

# ルールベース手法による 動作分析テキストからのPBPUの自動抽出

Automatic extraction of PBPU from movement analysis text using rule-based methods

櫛 力輔<sup>\*1</sup>      松下 光範<sup>\*1</sup>      堀 寛史<sup>\*2</sup>  
Rikisuke Ichii      Mitsunori Matsushita      Hirofumi Hori

<sup>\*1</sup>関西大学      <sup>\*2</sup>びわこリハビリテーション専門職大学  
Kansai University      Biwako Professional University of Rehabilitation

Movement analysis is an essential examination in planning physiotherapy. During this examination, the physiotherapist evaluates the patient's condition based on their experience and describes the test results as text. The text describing the examination outcome contains practical knowledge based on experience. Extracting and relating such knowledge to each other will help ensure objectivity and share best practices, thereby assisting physiotherapists in conducting movement analysis. However, these texts often contain linguistic and semantic ambiguities, making it difficult to extract knowledge uniformly by computer. A previous study defined the smallest unit of knowledge that constitutes physiotherapy (PBPU; Problem-Based Physiotherapy Unit). It contributes to organizing the movement analysis texts logically; however, extracting PBPU from the text requires enormous work time. In this study, we attempted to extract PBPU using a rule-based method to address the issue. As a result, we demonstrated that about half of the PBPU can be automatically extracted.

## 1. はじめに

理学療法とは、運動や機能障害の評価や治療を行うことで生活する上での身体機能の回復や維持、予防を目指す医学的リハビリテーションに位置付けられる。昨今の高齢社会において、高齢者が自立した生活を送るために、理学療法を提供する理学療法士の育成が重要視されている。

理学療法プログラムの立案において重要な検査のひとつに動作分析が挙げられる [木村 06]。動作分析は患者の動作を観察し、分析する検査手法であり、動作の特徴や構造、関与する筋肉や関節の動きなどを明確に捉え、その動作を改善することを目指す。理学療法士はこの検査により患者の動作上の機能不全や力学・生理学的な機能不全の特性を把握し、それらの因果関係を推測することで対処が必要な問題点を抽出する。動作分析は視覚的な観察によって実施されるため、理学療法士の主観に依存する部分が多く [木村 06]、その質は理学療法士の技能によって左右される。

動作分析の実施にあたっては、養成校で学ぶことのできる知識のみでなく、臨床経験によって会得される実践的な知識（以下、実践知と記す）を有していなければ、良質な理学療法を提供することは困難であり、経験の浅い理学療法士はその技能の低さ故に適切な評価や判断を行えない可能性がある指摘されている [堀 18]。技能の差を埋め、経験年数による理学療法の質の差を解消するためには、実践知を明示化していき、経験に左右されることのない論理的な思考に基づいた動作分析の方法論を確立する必要がある。そこで本研究では、動作分析の際に作成される文章（以下、動作分析テキストと記す）に着目する。理学療法士は自らの経験に基づいて患者の動作を観察するため、作成された動作分析テキストには、経験に基づく知識が多く含まれる。そのような知識を表出し、それらの知識同士を関係付けることが客観性の担保やベストプラクティスの共有につながり、経験の浅い理学療法士が動作分析を行う際の一助と

原文	PBPU
腸腰筋や大腿四頭筋の過緊張による、 並列表現 ↑ 股関節伸展制限が考えられる。	腸腰筋の過緊張が生じている 大腿四頭筋の過緊張が生じている 股関節伸展制限が考えられる
股関節は伸展位となっており、 主語① 述語① 立脚期を通して、ダブルニーアクションは、 確認できない 述語② 主語②	股関節は伸展位となっている 立脚期を通してダブルニーアクションは確認できない

図 1: PBPU の例（文献 [宮本 22] より引用し一部改変）

なることが期待される。

ただし、動作分析テキストは自然言語によって記述されているために、言語的・意味的あいまいさが多分に含まれ、計算機による知識の画一的な表出は困難である。この問題を解決することを企図して、宮本ら [宮本 22] は知識の表出を目指し、動作分析における知識の最小単位を「問題思考を基盤とした理学療法の構成素（Problem Based Physiotherapy Unit：以下 PBPU と記す）」と定義し、それを動作分析テキストから抽出することを試みた（図 1 右欄参照）。この手法では、得られた PBPU に観察・推測・教科書・予後・方針・診断・判断・その他といった 8 つのカテゴリを付与し、それらを因果関係に基づいて紐づけ、その因果関係をエッジとするネットワーク表現で可視化した。このネットワークを経験豊富な理学療法士に提示して定性的な評価を行った結果、記述者の実践知が豊富だと判断された動作分析には観察・推測・治療などの PBPU が満遍なく出現し、観察から推測・治療へと因果関係を示すエッジが適切に存在することが確認された。この結果は、生成された可視化表現が理学療法士の観察能力や論理構成力を把握する上で有効であることを示唆している。

しかし、宮本らの手法では PBPU を抽出する作業を人手で行う必要があるため、膨大な作業時間を要する。本研究ではこの課題を解決するために PBPU を動作分析テキストから機械的に抽出することを試みる。

連絡先: 櫛 力輔, 関西大学総合情報学部, 〒 569-1095, 大阪府高槻市霊山寺町 2-1-1, Tel:072-690-2437, Fax:072-690-2491,k244081@kansai-u.ac.jp

足関節や膝関節の動きが正常から逸脱しているために、  
足部のクリアランスが低下している



### 副詞節・並列節による分割

- ・足関節や膝関節の動きが正常から逸脱しているために、
- ・足部のクリアランスが低下している



### 文末表現の修正

- ・足関節や膝関節の動きが正常から逸脱している
- ・足部のクリアランスが低下している



### 名詞句の並列表現の分解

- ・足関節の動きが正常から逸脱している
- ・膝関節の動きが正常から逸脱している
- ・足部のクリアランスが低下している

図 2: PBPU 抽出の流れ

## 2. 提案手法

PBPU はタイミング・身体部位・状態 (項目)・状態 (程度) の 4 項目 \*1 で構成される知識単位である。このうち状態 (項目) は必須であるが、その他は欠落する場合もある。動作分析テキストから人手で PBPU を抽出する際には、文中に含まれる上記の 4 項目を単純に取り出すだけでなく、図 1 左欄のように主語・述語が複数含まれる重文や並列表現を伴う文を解釈し、ユーザである理学療法士が容易に理解することが可能な形式である単文に変換する。例えば、「腸腰筋や大腿四頭筋の過緊張による股関節伸展制限が考えられる」という文では、まず述語となる「による」といった語に着目することで「腸腰筋や大腿四頭筋の過緊張が生じている」「股関節伸展制限が考えられる」に分ける。その後、並列表現を表す接続詞「や」に着目することで「腸腰筋の過緊張が生じている」「大腿四頭筋の過緊張が生じている」を抽出する。ここでは「大腿四頭筋」「腸腰筋」が身体部位の記述にあたり、「過緊張」が状態 (項目) の記述にあたる。PBPU は抽出された後、その PBPU に記述されている内容の性質から 8 つのカテゴリに分類される。宮本らによって観察・推測といったカテゴリが全 PBPU のおよそ 8 割を占めていることが明らかとなっており [宮本 22]、これらのカテゴリの自動抽出を可能とすることが作業時間の短縮に大きく寄与すると考える。そのため、本研究での自動抽出の対象は観察・推測カテゴリの PBPU に絞ることとする。

### 2.1 手法の選定

重文や並列表現を含む動作分析テキストから PBPU のような知識単位を抽出する手段として、機械学習を用いる抽出手法とルールベースを用いる抽出手法の 2 つが存在する。

機械学習手法のひとつである教師あり学習は学習データとしてあらかじめ入力データに対する正解データを用意し、それを機械に学習させることで未知の入力データに対して抽出を可能にする手法である。近年、機械学習手法は多くの自然言語処理タスクで成果を上げている。しかし、この手法はその性質

\*1 文献 [宮本 22] では、状態 (項目) と状態 (程度) を状態、状態-value とそれぞれ呼称している。

表 1: 分割・文末表現修正ルール

表記 (依存関係ラベル)	修正後
こと (obl) から (case)	
ため (obl)	
為 (obl)	
に (case) 対し (fixed)	
が (mark)	
で (cop) あり (fixed)	である
て (mark) おり (fixed)	ている
さ (aux) せ (aux) ず (aux)	させない
せ (aux) ず (aux)	しない
おら (fixed) ず (aux)	いない
ず (aux)	ない
ない (aux)	ない
られ (aux)	られる
さ (aux) れ (aux)	される
わ (advcl) れ (aux)	われる
さ (aux) せ (aux)	させる
なり (advcl)	なる

上、大規模なデータを学習データとして使用することができなければ十分な精度を得ることが困難である。

ルールベース手法は文から分割点などを人手で特定し、それを分割する手順を定め、その手順に従って機械的に抽出するものである。利点としては、データ規模が大きくなるとも実装が可能であることが挙げられる。その一方で、ルールが煩雑になってしまう恐れもある。

本研究で扱うデータセットは動作分析データが 43 件、含まれる PBPU は 932unit であり、機械学習手法に適用するには小規模なものである。このような小規模データを学習データとして利用するため、Wei らはテキスト内で用いられている単語を同義語に置き換えることで新たなテキストを生成する手法やテキスト内から隣り合った 2 単語を任意に選び出し、その単語間に無作為に選ばれた単語を挿入することで新たなテキストを生成する手法といった水増し手法を提案している [Wei 19]。しかし、水増し手法を適用すると実際の理学療法士が作成したデータでないものがデータセットに混入してしまい、データセット自体の信頼性が損なわれてしまう恐れがある。本研究が扱うドメインの理学療法は医療分野のひとつであり、医療分野ではエビデンスに基づいた治療や説明が求められるため、データセットの信頼性を損なうことは避けるべきである。また、機械学習手法によって出力されたものがどのような過程を経て出力されたのかを説明するのは困難であり、これもエビデンスに基づいた治療や説明といったものを阻害する恐れがある。そのため、本研究において機械学習による手法は適さないと考え、ルールベース手法を採用した。

### 2.2 手順

本研究が想定する流れを図 2 に示し、具体例とともに述べる。PBPU を自動抽出するにあたって、PBPU は、多くの場合で動作分析テキストの 1 文から抽出されるという性質があるため、最初に動作分析テキストを文単位に分割する。次に、1 文から PBPU を抽出するための単文化を行うために、主部とそれに対応する述部を特定する必要がある。

#### 2.2.1 述部の特定

難解な文章に対して、文の区切れとなる副詞節や並列節といった文節で文を分割することで簡単な文構造に変換し、

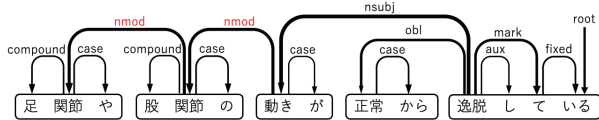


図 3: 名詞句並列の依存関係ラベルの例

読み手にわかりやすい文章を作成する研究が存在している [加藤 20][大野 20]。これらの研究と本研究で行う単文化は類似性が高いと考え、本研究においても述部の特定を行う際、副詞節・並列節といった文の区切れとなる箇所に着目し、構文解析によって得ることができる表記・依存関係などの情報から表 1 のように分割ルールを制定した。構文解析器には日本語自然言語処理ライブラリである GiNZA(ver. 5.1.2)\*<sup>2</sup> を用いた。また、動作分析テキストには専門用語が頻出するため、構文解析の精度が低下されることが懸念される。そのため、ユーザ辞書として、相良らが医療記録文の分かち書きを支援する目的で公開している ComeJisyo [相良 19] と先行研究で作成された理学療法辞書 [宮本 22] を用いた。ComeJisyo は MeCab 形式であり、そのままでは GiNZA で使用することができないため、GiNZA で使用されているパーサーである Sudachi\*<sup>3</sup> 形式に書き換えた。GiNZA で使用することのできる文節 API\*<sup>4</sup> を用いて文節単位に文を区切り、文節内に該当する箇所があった場合に分割ルールによる処理を適用した。分割箇所の欄には GiNZA による構文解析で文が形態素ごとに分けられたときに得られる表記とその形態素に付与される依存関係ラベルの情報を示した。表記以外に依存関係ラベルを使用する理由は、単純な文字列一致では分割箇所でない場所に同じ文字列が存在する場合に誤って分割されてしまうからである。それを回避するため、依存関係ラベルの情報を分割ルールに含めることで対処した。修正後の欄には分割した箇所をどのように修正するのかを示した。空欄であるものは分割箇所前の文節が終止形となっている文が多かったため、修正を行わず、そのまま切り捨てることとした。

図 2 の「足関節や膝関節の動きが正常から逸脱しているために、足部のクリアランスが低下している」という文では、「ため」という部分が副詞節・並列節の部分にあたり、分割ルールが適用され、この箇所が分割される。このとき、単に文を分割するのみでは、理学療法士が作成された PBPU を読む際に分かりにくい文になることが懸念されるため、分割箇所に基づくルールを適用することで文末表現を適切なものに修正する。前述の例の場合、「ために」という部分を切り捨てることで「～している」という終止形となる。ここまでの処理を行うことで、述部の特定が可能となる。

### 2.2.2 主部の特定

提案手法では、述部を特定した後、主部の特定を行う。主部の特定を行うためには、名詞句の並列表現を探し出し、それらを分解する必要がある。GiNZA では浅原ら [浅原 19] が作成したコーパスに基づいて名詞句の並列表現に図 3 のように nmod という依存関係ラベルが付与されることになっている。図 3 の例では、「足関節」と「股関節」が名詞句の並列表現にあたり、それぞれを結ぶようにして nmod ラベルが付与されている。これが付与されたものを目印とすることで名詞句の並列表現を分解する。まず、並列表現の開始位置を特定する。

文頭から nmod ラベルが付与されている形態素（以降、nmod 形態素と記す）を探索し、1 つ目の nmod 形態素から文頭へ向かって、複合語を示す compound ラベルが付与されていない形態素を探索し、その形態素を開始位置とする。文頭まで達した場合は文頭を開始位置とする。図 2 では、文頭の「足関節」が nmod 形態素にあたるため、開始位置は文頭となる。次に、並列となっている要素位置を特定する。文頭から nmod 形態素を順に探索していき、それぞれの要素番号を配列に格納する。この際、「といった」「という」「の」といった文字列が nmod 形態素の間に含まれる場合、分解すると PBPU としての意味が失われてしまうものがあつた。例えば、「股関節の動き」には「股関節」と「動き」が名詞句の並列表現のため nmod ラベルが付与されるが、分解すると PBPU としての意味が失われてしまう。そのため、「といった」「という」「の」といった文字列が含まれる場合は分解をせず、そのままの形とした。図 2 では、「足関節」、「股関節」が要素位置となる。

次に並列表現の終了位置を特定する。文頭から nmod 形態素の依存関係連先の形態素が nmod 形態素でない形態素まで探索し、その次の形態素を並列表現の終了位置とする。図 2 では「が」がこれにあたる。分解の手順は文頭から並列表現の開始位置までの文字列に並列となっている前の要素番号から次の要素番号までの形態素を接続する。このようにして「足関節の動きが正常から逸脱している」「膝関節の動きが正常から逸脱している」「足部のクリアランスが低下している」といった 3 つの PBPU が抽出される。以上の流れで PBPU を自動で抽出する。

## 3. データセット

本研究では、入力データとして理学療法士が記述した動作分析テキストデータを使用し、自動抽出された PBPU を評価するデータとして動作分析データを人手で PBPU 形式に構造化したものを用いた。本章ではそれぞれのデータに関して述べる。

### 3.1 動作分析テキストデータ

先行研究では、理学療法士に 3 症例（症例 A：脳卒中片麻痺男性患者、症例 B：脳卒中片麻痺女性患者、症例 C：パーキンソン病女性患者）の歩行・立ち上がり動画のうち 1 つの動画を提示し、動作分析を行ってもらい、動作分析テキストデータを計 42 件（症例 A：15 件、症例 B：14 件、症例 C：13 件）を取得した。対象者の年齢幅は 20 代前半から 40 代前半であり、臨床経験年数の幅は 2 年から 20 年以上であった。取得した動作分析テキスト 42 件全体の平均文字数は 577.3 文字、標準偏差は 243.5 文字であった。

### 3.2 構造化データ

自動抽出の精度評価を行うために、基準となる PBPU を人手で作成した。以下に作成手順を示す。まず、3.1 節で得られたデータを文単位に分割した。次に、それらが PBPU となるように単文化を行い、その上で観察・推測・教科書・予後・方針・診断・判断・その他といった 8 つのカテゴリを付与した。その後、PBPU のタイミング・身体部位・状態（項目）・状態（程度）の 4 項目に該当する各語句にアノテーションを行った。自動抽出の対象を観察・推測カテゴリに絞ることから、観察・推測カテゴリが付与された PBPU を列挙し、それらを構造化データとした。構造化データ作成後、経験豊富な理学療法士の精査を通じて、妥当性の担保を行った。

\*2 <https://megagonlabs.github.io/ginza>

\*3 <https://github.com/WorksApplications/Sudachi>

\*4 [https://megagonlabs.github.io/ginza/bunsetu\\_api.html](https://megagonlabs.github.io/ginza/bunsetu_api.html)

表 2: 抽出一致率

症例	完全一致率	部分一致率
A	45.0%	66.4%
B	52.9%	66.5%
C	43.0%	62.0%
全体	47.3%	65.3%

#### 4. 評価と考察

提案手法による自動抽出の精度を評価するために、人手で抽出した PBPU との比較を行った。評価は、理学療法士が記述した動作分析の 1 文単位で行った。評価の基準は文から自動抽出した PBPU と人手で作成した PBPU のタイミング・身体部位・状態（項目）・状態（程度）に該当する単語に相違があるかとした。動作分析テキストの 1 文から抽出すべき全ての PBPU が抽出された場合は完全一致、1 つでも抽出できていた場合は部分一致とした。結果を表 2 に示す。完全一致率の平均は 47.3%、部分一致率の平均は 65.3%であった。最も抽出精度が高かったのは、症例 B であり、一方で最も低かったのは症例 C であった。

副詞節・並列節での分割によって抽出が成功した例として、「遊脚期後期から踵接地時に足関節背屈不足であり、踵接地時には足底から接地しているため、足関節背屈制限が考えられる」という文が挙げられる。この文からは、「遊脚期後期から踵接地時に足関節背屈不足である」「踵接地時には足底から接地している」「足関節背屈制限が考えられる」という 3 つの PBPU が抽出されるべきであり、提案手法によってこれらを正しく抽出できることが確認できた。副詞節や並列節での分割が正しく行われたことで、これらの PBPU が抽出可能であったと考えられる。

副詞節・並列節での分割によって抽出が失敗した例として、「現状の伸展パターンでは立ち上がりにも困難感があることが予測でき、トランスファーが自立できず、生活範囲の狭小化も予測できる」という文が挙げられる。この文からは、「現状の伸展パターンでは立ち上がりにも困難感がある」「トランスファーが自立できない」「生活範囲の狭小化が予測できる」という 3 つの PBPU が抽出されるべきである。しかし、「現状の伸展パターンでは立ち上がりにも困難感があることが予測でき、トランスファーが自立できない」と本来 2 つの PBPU になるものが 1 つの PBPU として抽出されていた。正しく抽出するには「予測でき」の部分で文を分割する必要があるが、これに対する分割ルールを定めていなかったため、正しく抽出が行われなかった。

並列表現の分解によって抽出が成功した例として、「矢状面においては左立脚中期から後期にかけて体幹前傾と膝の過伸展が見られる」という文が挙げられる。この文からは、「矢状面においては左立脚中期から後期にかけて体幹前傾が見られる」「矢状面においては左立脚中期から後期にかけて膝の過伸展が見られる」という 2 つの PBPU が抽出されるべきであり、提案手法によってこれらを正しく抽出できることが確認できた。

並列表現の分解によって抽出が失敗した例として「麻痺側下肢に着目すると遊脚時に股関節外転、外旋、膝関節伸展位であるいわゆる分回し歩行と言われる歩様である」という文が挙げられる。この文からは、「遊脚時に股関節外転位である」「遊脚時に股関節外旋位である」「遊脚時に膝関節伸展位である」といった PBPU が抽出されるべきである。しかし、この文は複雑な入れ子構造で並列表現の部分が記述されており、これらの

係り受けが難解であるために適切に抽出できなかったと考えられる。

#### 5. おわりに

本研究では動作分析における知識の最小単位である PBPU を自動抽出するためのルールベース手法を提案した。抽出精度を評価した結果、半数程度の PBPU が自動抽出可能であることを示した。抽出結果の考察から分割ルールの見直しや複雑な入れ子構造の記述に対する対応の検討が必要であると考えられる。複雑な入れ子構造を用いて記述された文は計算機による構造の把握が困難であるだけでなく、人間が読んだ際にも分かりにくい文となっていることが予想される。そのため、その動作分析を記述した理学療法士にフィードバックを行い、わかりやすく文章を改善するように促す仕組みが必要だと考える。今後は、自動抽出の精度を向上させるため、文章の分割・並列処理ルールの改良や専門辞書の作成を進めるとともに文章のわかりやすさを改善させるためのフィードバック機構の実現を目指す。

#### 参考文献

- [Wei 19] Wei, J. and Zou, K.: EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks, in *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 6382–6388 (2019)
- [加藤 20] 加藤 汰一, 宮田 玲, 立見 みどり, 佐藤 理史: 文化財説明文を対象とした平易化支援システムの設計と実装, 人工知能学会全国大会 (第 34 回) 論文集, 4Rin1-49 (2020)
- [宮本 22] 宮本 誠人, 松下 光範, 高岡 良行, 堀 寛史: 理学療法初学者の支援を目的とした動作分析テキストの構造の可視化, 人工知能学会全国大会 (第 36 回) 論文集, 111-OS-6-04 (2022)
- [浅原 19] 浅原 正幸, 金山 博, 宮尾 祐介, 田中 貴秋, 大村 舞, 村脇 有吾, 松本 裕治: Universal Dependencies 日本語コーパス, 自然言語処理, Vol. 26, No. 1, pp. 3–36 (2019)
- [相良 19] 相良 かおる, 山崎 誠, 麻子 軒, 東条 佳奈, 小野 正子, 内山 清子: 実践医療用語の語構成要素—意味を基準とした分割, じんもんこん 2019 論文集, pp. 57–64 (2019)
- [大野 20] 大野 博之, 稲積 宏誠: 修飾節に着目した長文における短文化への試み, 情報処理学会研究報告, Vol. 2020-CE-154, No. 12, pp. 1–6 (2020)
- [堀 18] 堀 寛史: 科学的根拠と技能: 理学療法哲学試論, 臨床哲学, Vol. 19, pp. 45–63 (2018)
- [木村 06] 木村 貞治: 理学療法における動作分析の現状と今後の課題, 理学療法学, Vol. 33, No. 7, pp. 396–403 (2006)