

# 抄録からのキーワード抽出時におけるバイアスの検討

佐川輝高 岡田ルリ子 青木光子  
愛媛県立医療技術大学 保健科学部

**背景と目的：**昨年度までの研究で、抄録のキーワード抽出とそれに続くチャート化を行う場合、候補単語の共起度を基準にし共起分解度数などを用いることでチャート化が可能だった抄録がある一方、ある程度のチャート化は可能であったが完全では無かったものも存在し、全ての抄録の主張をチャート化することが可能ではなかった。今回は全ての抄録でその主張がチャート化されることを目標に、その方法の検討を行った。

**方法：**

## 1. 素材

モデル文章は本学学生の看護研究抄録を用い、コントロールとして授業中に出された課題に対し学生から提出された文章を用いた。

## 2. 解析手順

WinCha2000R2 を使用し形態素解析を行った。ノイズ除去後、各キーワード候補単語の全文での出現度数を基本度数として、一文中での任意2単語組合せの共起度を求めた。この単語組合せを分解し、それぞれの単語に共起度を持ち越し、その単語の全文での持ち越し共起度を合計、これを共起分解度数として用いた。

## 3. バイアスの検討

題名、目的、方法、結果、考察のそれぞれに表われるキーワード候補単語にバイアス度数（整数）を設定し共起分解度数にこれを掛け合わせ、その時の上位キーワード候補単語により抄録の内容をチャート化することが可能か否を検討した。

**結果：**全候補単語の度数初期値を1とした場合、つまりバイアスを掛けない、共起分解度数のみに頼ったチャート化では全ての抄録をチャート化することは出来なかった。題名・目的に表れるキーワード候補単語のバイアス度を5とすると、全ての抄録で適切なチャート化が可能であった。今回調べた抄録ではこのバイアス度数が5未満でも6以上でも全ての抄録をチャート化することは出来なくなった。

**考察：**今回、抄録内容検討において抄録の題名・目的・考察から総合的に判断するという方法を行ったので、これを反映している結果と言えるのかもしれない。抄録中、その主張が一文中にまとまっている場合、大部分はまとまっているが重要な単語が抜けている場合、まとまっていない場合の3パターンが存在した。現在、これが文書解析にどのような影響を与えるか更に検討を加えている。私達が文章から総合的に判断した抄録の内容をチャート化することは、今回、題名・目的から得られるキーワード候補単語にバイアス度数を設定することで可能となったが、著者の主張からは外れるが抄録に秘められた潜在的因果関係を持つキーワード候補単語の抽出にはバイアス度数を掛け合わせない生の共起分解度数の方が適当であるなど、抄録の意味あるチャート化には更なる検討の必要がある。