

『現代日本語書き言葉均衡コーパス』新聞・書籍・雑誌データの助動詞に対する用法情報付与
加藤祥, 浅原正幸, 山崎誠

2018 年までに、著者によって『現代日本語書き言葉均衡コーパス』のコアデータに含まれる書籍サンプル (PB)、雑誌サンプル (PM)、新聞サンプル (PN) の部分集合 (347,094 語) に対し、『分類語彙表増補改訂版』(国立国語研究所, 2004) の分類語彙表番号を付与する作業が行われた。しかし、分類語彙表には機能語の情報がほとんどないため、自立語のみに意味情報が付与される結果となっていた。そこで、自立語に意味情報(分類語彙表番号)が付与されたコーパスの助動詞にも意味情報(用法情報)を付与することにした。本発表は、作業内容と基礎データのほか、データを用いた調査例を報告する。

作業対象の助動詞は、『現代語の助詞・助動詞』(国立国語研究所報告 3, 1951) の第二部(助動詞)に掲載された 27 種類とし、分類(用法の詳細説明)や用例の確認は同書に拠った。助動詞の用例を収集するにあたっては、辞書見出し語にあたる語彙素を用いた。本作業により、25,892 短単位に助動詞の用法分類が付与された。

この結果、助動詞を用法によって調査することが可能となった。たとえば、本作業の結果を用いて、助動詞が用法別にどのような分布を示しているのか、書籍・雑誌・新聞のレジスタ情報を用いた調査が可能となった。また、自立語に付与されている分類語彙表番号(意味情報)を用いることで、助動詞がどのような意味(分類)の語と接続するときどのような用法で用いられる傾向があるのかを調査することもできる。このほか、用法に特化した助動詞用例の収集が可能となったため、たとえば「類似」用法の助動詞(「ようだ」「みたいだ」「ふうだ」「らしい」など)を手掛かりとした比喩表現の可能性の高い用例の収集などもできるようになった。これらの集計例を示すとともに、分類語彙表番号と助動詞の用法情報を付与したデータを用いたデモンストレーションを行う。