「日本語話し言葉コーパス(CSJ)形態素解析結果の分析」

2021/5/2

学籍番号：2J19F508-7

氏名：川崎咲希

　日本語話し言葉コーパス(CSJ)の形態素解析結果から考察した内容を纏める。

**使用データ：**

　分析に使用したデータは以下のとおりである。

1. A01F0067/学会講演(女)
2. S00F0014/模擬講演(女)
3. A01M0020/学会講演(男)
4. S00M0053/模擬講演(男)

**【結果】**

　Mecabを用いて形態素解析を行った結果から、全形態素数に対する品詞の出現頻度割合を調べた。その結果を図１に示す。

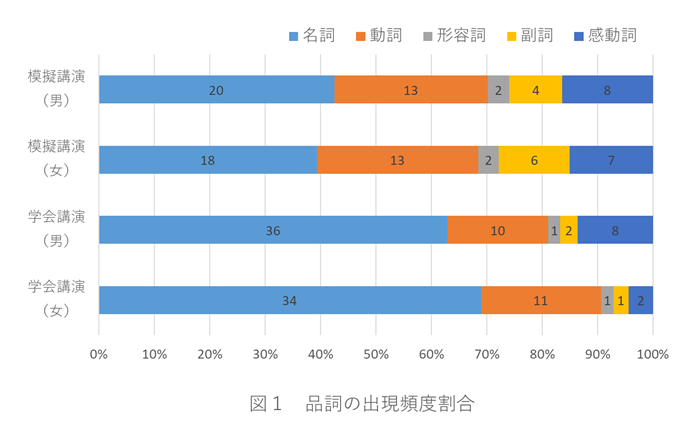


　図１より、学会講演において名詞の出現頻度が模擬講演よりも高くなっている。また、動詞および形容詞、副詞の出現頻度は模擬講演の方が高くなっている。これらの結果は男性、女性の両方において同様であった。感動詞においては講演種類による差はみられなかった。

**【考察】**

講演種類による品詞の出現頻度割合の違いについて調べた上記の結果より、学会講演では模擬講演に比べ名詞が多く使われているといえる。このことから、学会講演の特徴として、専門用語や数字、アルファベットや単位が多用されていることが考えられる。また、模擬講演と比較するとあらたまった話し方であることも名詞の頻度が高い要因と思われる。あらたまった話し方では「さっき」を「先程」と表現し、また「揺れる」「変わる」を「振動」「変換」というように、日常会話では副詞や動詞が使われるところを名詞が用いられることで頻度が高くなると思われた。第２回の考察にて同データで転記分析を行い、学会講演の特徴として、模擬講演に比べ自発性が低く説明的な内容であることを挙げたが、形態素解析の結果からも同様のことがいえる。

　また、模擬講演においては特に副詞が多く使われている。学会講演では数値を用いて物事の程度を表すのに対し、模擬講演では「結構」「ちょっと」「だんだん」といった大まかな表現が用いられていることが要因として挙げられる。「とても」「とにかく」「あんまり」といった強弱を表す副詞も多用されている。このことは、学会講演が実験等から得た事実を客観的に伝えるのに対し、模擬講演の特徴として個人の主観が多く含まれることを表している。第２回の転記分析でも、模擬講演の内容には個人の主観が多く含まれており学会講演と比較すると日常会話に近いものとなっていることを挙げたが、形態素解析の結果からも同様のことが読み取れた。