

言語資源研究系  
萌芽・発掘型共同研究プロジェクト

テキストの多様性を捉える  
分類指標の策定

リーダー： 准教授 柏野和佳子  
〔共同研究員：6名〕

コーパスに収録されるテキストに付与する、適切で有用な分類指標を設ける。

- 『現代日本語書き言葉均衡コーパス』(BCCWJ)収録書籍テキストを対象に、「**難易**、**主観・客観性**、**硬軟**、**丁寧さ**、**直接的な語り性**」という、5観点の人手分類が重点課題。
- 難易度**に関しては、機械処理と比較分析をし、自動付与の精度向上を目指す。

BCCWJの収録書籍  
テキストに付与済みの  
情報

書誌情報

書名, 著者名,  
出版社, 出版年  
NDC, Cコード

電子化形式のタグ情報

[言語的な階層構造]

記事, 段落, 文,  
見出し, 引用, 発話,  
箇条書き など



新たに付与する分類指標

対象読者(←難易)

- 1 専門家向き
- 2 やや専門的な一般向き
- 3 一般向き
- 4 中高生向き
- 5 小学生・幼児向き

主観・客観性

- 1 とても客観的
- 2 どちらかといえば客観的
- 3 どちらかといえば主観的
- 4 とても主観的

文章の硬軟

- 1 とても硬い
- 2 どちらかといえば硬い
- 3 どちらかといえば軟らかい
- 4 とても軟らかい

文章の丁寧さ

- 1 とても丁寧
- 2 どちらかといえば丁寧
- 3 どちらかといえばくだけている
- 4 とてもくだけている

直接的な語り性

- 1 直接的な語り性あり
- 2 直接的な語り性なし



★10,000テキスト分の分類指標付与情報を公開。  
★テキスト研究や、コーパス活用に寄与。

