

特許出願におけるリーダビリティと被引用数の関係

朴峻基

テキストの読みやすさ、読みにくさをリーダビリティという。これは文の長さや複雑さなどに依存する。現在主に判定に用いられる方法は文の内容ではなく、形式、長さ、音節、使われる単語の難しさなどで判定する。本研究においても上記のものを判断材料に読みやすさ、読みにくさを判定していく。対象は日本の特許出願の本文とする。特許文書の読みやすさと、その特許文書の被引用数の関係を明らかにすることが、本研究の目的である。企業や研究所において、出願あるいは保持する特許の価値を評価することは非常に重要な課題である。特許の代表的な評価方法のひとつに引用分析があり、被引用数が多い特許ほど価値が高いとされる。革新をもたらした発明の特許は被引用数が多いという報告もある。しかしながら、文献の引用は、発明・発見の内容だけでなく、形式上の要因（ページ数、図表数、参考文献数など）によっても左右されることが知られている。本研究では、外形的要因の中でも、これまであまり調査されてこなかったリーダビリティに特に注目して、被引用数との関係を分析する。被引用数へのリーダビリティの影響を明らかにすることは、その影響を除外して、発明内容と被引用数との関係を把握することにつながられる点で意義があると考えられる。

1993年公開の特許出願から抽出した800件のデータを利用した。リーダビリティ・リサーチ・ラボが提供するツールを利用し、特許出願の本文のリーダビリティを算出した。800件の出願は、国際特許分類(International Patent Classification)で分けられた八つのセクション(分野)から、それぞれ100件ずつ抽出した。それら800件の出願について、被引用数とリーダビリティとの関係をウィルコクソンの順位和検定をもって調査した。具体的には、被引用数の多い群のリーダビリティの平均値と被引用数が少ない群のリーダビリティの平均値に違いがあるかどうかについてを検定を行った。本研究では、複数の閾値によりデータ群に分けて、その平均値を比較した。データ全体だけではなく、各セクション別の調査も行い、各セクションごとの特徴も発見した。

結果として、各セクションについて、単語の羅列の頻度、専門的な単語の属性などに関する特徴を発見することができた。ただし、それぞれのデータ群の間に有意差は認められなかった。その原因として、データ数の少なさ、本文の長さのばらつき、純粋な文章ではなく単なる単語の羅列も特別に扱うことなくリーダビリティを算出したためのリーダビリティの減少などが考えられる。

指導教員 芳鐘冬樹