

明治初期の口語語彙と文語語彙
— 『日本語歴史コーパス』の「明治初期口語資料」と『明六雑誌』の語彙比較—
田中牧郎

口語語彙と文語語彙との異なりや相互関係については、平安時代や室町時代の、口語体資料と文語体資料を使って、ほぼ同義で文体対立のある語を抽出する形で、研究が進められてきた。こうした研究は、同時代の口語体資料と文語体資料を形態論情報付きコーパスにして、相互の語彙頻度を比較することや、分類語彙表番号などの意味情報を利用することによって、大規模かつ精確に行うことが期待できる。このような考えから、発表者は、国立国語研究所編『日本語歴史コーパス 明治・大正編』に収録されている、「明治初期口語資料」と『明六雑誌』を、それぞれ口語体資料と文語体資料と扱って、明治初期の口語語彙と文語語彙をとらえる研究を行っている。本発表では、その研究状況を報告し、今後の研究の方向性について議論したい。

上記の2資料の固有名詞を除く全自立語をダウンロードして語彙頻度表を作成し、その語彙素 ID に分類語彙表番号を対応付け、同一の分類項目に属する語彙に焦点をあてていく。総頻度が10以上の語について、口語体資料に特徴的な「口語語彙」、文語体資料に特徴的な「文語語彙」、双方によく見られる「共通語彙」、その中でも特に頻度の高い「基本語彙」の4種に類別する。その上で、同義性の高い語群で、各類別どうしを比較しながら考察する語誌研究を展開する。その考察の際に、コーパスから得られる用例を確認して、多義語の意味区分を行って、分類語彙表番号を用例ごとに引き当てていく作業を行う必要がある。また、頻度の低い語を対象にすることも望まれ、『分類語彙表』非収録語に分類語彙表番号を与える作業も重要である。

以上のように課題は多いものの、有益な研究の展開が見込まれることから、現状では口語体資料に偏りがちなコーパスに、文語体資料を加えていくことが期待される。