

## コーパスとクラスター分析を用いた副詞の文体調査

中俣尚己

本研究では、日本語の副詞の文体差について、クラスター分析を用いて可視化し、日本語教育などの実用面で利用できる分類方法を提案することを目的とする。

手法としては、教科書やコーパスの調査を通じて初級で導入すべき副詞を選定した島崎英香 (2015) 「初級日本語学習者のための副詞 90 語の選定—日本語母語話者の副詞の使用実態を通して—」(『日本語／日本語教育研究』6) を基準に、語の文体が異なる類義語を加え、『現代日本語書き言葉均衡コーパス』(BCCWJ) の頻度を参考に 164 語を選定した。続いて、BCCWJ で検索を行い、文書ジャンルごとの pmw を計算、さらに副詞間の頻度差をならすことで、単純な粗頻度の影響を排除した。このデータに対して階層的クラスター分析を行った。

結果、6 クラスターに分割することが妥当とされたが、うち 2 クラスターは含まれる語が少なく、元となるデータの歪みが反映されたものであると考えられるため、4 クラスターを考察の対象とする。この 4 つは文体について「あらたまり—くだけ」と「かたい—やわらかい」の 2 軸で論じるべきであると主張した石黒圭 (2015) 「書き言葉・話し言葉」と「硬さ／軟らかさ」(『日本語学』34-1) によって下表のように位置づけられる。

	多いレジスター	語数	例	文体の特性	論説文での使用
C1	知恵袋, ブログ	46	あまり, ちょっと, かなり	あらたまり・軟	不可
C2	国会会議録	10	だいたい, 若干, ぜひ	あらたまり・硬	?
C3	文学	61	やがて, しばらく, もっと	くだけ・軟	主観部のみ可
C4	特徴なし	42	最も, ほぼ, 特に	くだけ・硬	客観部でも可