

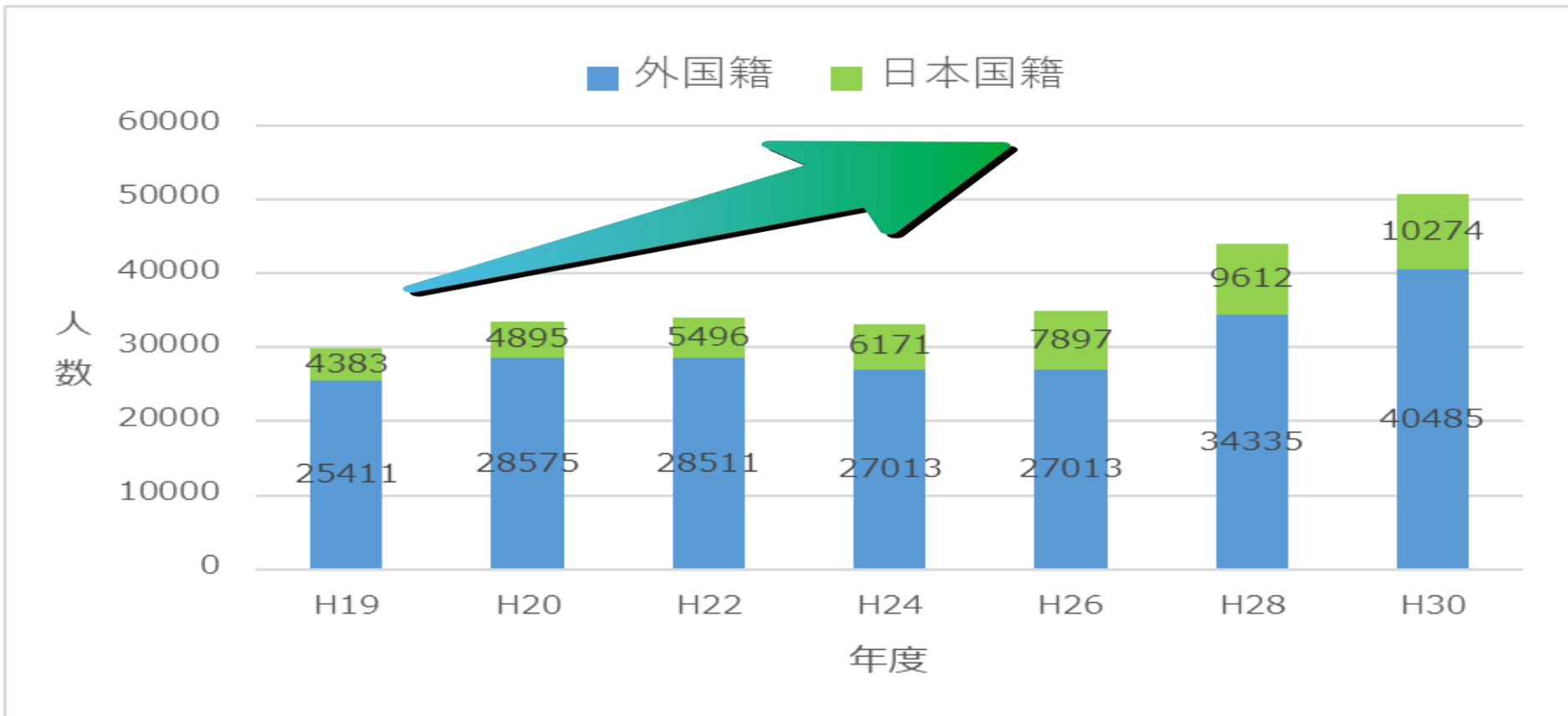
# 小学生向けテレビ教育番組の音声に用いられる語彙の調査

○浅井優介 北村達也（甲南大学） 川村よし子（東京国際大学）

## 背景と目的

近年、日本語指導が必要な児童が増え続けているが、指導方法が明確でないため現場では十分な指導ができていない（文部科学省,2018）.学級活動でよく使われる語彙や表現をリストアップすることができれば、効率的な指導が可能になる。

表1.公立学校における日本語指導が必要な児童生徒数の推移



出展：文部科学省「日本語指導が必要な児童生徒の受入状況等に関する調査」（平成30年度）

## 先行研究

本田（2016）

- 大規模コーパスBCCWJを用いた重要語彙の選出
- 出現頻度と散布度を掛け合わせた有用度指標の提案
- 総語数9463万語
- 総異なり語数：3万6千語

問題点

- BCCWJに収録されているテキストの内容の難易度が小学生向きではない。
- 書き言葉データベースを対象とした調査のため学級活動での会話で使われる語彙と一致しない可能性がある。

NHKのテレビ教育番組は対象学年が設定されているため、その中で使われる語彙や表現は小学生の学級活動で使われるものと対応すると仮定し、語彙の調査を行う。

## 研究方法

1.番組音声の書き起こし



2.形態素解析



3.出現頻度,出現する番組数を用いて重要度のランキング

- NHK for schoolの番組を音声認識を利用して書き起こしたものを手直し

データ抽出条件

- 1番組：30分
- 1教科：最大3番組
- 低学年：1年生から3年生，または1年生から6年生を対象にした番組
- 高学年：4年生から6年生を対象にした番組

形態素解析ソフトウェア「Web茶豆」[4]を利用

- 除外した品詞  
助詞,助動詞,接頭辞,接尾辞,固有名詞,数詞,キャラ語, 接尾辞的形狀詞

語彙の重要度

$$\log f \times \frac{\{d:d \ni t_i\}}{D}$$

$f$ :出現回数  $\{d:d \ni t_i\}$ :単語 $t_i$ が出現する番組数  
 $D$ :全番組数



図1.NHK for schoolのwebページ

書き起こした番組

- 低学年：理科(3), 社会(1), 国語(2), 算数(1), 道徳(2), 生活(1), 特活(2), 特支(3), その他(2), 生活(1)
- 高学年：理科(3), 社会(3), 国語(1), 算数(2), 道徳(2), 音楽など(3), 総合(3), 特活(2)

## 結果

表2.得られた語彙データ数[語]

|     | 番組数 | 総語数    | 総異なり語数 | 1文の平均語彙数 |
|-----|-----|--------|--------|----------|
| 低学年 | 17  | 21,306 | 2,960  | 6.80     |
| 高学年 | 19  | 29,258 | 3,807  | 8.14     |

- 総異なり語は高学年の方が28%多い。
- 1文に含まれる平均語彙数の低学年と高学年の差は有意傾向である。 $t(34)=2.032, p=0.0505$

表3.テキストカバー率[%]

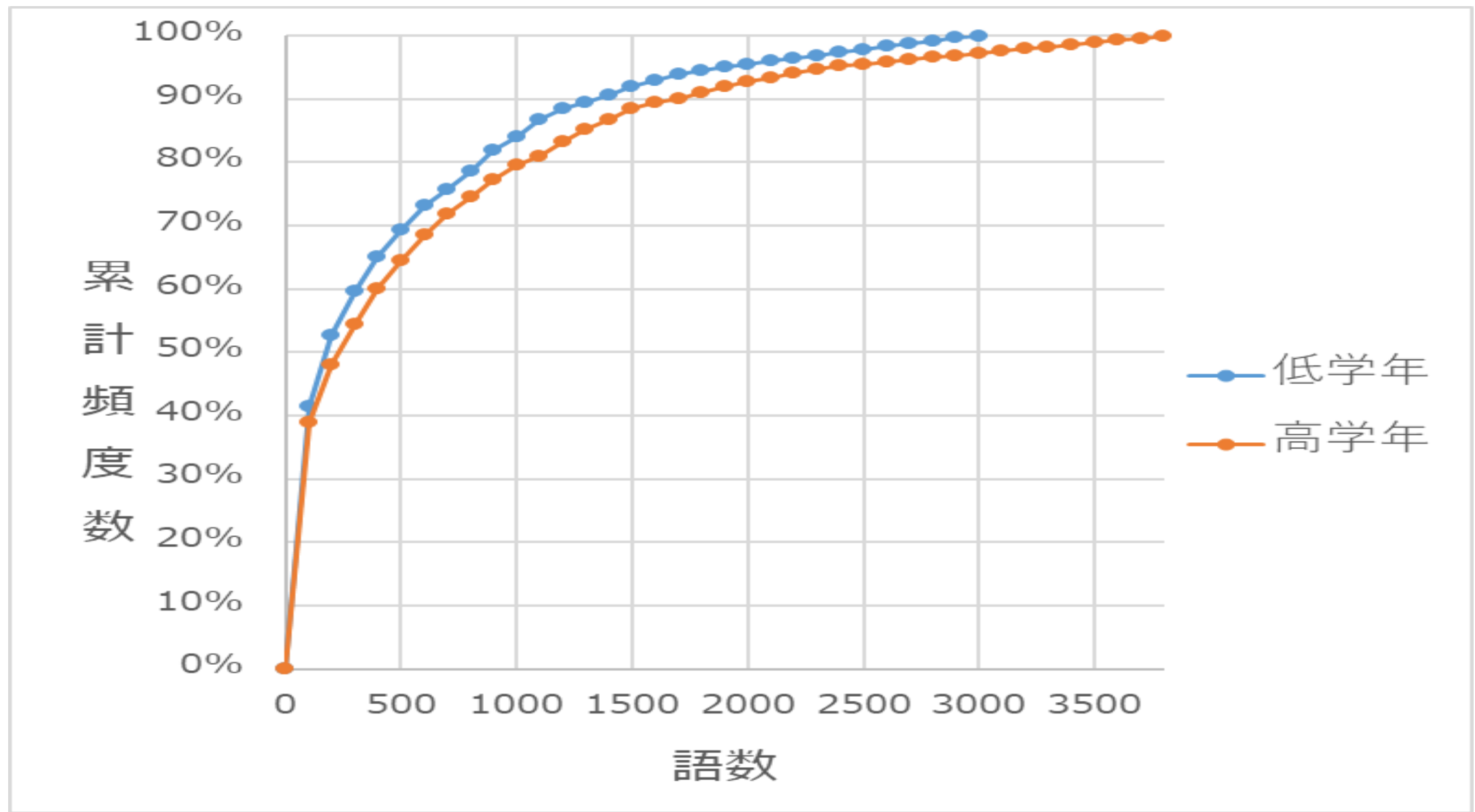


表4.低学年上位30語

| No | 語彙素  | 品詞  | 頻度  | 音声単語親密度 | 級 |
|----|------|-----|-----|---------|---|
| 1  | する   | 動詞  | 701 | 5.188   | 4 |
| 2  | いる   | 動詞  | 328 | 4.875   | 4 |
| 3  | 何    | 代名詞 | 297 | 5.844   | 4 |
| 4  | 見る   | 動詞  | 283 | 6.375   | 4 |
| 5  | 無い   | 形容詞 | 242 | 5.812   | 4 |
| 6  | なる   | 動詞  | 235 | 5.469   | 4 |
| 7  | これ   | 代名詞 | 220 | 4.969   | 4 |
| 8  | 有る   | 動詞  | 217 | 5.688   | 4 |
| 9  | この   | 連体詞 | 187 | 5.5     | 4 |
| 10 | 行く   | 動詞  | 181 | 5.75    | 4 |
| 11 | どう   | 副詞  | 177 | 4.906   | 4 |
| 12 | 事    | 名詞  | 216 | 5.062   | 3 |
| 13 | 言う   | 動詞  | 214 | 5.438   | 4 |
| 14 | 来る   | 動詞  | 195 | 5.781   | 4 |
| 15 | そう   | 副詞  | 184 | 4.875   | 4 |
| 16 | 良い   | 形容詞 | 208 | 4.875   | 4 |
| 17 | それ   | 代名詞 | 110 | 5.719   | 4 |
| 18 | やる   | 動詞  | 141 | 6       | 4 |
| 19 | もう   | 副詞  | 103 | 5.594   | 4 |
| 20 | 皆    | 名詞  | 165 | 5.781   | 4 |
| 21 | 分かる  | 動詞  | 103 | 6.125   | 4 |
| 22 | ここ   | 代名詞 | 94  | 4.781   | 4 |
| 23 | 出来る  | 動詞  | 162 | 6.188   | 4 |
| 24 | くれる  | 動詞  | 76  | 5.625   | 3 |
| 25 | うん   | 感動詞 | 113 | 6.031   | 4 |
| 26 | 物    | 名詞  | 80  | 5.969   | 4 |
| 27 | 僕    | 代名詞 | 134 | 4.844   | 3 |
| 28 | あつ   | 感動詞 | 127 | 5.594   | 2 |
| 29 | 時    | 名詞  | 64  | 5.156   | 4 |
| 30 | ちょっと | 副詞  | 80  | 6.281   | 4 |

表5.高学年上位30語

| No | 語彙素  | 品詞  | 頻度   | 音声単語親密度 | 級 |
|----|------|-----|------|---------|---|
| 1  | する   | 動詞  | 1069 | 5.188   | 4 |
| 2  | いる   | 動詞  | 628  | 4.875   | 4 |
| 3  | 事    | 名詞  | 390  | 5.062   | 3 |
| 4  | 有る   | 動詞  | 349  | 5.688   | 4 |
| 5  | なる   | 動詞  | 338  | 5.469   | 4 |
| 6  | 言う   | 動詞  | 335  | 5.438   | 4 |
| 7  | 見る   | 動詞  | 330  | 6.375   | 4 |
| 8  | この   | 連体詞 | 295  | 5.5     | 4 |
| 9  | これ   | 代名詞 | 284  | 4.969   | 4 |
| 10 | どう   | 副詞  | 246  | 4.906   | 4 |
| 11 | 来る   | 動詞  | 238  | 5.781   | 4 |
| 12 | 良い   | 形容詞 | 218  | 4.875   | 4 |
| 13 | 何    | 代名詞 | 209  | 5.844   | 4 |
| 14 | 無い   | 形容詞 | 249  | 5.812   | 4 |
| 15 | 行く   | 動詞  | 240  | 5.75    | 4 |
| 16 | それ   | 代名詞 | 169  | 5.719   | 4 |
| 17 | 出来る  | 動詞  | 156  | 6.188   | 4 |
| 18 | ここ   | 代名詞 | 127  | 4.781   | 4 |
| 19 | やる   | 動詞  | 121  | 6       | 4 |
| 20 | その   | 連体詞 | 147  | 5.844   | 4 |
| 21 | 皆    | 名詞  | 141  | 5.781   | 4 |
| 22 | そう   | 副詞  | 171  | 4.875   | 4 |
| 23 | ちょっと | 副詞  | 112  | 6.281   | 4 |
| 24 | 物    | 名詞  | 109  | 5.969   | 4 |
| 25 | もう   | 副詞  | 103  | 5.594   | 4 |
| 26 | 分かる  | 動詞  | 114  | 6.125   | 4 |
| 27 | 時    | 名詞  | 86   | 5.156   | 4 |
| 28 | 出る   | 動詞  | 84   | 5.719   | 4 |
| 29 | 所    | 名詞  | 66   | 5.375   | 4 |
| 30 | こちら  | 代名詞 | 76   | 5.781   | 4 |

※級は旧日本語能力試験での難易度であり、4級が最も易しいレベルである。級はリーディングチュウ太[5]を用いて判定した。

## 考察

- 上位で見られた語は旧日本語能力試験の4級かつ、音声親密度5.5以上のものがほとんどであった。したがって小学生が優先的に学ぶべき語彙として適した結果が得られたと考えられる。
- 表2の結果は高学年と低学年では言語能力の違いが現れていると推測できる。

## 課題

- データ量が少ないため、重要語が上位に見られないケースが多少見られた。
- 語彙だけでなく文の係り受け等の構造の調査も必要。

## 参考文献

- 本田ゆかり（2015）「大規模コーパスに基づく日本語教育語彙表の作成」東京外国語大学博士論文
- 天野成昭, 近藤公久（1999）『NTTデータベースシリーズ 日本語の語彙特性（第1期）』三省堂
- 文部科学省初等中等教育局国際教育課（2018）「外国人児童生徒等教育の現状と課題」
- Web茶豆, <http://chamame.ninjal.ac.jp/>
- リーディング・チュウ太, <http://language.tiu.ac.jp/>