	行動 (6)	集中 (4)
参加 (38)	対応 (11)	実施 (6)	設定 (4)
担当 (32)	終了 (10)	衝突 (6)	追加 (4)
理解 (30)	工夫 (9)	提供 (6)	配慮 (4)
指導 (24)	運動 (7)	開催 (5)	判断 (4)
注意 (14)	協力 (7)	獲得 (4)	誘導 (4)
タッチ (13)	実践 (7)	観察 (4)	用意 (4)
意識 (12)	進行 (7)	緊張 (4)	

(出現数)

表1-2. 「できる」と関連する名詞

理解 (7)	活動 (2)	加勢 (1)	実感 (1)
交流 (3)	実践 (2)	確認 (1)	実施 (1)
参加 (3)	説明 (2)	還元 (1)	成長 (1)
アプローチ (2)	復活 (2)	吸収 (1)	体験 (1)
プレー (2)	イメージ (1)	行動 (1)	対応 (1)
運動 (2)	チャレンジ (1)	指導 (1)	

(出現数)

また, アクティブ・チャイルド・プログラムに関係する語「遊び」, 「子ども」を node word とし, その前後5語以内に出現した形容詞を抽出した(表3)。なお, 抽出した語は, node word と語との結びつきの強さを表すスコアが1.00を超えるものとした。コロケーション統計の結果, 「遊び」については, 「親子」, 「伝承」, 「外」, 「良い」, 「楽しい」という語のスコアが高く, 実施した運動遊びが伝承遊びや外遊びであり, 親子で行うものもあったことを反映していた。また, 遊びが楽しくて, 良い運動であったことを示していた。「子ども」については, 「体力」, 「運動」, 「親」, 「遊び」のスコアが高く, 運動遊びを行うことによって, 子どもの体力向上が期待されていることが示された。

3-1. 「遊び」に関連する語

親子 (4.12)	ルール (2.08)	担当 (1.82)	学校 (1.20)
プログラム (4.00)	一緒 (2.08)	遊び (1.60)	活動 (1.20)
伝承 (2.50)	提供 (2.08)	身体 (1.58)	小学生 (1.03)
外 (2.33)	良い (2.00)	パート (1.50)	
子ども (2.25)	楽しい (1.85)	運動 (1.33)	

(スコア)

3-2. 「子ども」に関連する語

体力 (4.67)	大人 (1.83)	他 (1.50)	少ない (1.25)
運動 (2.78)	ルール (1.73)	低い (1.50)	チーム (1.20)
親 (2.45)	多い (1.70)	ゲーム (1.45)	一緒 (1.20)
遊び (2.25)	子ども (1.67)	自分 (1.33)	片方 (1.20)
活動 (1.88)	対応 (1.67)	理解 (1.32)	楽しい (1.12)
外 (1.83)	身体 (1.50)	高い (1.25)	機会 (1.07)

(スコア)

3-4. 対応分析

出現回数15以上の名詞, 動詞, 形容詞, 100語を対象語として, 性別 (m・f) とスポーツ教室での指導対象 (G1・G2) を外部変数とした対応分析を行った(図3)。性別と指導対象のクロス集計では, 普段, 子どもを指導対象としている学生では, 「動き」, 「遊び」, 「体力」, 「楽しい」といった運動遊びの楽しさや効果に関わる語が近くにプロットされた。一方, 大人を指導対象としている学生では, 「指導」, 「伝える」, 「工夫」, 「注意」などの語が近くにプロットされ, 子どもに運動遊びを指導する経験が少ないことから, 指導に対する戸惑いがあり, 子どもへの接し方や指導方法を工夫していることが明らかとなった。

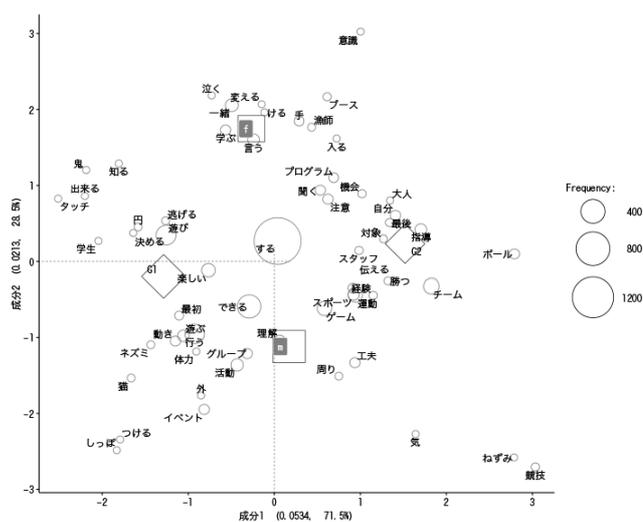


図3. 対応分析 (性別×指導対象)

4. まとめ

本研究では, テキストマイニングを用いてアクティブ・チャイルド・プログラムに参加したのちの「振り返り」に関する記述を分析した。なお, テキストマイニングには, 形態素解析, データベース, 統計解析, グラフ理論による分析と視覚化が可能なフリーソフト

ウェアである「KH Coder」を用いた。結果は、以下の通りである。

名詞、動詞、形容詞の3つの品詞を分析の対象とし、各品詞別に出現回数が高かった語を抽出したのち、1) 階層的クラスタ分析、2) 共起ネットワーク、3) KWICコンコーダンス検索、4) 対応分析を行った。いずれも、「振り返り」のレポートで記述された情報を視覚的に理解できる有効な方法であった。階層的クラスタ分析により、「振り返り」の内容は、運動遊びの実施方法や運動遊びが子どもに与える影響に関する語と指導方法や子どもへの関わり方に関する語に大別されることが明らかとなった。また、共起ネットワークを作成することによって、子どもの運動遊びに対する感情、遊びを通して身体を動かすことの楽しさ、遊びのやり方を子どもに説明することの難しさなどが存在することが示された。また、普段のスポーツ教室で指導を行っている対象によって、現れる特徴語が異なることが明らかとなった。さらに、対応分析を行った結果、指導対象が異なることによって、関連する語が異なることが示された。子どもを対象としたスポーツ教室で指導を行っている学生は、運動遊びの楽しさが、大人を対象としている学生は、運動遊びを指導する際の戸惑いを示していた。自由記述形式で得られたテキストデータを計量分析し視覚化することは、「振り返り」の内容を客観的に示す上で、有効な方法であると考えられる。

文献

- 樋口耕一 (2017) 言語研究の分野における KH Coder 活用の可能性. 計量国語学, 31 (1) : 36-45.
- 樋口耕一 (2014) 社会調査のための計量テキスト分析 -内容分析の継承と発展を目指して-. ナカニシヤ出版: 東京.
- 樋口耕一 (2011) 計量テキスト分析の提案と必要なソフトウェアの開発. ソシオロジ, 55 (3) : 102-108.
- 樋口耕一 (2004) テキスト型データの計量的分析 -2つのアプローチの峻別と統合-. 理論と方法 19 (1) : 101-115.
- 星野崇宏・荘島宏二郎・樋口耕一・富田英司 (2016) 教育心理学研究のためのテキストデータの計量分析. 教育心理学年報, 55 : 313-321.
- 野口和行・村山光義・村松 憲・板垣悦子・東海林祐子 (2018) シーズンスポーツ「アウトドアレクリエーション」受講者のライフスキルの獲得 -ふりかえりの記述による質的検討-. 慶應義塾大学体育研究所, 57-1 : 19-31.
- 佐藤郁哉 (2011) 質的データ分析法. 新曜社: 東京.
- 田中雅人 (2018) 計量テキスト分析を用いた野外活動の振り返りの分析. 愛媛大学教育学部保健体育紀要, 10 : 37-43.
- 牛澤賢二 (2018) やってみよう テキストマイニング. 朝倉書店: 東京.